

The Simplicity of Transport

Triangulating the First Light

The Simplicity of Transport

Triangulating the First Light

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus prof.mr.dr. P.F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 14 februari 2007
klokke 16.15 uur

door

Nail Guillaume Hubertus Ritzerveld

geboren te Maastricht in 1981

Promotiecommissie

Promotor: Prof. dr. V. Icke

Referent: Dr. M. Sambridge (Research School of Earth Sciences, Canberra)

Overige leden: Dr. G. Mellema (Stockholm Observatory)
Prof. dr. W. van Saarloos (Instituut-Lorentz, Leiden)
Dr. J. Schaye
Dr. T. Theuns (Institute for Computational Cosmology, Durham)
Prof. dr. M. A. M. van de Weygaert (Kapteyn Astronomical Institute)
Prof. dr. P. T. de Zeeuw

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 1 |
| 1.1 | Transport Theory | 1 |
| 1.2 | Radiative Transfer | 2 |
| 1.3 | Numerical Methods | 2 |
| 1.4 | A New Method | 3 |
| 1.5 | Triangulating the First Light | 4 |
| 1.6 | Thesis outline | 4 |
| 1.6.1 | Part I | 4 |
| 1.6.2 | Part II | 5 |
| 1.6.3 | Part III | 5 |
| I | Simplicial Transport | 7 |
| 2 | Transport Theory | 9 |
| 2.1 | Introduction | 9 |
| 2.2 | Kinetic Theory | 10 |
| 2.2.1 | The Microscopic Level | 11 |
| 2.2.2 | Coarse Graining | 12 |
| 2.2.3 | Contraction | 13 |
| 2.3 | Transport Equations | 13 |
| 2.4 | Collision Phenomena | 14 |
| 2.4.1 | Linear Collision Operators | 15 |
| 2.4.2 | The Boltzmann Collision Operator | 16 |
| 2.5 | A Stochastic Interpretation | 19 |
| 2.5.1 | The Master Equation | 19 |
| 2.5.2 | Stochastic Transport | 21 |
| 2.6 | Monte Carlo Methods | 23 |
| 2.7 | Conclusions | 25 |
| 3 | Random Lattices | 27 |
| 3.1 | Introduction | 27 |
| 3.2 | Regular Meshes | 28 |
| 3.3 | Random Lattices | 29 |
| 3.3.1 | Point Processes | 30 |
| 3.3.2 | Voronoi Diagram | 32 |
| 3.3.3 | Delaunay Triangulation | 32 |

| | | |
|-----------|--|-----------|
| 3.3.4 | Poisson Delaunay | 34 |
| 3.4 | Construction Algorithms | 40 |
| 3.4.1 | Quickhull | 40 |
| 3.5 | Applications | 42 |
| 3.6 | Conclusions | 45 |
| 4 | Transport on Adaptive Random Lattices | 47 |
| 4.1 | Introduction | 47 |
| 4.2 | Adaptive Random Lattices | 48 |
| 4.2.1 | The correlation function | 49 |
| 4.3 | Lattice Properties | 51 |
| 4.3.1 | Distributional Equivalence | 51 |
| 4.3.2 | Length Sampling | 51 |
| 4.3.3 | Angular Sampling | 52 |
| 4.4 | Transport | 54 |
| 4.4.1 | Ballistic Transport | 54 |
| 4.4.2 | Transport through Scattering Media | 56 |
| 4.4.3 | General Interactions | 57 |
| 4.4.4 | Time Stepping | 58 |
| 4.5 | 3D and Beyond | 59 |
| 4.6 | Efficiency | 60 |
| 4.7 | Conclusion | 61 |
| 4.A | Fractal Point Processes | 62 |
| 4.B | Conservation of Momentum | 63 |
| 5 | Fields of Application | 67 |
| 5.1 | Introduction | 67 |
| 5.2 | Nonequilibrium Statistical Physics | 68 |
| 5.3 | Graph Theory | 69 |
| 5.4 | Cellular Automata | 71 |
| 5.4.1 | Lattice Boltzmann Methods | 72 |
| 5.5 | Conclusions | 76 |
| 6 | Sampling Theorems for Spatial Point Processes | 79 |
| 6.1 | Introduction | 79 |
| 6.2 | Sampling Theorems | 81 |
| 6.2.1 | Nyquist Sampling | 82 |
| 6.2.2 | Analysis | 83 |
| 6.2.3 | Remarks | 85 |
| 6.3 | Ensemble Approach | 86 |
| 6.4 | Conclusions | 87 |
| II | Triangulating Radiation | 89 |
| 7 | Cosmological Radiative Transfer | 91 |
| 7.1 | Introduction | 91 |
| 7.1.1 | Radiative Transfer | 91 |

| | | |
|------------|---|------------|
| 7.1.2 | A Cosmological Setting | 92 |
| 7.1.3 | Overview | 92 |
| 7.2 | Cosmological Radiative Transfer Equations | 93 |
| 7.3 | Simplifications | 94 |
| 7.3.1 | Local Approximation | 94 |
| 7.3.2 | Quasi-static Approximation | 94 |
| 7.3.3 | Frequency Reduction | 95 |
| 7.3.4 | Front Tracking | 95 |
| 7.4 | Numerical Methods | 97 |
| 7.4.1 | Long Characteristics Methods | 97 |
| 7.4.2 | Short Characteristics Methods | 98 |
| 7.4.3 | Hybrid Methods | 99 |
| 7.4.4 | Monte Carlo Methods | 99 |
| 7.4.5 | Moment Methods | 100 |
| 7.5 | Conclusions | 100 |
| 7.A | Ionization Front Thickness | 102 |
| 8 | SimpleX | 105 |
| 8.1 | Introduction | 105 |
| 8.2 | Code Overview | 106 |
| 8.3 | Implementation | 108 |
| 8.3.1 | Preprocessing Steps | 108 |
| 8.3.2 | Transfer Steps | 111 |
| 8.3.3 | Output | 114 |
| 8.4 | Conclusions | 115 |
| 9 | Code Comparison Project | 117 |
| 9.1 | Introduction | 117 |
| 9.2 | Basic Physics | 119 |
| 9.3 | Simple Ionization Front | 121 |
| 9.4 | A Cosmological Density Field | 126 |
| 9.5 | Conclusions | 129 |
| 9.A | Solving the Rate Equation | 131 |
| III | The First Light | 133 |
| 10 | The Epoch of Reionization | 135 |
| 10.1 | Introduction | 135 |
| 10.2 | Requirements | 137 |
| 10.2.1 | Hydrodynamics | 137 |
| 10.2.2 | Radiative Transfer | 137 |
| 10.2.3 | Box Size and Resolution | 138 |
| 10.3 | Aims | 138 |
| 10.3.1 | The Different Aspects | 139 |

| | |
|---|------------|
| 11 The Diffuse Nature of Strömgren Spheres | 143 |
| 11.1 Introduction | 143 |
| 11.2 Analytic Derivation of the Diffuse Radiation Field | 145 |
| 11.2.1 Homogeneous HI Matter Distribution | 146 |
| 11.2.2 An r^{-1} HI Matter Distribution | 148 |
| 11.2.3 An r^{-2} HI Matter Distribution | 148 |
| 11.3 Numerical Analysis | 149 |
| 11.4 Counter Effects | 152 |
| 11.4.1 Hard Photons | 152 |
| 11.4.2 Redshifting | 152 |
| 11.4.3 Diffusion Speed | 153 |
| 11.5 Conclusions | 154 |
| 11.A General Derivation of the Diffuse Radiation Field | 156 |
| 11.A.1 Case A: $p > d/2$ | 156 |
| 11.A.2 Case B: $p = d/2$ | 157 |
| 11.A.3 Case C: $p < d/2$ | 157 |
| 11.A.4 Overview | 158 |
| 12 Refining Reionization Simulations | 159 |
| 12.1 Introduction | 159 |
| 12.2 Simulation Setup | 160 |
| 12.2.1 Hydro Input | 160 |
| 12.2.2 The SimpleX Lattice | 161 |
| 12.2.3 Source Luminosities | 161 |
| 12.2.4 The Transfer | 162 |
| 12.3 Numerical Effects | 164 |
| 12.3.1 The Effect of Resolution | 164 |
| 12.3.2 Star Formation Histories | 166 |
| 12.3.3 Results | 167 |
| 12.4 Reheating | 169 |
| 12.4.1 Star Formation Rate Effect | 170 |
| 12.4.2 Clumping Effect | 171 |
| 12.4.3 Total Effect | 173 |
| 12.5 Diffuse Photons | 175 |
| 12.5.1 Power Spectrum | 176 |
| 12.5.2 Morphology | 177 |
| 12.6 Conclusions | 180 |
| Nederlandse Samenvatting | 183 |
| Curriculum Vitae | 189 |
| Nawoord | 191 |

Introduction

Since the dawn of time, the Universe has had deviations from being perfectly homogeneous. The exact spectrum of the earliest observable inhomogeneities is imprinted in a pervasive field of prehistoric photons that we see now as the Cosmic Microwave Background (CMB). The local observable Universe is nothing like what is seen in this CMB: the distribution of baryons and dark matter has changed from a nearly smooth primordial soup, to a geometrically complex configuration of filamentary structures in which one may witness the birth of a star, that of a planet, and maybe even that of life.

How does this change come about? Herakleitos' aphorism conveys it most concisely: ΠΑΝΤΑ ΡΕΙ ΚΑΙ ΟΥΔΕΝ ΜΕΝΕΙ. The Universe's content is far from static; its elementary constituents are in constant motion. Whenever there is a gradient, in either pressure or potential, baryonic and dark matter particles move from one location to the next. Photons and other massless particles are constantly on the move, required to travel at the speed of light. If it were not for this natural process, the Universe would still be in the same state as it was as we observe it in the CMB. The transport of particles is therefore *the* essential ingredient in understanding how our Universe changes from one state into the next.

1.1

Transport Theory

The movement of a particle can be tracked by solving its equation of motion, the result of which is a geodesic trajectory along which it will travel unhampered until some interaction (annihilation, scattering, dissociation, etc.) takes place. Most particles travel in groups, though. A drop of water consists of very many individual molecules (of the order of Avogadro's number, i.e. $\sim 10^{23}$), and a laser beam is usually made up out of at least as many photons. This drastically complicates matters, because not only do we need to solve the equations of motion for each and every particle, the particles will also interact amongst themselves, entangling the dynamical history of each individual particle into a complex tapestry of intertwined trajectories. It is therefore near to impossible to describe the behavior of a collection of particles by solving for the trajectory of each individual particle, even though that would be the most accurate way of obtaining their dynamics.

Physicists, confronted with this immense problem, have found a partial way out: instead of trying to describe the behavior of each particle, one can aim for describing the collective behavior of the group of particles as a whole, coarse graining from the microscopic level to a mesoscopic or macroscopic one. The equations of motion for the individual particles will henceforth be replaced by dynamic equations for system averages, quantities that describe the physical properties of the collection of particles. These equations can come in a wide

variety of forms, depending on the problem at hand. The collective behavior of gas particles in a fluid, for example, are described by the Euler or Navier-Stokes equations (Landau & Lifschitz 1987), while the collective behavior of photons traveling through a cloud can be described by the radiative transfer equations (Chandrasekhar 1950).

The general mathematical framework that is used as a context within one can follow the dynamical behavior of a collection of particles is known as *nonequilibrium statistical mechanics* (Liboff 1969; Balescu 1975; Reichl 1998). More specifically, the collective movement of particles, possibly through a host medium, is described by the extremely versatile *transport theory* (Duderstadt & Martin 1979). Both of these subjects are cornerstones of modern physics, having proven their use in the description of gases, liquids, plasmas, and radiation fields, just to name a few. Their principles are much more universal, though. Every system that can be dissected into a collection of individual constituents that interact via a set of (often simple) rules can be understood on a macroscopic scale by using the same machinery. Whether is the flow of traffic on highways (Chowdhury et al. 1997), the exchange of money on the financial market (Dragulescu & Yakovenko 2000), the flow of data on the World Wide Web (Callaway et al. 2000), that of words in a verbally communicating society (Schulze & Stauffer 2005), or even the movement of schools of fish (Czirok & Vicsek 2000), the same principles of transport theory can be used to understand, or at least model, the specific process.

1.2

Radiative Transfer

Historically, one of the first transport problems that was studied was that of photons through a medium, with which interaction may take place. This subject of radiative transfer is still one of the most demanding and intricate problems in modern day physics. The transported entities are photons, particles that move with the highest speed possible, thereby reaching many different parts of space, interacting with it, and being reemitted by it, before arriving at a certain location. The problem is therefore highly non-local, so that its solution is far from trivial. The photons have a certain location, travel into a certain direction, and have a certain frequency, all of which can be time-dependent. Thus, the problem is seven-dimensional: a daunting task.

Its applications are widespread, ranging from laser physics, to understanding nuclear detonations, to even creating convincing 3D animations. In astrophysics, radiative transfer is one of the essential ingredients in understanding cosmological processes. The formation of cosmic structures, such as galaxies and stars, is influenced, indeed sometimes even dominated, by radiative effects. It is therefore mandatory to incorporate this specific transport theoretic process into any analysis of physical cosmology.

1.3

Numerical Methods

The equations that govern transport theoretic problems are often very difficult to solve analytically, closed solutions only existing in very specific, physically and geometrically simple problems. In almost all cases, numerical methods are necessary to obtain an indication of the actual solution, if there exists one at all. Most existing numerical methods start

with the macroscopic equations describing the transport problem, and try to solve these by discretizing both the differential operators and the properties of the particles and medium themselves. In radiative transfer methods, for example, not only the macroscopic radiative transfer equations, but also the properties of the medium through which the photons will propagate are discretized. By using a finite differencing method, solutions are obtained.

Almost all of these methods use a fixed rectangular computational mesh as a basis for the discretization procedure. Choosing this type of mesh has nothing to do with the physical problem at hand. Indeed, using these meshes has been known to introduce symmetry breaking effects, because both rotation and translation invariance are broken. Moreover, the medium distribution itself is often highly inhomogeneous, so that fixed grids are either too coarse in detail-rich regions, or too refined in nearly homogeneous area. The resultant methods can be designed to be very accurate, but most often they become very involved and computationally very laborious.

1.4

A New Method

In this thesis, we present a new method that is able to solve radiative transfer problems, and other processes in which a collection of particles is transported through a background medium. It is radically different from existing numerical methods in two ways.

First, it does not solve the macroscopic, coarse-grained, differential equations that describe the dynamical behavior of a collection of particles as a whole. Instead, it goes back to the original mesoscopic perspective of individual particles moving through the medium along a trajectory taking it from one interaction to the next.

Second, it uses a computational mesh that is drastically different from the usual ones. A point process is used to construct a distribution of points that directly represents the background medium distribution, and, from this, a mesh is constructed by performing a tessellation procedure. The resultant mesh is a graph along which the particles can move, emulating the actual physical Markov process of particles moving from one interaction to the other. Indeed, one can show that, by construction, the line lengths of the adaptive mesh correlate with the local mean free paths of the background medium. Moreover, the choice of the tessellation procedure ensures that both rotation and translation symmetry are preserved, by which no unwanted spurious invariants are introduced into the numerical solution.

Thus, we have defined the method to operate on a mesh that is very physical indeed, and is homogeneous from the perspective of the particles themselves: every mesh line, whether it is short or long, is identical for the transported particle, namely a fixed number of mean free paths. The result is a method that not only has a wide range in resolution, because of its adaptive properties, thereby being able to solve substructure that was previously unresolved, but it can also be shown to be much more efficient than other existing methods. Those scale with the number of sources of particles in the problem; ours does not.

The method combines techniques from different branches of the sciences, and because of its general framework, it has many diverse applications. We shall therefore spend quite some time in this thesis on going into these matters in considerable detail. Of all possible applications, we have singled out one particular one for which we will use our transfer method, and that is the era of cosmological reionization.

1.5

Triangulating the First Light

Largely due to the transport of particles, the Universe changes state. Starting nearly homogeneous at recombination, collections of baryonic and dark matter particles coalesce into a sponge-like large scale structure, under the influence of the gravitational potential of the initial seeds of overdensity. In the high density walls of these structures, the circumstances can be such that the present gas can contract and cool sufficiently to host the formation of the first stars.

These sources provide the first new supply of photons, and their impact on the matter in the Universe is drastic. Being very massive, the first stars form photons that are energetic enough to ionize the hydrogen and helium that had recombined earlier. Around each source HII regions are blown, which grow with increasing time. This process culminates in the HII bubbles overlapping, so that the Universe is filled and again becomes transparent to ionizing radiation. This process occurs at the end of the so-called *Dark Ages*, and is known as the *Epoch of Reionization*. Currently, it is one of the most studied topics in cosmology.

When wanting to study this epoch numerically, it is immediately evident that transport methods are essential. The collective movement of individual gas and dark matter particles can be simulated by the use of cosmological hydrodynamics methods, and the transport of ionizing photons needs a numerical radiative transfer method. Of these, the hydrodynamics methods are the most efficient. The cosmological radiative transfer methods form the bottleneck in current reionization simulations: they scale with the number of sources. Reionization is induced by very many sources that are distributed very inhomogeneously, which is why standard radiative transfer computation is very expensive.

Our new method does not scale with the number of sources, and uses a mesh that adapts to the medium distribution. With an implementation of this method, we are therefore in the fortunate position to be able to more systematically study several aspects of this intricate cosmological era, that were previously beyond computational feasibility. We shall describe the results hereof in this thesis.

1.6

Thesis outline

This thesis is divided into three distinct parts: definition, implementation, and application.

1.6.1 Part I

The first part gives a general introduction of our new transfer method, putting emphasis on its inner workings, but also on its versatility and generality.

We commence by giving a general introduction into the topic of transport theory in Chapter 2, emphasizing certain aspects that will serve as a general framework within which we can more aptly explain our method. Hereafter, in Chapter 3, we will discuss the properties and construction techniques of the adaptive mesh used by our method, and, that being done, we are in a position to describe in detail how the method works in Chapter 4. In Chapter 5, we emphasize the method's versatility and generality by going into its link to other branches of science, pointing out several of its possible uses, and we finish this part of the thesis by

quantitatively deriving sampling criteria for the method's adaptive random lattice in Chapter 6.

1.6.2 Part II

The second part of this thesis describes a specific implementation of the more general method, aimed at doing cosmological radiative transfer.

We first give a general introduction of the current status of cosmological radiative transfer in Chapter 7, pointing out the ingredients essential for any numerical method. Hereafter, we describe in Chapter 8 how our new method was implemented into a C++ package, called `SimpleX`, specifically designed to be used within the context of reionization simulations. Using several test cases, `SimpleX` was compared with other cosmological radiative transfer code in an international comparison project, the results of which are presented in Chapter 9.

1.6.3 Part III

The third and final part of this thesis describes the results of using the cosmological radiative transfer implementation, `SimpleX`, to perform cosmological reionization simulations.

Because the radiative transfer used to be the bottleneck in reionization simulations, several physical aspects of the process of reionization are usually ignored, in order to make the simulations computationally feasible. In Chapter 10, we give a brief introduction of reionization simulations, describing their numerical and physical requirements, and we discuss several of the aspects that were usually underemphasized. One of these aspects is the influence of the diffuse radiation field. In Chapter 11, we show quantitatively that diffuse photons may even dominate the radiation field within HII regions, so that their influence cannot be ignored. We finish this thesis by presenting the results of using `SimpleX` to perform many different reionization simulations in Chapter 12, showing that several numerical and physical aspects have a profound influence on the overall reionization history.

Bibliography

- Balescu, R. 1975, *Equilibrium and Nonequilibrium Statistical Mechanics* (New York: Wiley-Interscience)
- Callaway, D. S., Newman, M. E. J., Strogatz, S. H., & Watts, D. J. 2000, *Physical Review Letters*, 85, 5468
- Chandrasekhar, S. 1950, *Radiative Transfer* (London: Oxford University Press)
- Chowdhury, D., Ghosh, K., Majumdar, A., Sinha, S., & Stinchcombe, R. B. 1997, *Physica A Statistical Mechanics and its Applications*, 246, 471
- Czirok, A. & Vicsek, T. 2000, *Physica A Statistical Mechanics and its Applications*, 281, 17
- Dragulescu, A. & Yakovenko, V. M. 2000, *European Physical Journal B*, 17, 723
- Duderstadt, J. J. & Martin, W. R. 1979, *Transport Theory* (New York: J. Wiley)
- Landau, L. D. & Lifschitz, E. M. 1987, *Fluid Dynamics* (Oxford: Pergamon)
- Liboff, R. L. 1969, *Introduction to the Theory of Kinetic Equations* (New York: J. Wiley)
- Reichl, L. E. 1998, *A Modern Course in Statistical Physics*, 2nd edn. (New York: Wiley-Interscience)
- Schulze, C. & Stauffer, D. 2005, *International Journal of Modern Physics C*, 16, 781

Part I

Simplicial Transport

Simplicity is the final achievement.
After one has played a vast quantity of
notes and more notes, it is simplicity that
emerges as the crowning reward of art.

FREDERIC CHOPIN

Transport Theory

In this chapter, we will lay a foundation for our new numerical method, by delving into the particulars of the physics of transport processes. We will give two different perspectives, one based on a deterministic approach, the other on a stochastic one. It is the latter class that has led to a class of very versatile numerical methods for solving the transport equations, now known as Monte Carlo methods. It is this stochastic approach that we will use as a basis for our new transport method, as discussed in Chapter 4.

2.1

Introduction

Transport theory is most commonly defined as the mathematical description of the transport of particles through a host medium (Duderstadt & Martin 1979). The theory has proven its use in modeling a wide variety of physical, or physics related, phenomena, in which one wants to study the behavior of a large number of particles interacting with a medium. Examples hereof are abundant, from the transport of neutrons through the uranium fuel elements of a nuclear reactor, to the transport of photons through the intergalactic medium, from the analysis of traffic flowing along a predefined system of highways, to the motion of a gas within a wind tunnel (a more extensive range of possible uses will be given in Chapter 5). There is a subtle distinction between the first two examples, in which the medium can be assumed to be distinct from the transported particles, and the latter two, in which the medium consists of the particles themselves. This distinction will give rise to very different properties of the governing equations, cf. Sect. 2.4.

In almost all cases, the particles that are transported are very numerous, i.e. on the order of Avogadro's number ($\sim 10^{23}$). This makes it impossible to track all particles individually. As such, transport equations are derived making use of the same machinery developed within the discipline of *nonequilibrium statistical mechanics*. Strictly speaking, transport theory is a restricted subset of *kinetic theory*, in which one wants to derive equations for macroscopic observables based on what we know of the microscopic details of the process. It is therefore appropriate to start this chapter with a brief introduction of kinetic theory in Sect. 2.2. Hereafter, we derive the general form of the transport equation in Sect. 2.3. The set of transport equations is not closed until we give a, preferably detailed, description of the interactions between particles and medium (or amongst the particles themselves). It is this collision term that gives rise to the wide diversity of forms of the transport equation, but we shall categorize them into two distinct classes in Sect. 2.4.

Although the transport equations were, chronologically speaking, originally derived by coarse graining a many body problem into more tractable equations for particle distribution functions, our understanding of these transport phenomena has deepened tremendously by taking a different approach. Because kinetic theory has an inherently stochastic character, it is obvious that we can take an alternative approach to mathematically describing transport problems by the use of stochastic differential equations, used in the field of *stochastic processes*. We shall describe some aspects of this approach in Sect. 2.5.

This stochastic approach is very useful, not only from a theoretical perspective, but also because from it stems one of the most versatile ways of numerically tackling the transport equations. This branch of transport solvers has been given the apt name of *Monte Carlo* methods, and because these lie at the basis of our new method described in Chapter 4, we will describe the main idea and some of the details in Sect. 2.6.

2.2

Kinetic Theory

Kinetic theory can be defined as the analysis of nonequilibrium physical phenomena that emerge from the collective behavior of a large number of particles. Many excellent introductions have been written on the subject, some of the definite ones being Balescu (1975) and Reichl (1998). Here, we briefly review this subject from the more general perspective of nonequilibrium statistical mechanics, mainly along the lines followed in Duderstadt & Martin (1979).

Any transport process can in principle be solved by keeping track of all the individual particles involved, and solving for them individually when possible, or for the collective transport when the particles can interact amongst themselves. However, in practice, it is impossible to do this, given the computational task involved, not to mention the neigh to impossible task of determining the initial conditions of each individual particle. The primary goal of kinetic theory is to derive the macroscopic behavior of such many particle systems starting from the microscopic dynamics of the particles involved, by performing one or more steps of coarse graining.

In analyzing the flow of gas, for example, we might study the actual equations of motion for each and every gas particle, or we could 'zoom out' (i.e. coarse grain) and give a more approximate description using the well-known *Boltzmann equation*. In a similar manner, we could zoom out even further and describe the gas on a hydrodynamical level (e.g. by using the *Navier-Stokes equations*). Which level of coarse graining is acceptable depends on the problem at hand, and we will give a more quantitative measure for the applicability of each level in Sect. 2.4.2.

The mechanisms for passing from the level of the microscopic equations of motion to the level of the kinetic transport equations and eventually to the hydrodynamical limit make up the discipline of nonequilibrium statistical mechanics. In each step of coarse graining, averages are taken over the possible microscopic motions of the particles in the system. These averages then emerge as macroscopic properties of the system, such as the pressure or the temperature.

2.2.1 The Microscopic Level

We start by focusing on the lowest level. On a microscopic level, the dynamical evolution of particles is governed by the well-known laws of mechanics:

$$\begin{aligned}\frac{d\mathbf{x}_i}{dt} &= \mathbf{v}_i(t) \\ \frac{d\mathbf{v}_i}{dt} &= \frac{1}{m}\mathbf{F}(\mathbf{x}_i(t), t),\end{aligned}\quad (2.1)$$

for every particle $i = 1, \dots, N$. Given the initial particle and momentum value of all N particles, $\Gamma_N(0) \equiv (\mathbf{x}_1, \mathbf{v}_1, \dots, \mathbf{x}_N, \mathbf{v}_N)$, defined as a point in $6N$ -dimensional *phase space*, the linear system Eq.(2.1) fully determines the trajectories $\Gamma_N(t) = (\mathbf{x}_1(t), \mathbf{v}_1(t), \dots, \mathbf{x}_N(t), \mathbf{v}_N(t))$ of all N particles through phase space.

Any measurement on the system will most probably be a time average over a part of one of the trajectories. If A is a function of the state of the system, i.e. $A = A(\Gamma_N)$, a physical measurement is associated with the time average of A , defined as

$$\langle A \rangle_T = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt A(\Gamma_N(t)). \quad (2.2)$$

It is impossible to solve the equations of motions for a system of N interacting particles directly, if N even remotely approaches, say, the number of atoms in a kilogram of gas. Moreover, physical observations assuredly do not match that level of detail. Therefore, we could take a different approach, due to Gibbs, in which we do not consider just one single dynamical system, but an *ensemble* of systems. Each member of this ensemble is identical with respect to the gross macroscopic variables that can be specified (such as the total energy of the system), but have an otherwise unspecified phase space distribution. The ensemble can be fully specified by the ensemble distribution function $\rho(\Gamma_N, t)$. Given this distribution function, we can define another average of the dynamical variable A , namely the *ensemble average*, i.e.

$$\langle A \rangle \equiv \int d\Gamma_N \rho(\Gamma_N, t) A(\Gamma_N). \quad (2.3)$$

The ensemble average and the time average of a dynamic variable A can be related to each other, when one assumes that the ensemble averages are representative for the measured macroscopic properties of the system. The modern theory of statistical mechanics hinges on this so-called *ergodic theorem*, which can be represented as

$$\int d\Gamma_N \rho(\Gamma_N, t) A(\Gamma_N) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt A(\Gamma_N(t)). \quad (2.4)$$

One can circumvent all the issues involving the rigorous justification of the ergodic theorem by noting that all measured properties of the system can always be expressed in terms of the ensemble averages.

Thus, all information of the many body system is encoded into the distribution function $\rho(\Gamma_N, t)$. As such, the N equations of motion in the system Eq.(2.1) can be replaced by one first order partial differential equation for the ensemble distribution function. Using Hamilton's equations, we obtain the *Liouville equation*

$$\begin{aligned}
\frac{\partial \rho}{\partial t} &= \{H, \rho\} & (2.5) \\
&= -\sum_{i=1}^N \left(\mathbf{v}_i \cdot \frac{\partial}{\partial \mathbf{x}_i} + \frac{1}{m} \mathbf{F}_i \cdot \frac{\partial}{\partial \mathbf{v}_i} \right) \rho(\Gamma_N, t) \\
&\equiv -iL\rho,
\end{aligned}$$

in which $\{ \}$ are the Poisson brackets, and in which we have defined the Liouville operator L . It is important to stress that the solution of the Liouville equation in Eq.(2.5) is formally equivalent to that of the exact equations of motion.

In quantum mechanics similar considerations apply. If ρ is interpreted as a density operator, one obtains the *Neumann equation*

$$\frac{\partial \rho}{\partial t} = [H, \rho] = -iL\rho, \quad (2.6)$$

with an ensemble average for an observable A defined as $\langle A(t) \rangle = \text{Tr} \{ A\rho(t) \}$

2.2.2 Coarse Graining

Boltzmann derived his kinetic equation (cf. Sect. 2.4.2) on a somewhat phenomenological basis. During the last 60 years, however, many elaborate, albeit sometimes very abstract, schemes have been developed to systematically move from the complicated microscopic equations of motion (e.g. the Liouville equation) to a more simple, tractable macroscopic picture of the many body problem. Each and every one of these approaches involve approximations to some extent, depending on the properties of the physical problem at hand. Most notably, the BBGKY hierarchy procedure (Balescu 1975; Reichl 1998) has proven its virtue in rigorously deriving Boltzmann-like equations from the microscopic dynamics.

Many subtle difficulties arise when taking these coarse graining steps. The equations of motion of the microscopic many body system in the form of the Liouville or Von Neumann equation are fully time-reversible. That is, the transformation $t \rightarrow -t$ leaves the form of the equations, and solutions, the same (this is simply because in mechanics all time derivatives are second order). Indeed, it is this time symmetry that enforces the conservation of energy within the system, via Noether's theorem (Noether 1918). When coarse graining these equations of motion to the evolution of the distribution function for just one particle (as in the Boltzmann equation), time symmetry is broken. Using his famous *H-theorem*, Boltzmann's equation can be used to show that the system will evolve towards an equilibrium state of maximum entropy, and *not* vice versa. Note, that the system *can* fluctuate towards a state of lower entropy, but that the overall trend will always be towards the maximum entropy state. These fluctuations may resolve the apparent dichotomy between the second law of thermodynamics and Poincaré's recurrence theorem¹ (Barreira 2005). When moving one step further, coarse graining from the kinetic level to the hydrodynamic limit, new irreversible processes are introduced. There are many interpretations as to how this time symmetry breaking can be understood, and even be linked to the *arrow of time* (Coveney 1988), but that is beyond the scope of this thesis. In the following, we will just briefly sketch the procedure of coarse graining from one level to the next.

¹Loosely defined, it states that all conservative dynamical systems with finite energy are quasi-periodic. That is, states tend to recur.

2.2.3 Contraction

Most dynamical variables of interest depend only on the phase space coordinates of one particle. As such, we can contract the phase space distribution function $\rho(\Gamma_N)$ for all the N particles onto a single-particle distribution function via

$$f(\mathbf{x}_1, \mathbf{v}_1) \equiv \int d\Gamma_{N-1} \rho(\Gamma_N, t). \quad (2.7)$$

The ensemble average for the dynamic variable A in Eq.(2.3) is replaced by

$$\langle A(t) \rangle = \int d^3x_1 \int d^3v_1 f(\mathbf{x}_1, \mathbf{v}_1, t) A(\mathbf{x}_1, \mathbf{v}_1). \quad (2.8)$$

Equations that describe the time evolution of the single-particle distribution function $f(\mathbf{x}_1, \mathbf{v}_1, t)$ are called *kinetic equations*, of which the Boltzmann equation is the most famous.

We can take one step further by obtaining equations for the ensemble averages $\langle A(t) \rangle$ themselves. Such system of equations can take the form of a diffusion equation, or of the equations of hydrodynamics, which will be derived more explicitly in Sect. 2.4.2. For now, we shall proceed by deriving the general form of the transport equations.

2.3

Transport Equations

Transport equations are usually derived on the kinetic level, giving rise to kinetic equations, such as the Boltzmann equation. There are several procedures to derive a general form of these transport equations. The simplest one, in our opinion, equates the substantial derivative, that describes the local time rate of change of the particle phase space density along a trajectory, to the change in density due to collisions and sources:

$$\frac{Df}{Dt} = \left(\frac{\partial f}{\partial t} \right)_{\text{coll}} + s. \quad (2.9)$$

Explicitly expanding DN/Dt as

$$\begin{aligned} \frac{Df}{Dt} &= \frac{\partial f}{\partial t} + \frac{\partial \mathbf{r}}{\partial t} \cdot \frac{\partial f}{\partial \mathbf{r}} + \frac{\partial \mathbf{v}}{\partial t} \cdot \frac{\partial f}{\partial \mathbf{v}} \\ &= \frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} + \frac{\mathbf{F}}{m} \cdot \frac{\partial f}{\partial \mathbf{v}}, \end{aligned} \quad (2.10)$$

we obtain the general form of the transport equation for particles moving in phase space:

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} + \frac{\mathbf{F}}{m} \cdot \frac{\partial f}{\partial \mathbf{v}} = \left(\frac{\partial f}{\partial t} \right)_{\text{coll}} + s(\mathbf{r}, \mathbf{v}, t). \quad (2.11)$$

The source term $s(\mathbf{r}, \mathbf{v}, t)$ is mostly assumed to be given, but may depend on the phase space density $f(\mathbf{r}, \mathbf{v}, t)$ itself, as is mostly the case in radiation transport.

To proceed, we must specify the exact form of the collision term in order to adequately describe the interaction processes of the particles with the medium or amongst the particles themselves. A detailed description of the possible forms of the collision operator will be given in the next section. For now, we introduce several useful parameters that can be used to characterize the interaction terms.

When assuming that the interactions are localized, i.e. long range forces can be neglected, we can introduce the concept of a *mean free path*, which is the average path length of a particle between two interaction, or collision, events. Assuming, for now, that particles move through a static background medium, the mean free path λ can be related to the medium density through

$$\lambda^{-1}(\mathbf{r}, \mathbf{v}) = n(\mathbf{r})\sigma(\mathbf{v}) \equiv \Sigma(\mathbf{r}, \mathbf{v}), \quad (2.12)$$

in which $n(\mathbf{r})$ is the background medium density, and $\sigma(\mathbf{v})$ is the total microscopic cross section for all the relevant interactions, and in which we have defined $\Sigma(\mathbf{r}, \mathbf{v})$ as the macroscopic cross section. From this macroscopic cross section, we can infer several other parameters of the process, such as the *collision frequency* $v\Sigma(\mathbf{r}, \mathbf{v})$ and the *reaction rate density* $v\Sigma(\mathbf{r}, \mathbf{v})f(\mathbf{r}, \mathbf{v}, t)$.

As an example, we examine a scattering process, in which particles move from one scattering event in the medium to another. We can introduce the scattering probability function $f(\mathbf{v}' \rightarrow \mathbf{v})$, which defines the probability that a particle with velocity \mathbf{v}' will be scattered and end up with a velocity \mathbf{v} . Thus, we obtain the collision kernel

$$\Sigma(\mathbf{r}, \mathbf{v}' \rightarrow \mathbf{v}) = \Sigma(\mathbf{r}, \mathbf{v}')f(\mathbf{r}, \mathbf{v}' \rightarrow \mathbf{v}), \quad (2.13)$$

which satisfies $\Sigma(\mathbf{r}, \mathbf{v}) = \int d^3v' \Sigma(\mathbf{r}, \mathbf{v}' \rightarrow \mathbf{v})$ by definition. The collision term can now easily be obtained, by incorporating the reaction rate density for particles scattering from velocity \mathbf{v}' into \mathbf{v} (gain term) and vice versa (loss term). Thus,

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \int d^3v' [v'\Sigma(\mathbf{r}, \mathbf{v}' \rightarrow \mathbf{v})f(\mathbf{r}, \mathbf{v}', t) - v\Sigma(\mathbf{r}, \mathbf{v} \rightarrow \mathbf{v}')f(\mathbf{r}, \mathbf{v}, t)]. \quad (2.14)$$

This equation has the well-known form of a Master Equation for Markov stochastic processes (Van Kampen 1981). This correspondence will form the basis for the stochastic interpretation of the transport equations in Sect. 2.5.

2.4

Collision Phenomena

Transport equations in the form of Eq.(2.11) are exact, in the sense that they can be easily obtained by manipulation of the microscopic equations of motion, such as the Hamilton equation or the Liouville equation. It is the collision term in Eq.(2.11) that incorporates all the detailed physics of the microscopic interactions. To make the problem more tractable, one is usually forced to introduce certain approximations, depending on the relevant physics. In general, there are two distinct classes of problems, and thus two classes of collision operators.

First, there is the case in which test particles (neutrons, photons, etc.) move through a background medium. This process can, in general, be described by linear collision terms, and the resulting transport equations can be readily tackled with the techniques of linear mathematical analysis. This linearity breaks down, when the test particles can interact amongst themselves. This can happen, either when the density of particles becomes high enough to allow for this self-interaction, or when source terms are introduced that depend on the particle distribution, such as can be the case in radiation transport.

This self interaction is not an exception, but the rule, when it comes to the transport of gases, or a plasma (or other collections of particles, which behave like a gas, such as traffic). In this case, the background medium is the gas itself, and the resulting collision terms will

be nonlinear. Clearly, the rate of interaction will depend on the probability that at least *two* particles will find themselves near the same location in (phase) space.

There is a subtle distinction between these two cases. Because in the first case, the feedback of the particles on the medium, such as recoil effect or heating, is mostly neglected, the governing equations will only conserve particle numbers, but not momentum and energy. In the second case, however, these feedback effects are fully accounted for, and the resultant transport equations therefore *do* conserve momentum and energy. This distinction becomes apparent when moving to the continuum limit: the equations of hydrodynamics are nothing more than macroscopic equations describing the conservation of mass (particle number), momentum and energy.

2.4.1 Linear Collision Operators

When the mean free path can be considered to be independent of the phase space density, the transport equation will transform into a linear partial differential equation, which can be solved relatively easily, given the proper initial conditions and the exact geometry of any background medium. Linear transport equations are used in a wide variety of problems, from the transport of neutrons in a reactor, to the transport of high energy electrons through an atmosphere, to name a few. Here, we single out one example, mainly because the latter two parts of this thesis revolve around this subject.

Radiative transfer can be defined to encompass all phenomena related to the propagation of electromagnetic radiation through and its interaction with matter, as long as it can be described by a transport equation. The transport of photons through a medium is often considered to be one of the most difficult linear transport problems. This is mainly because the complexity is enlarged by the fact that the photon mean free paths usually depend on the frequency of the radiation. Moreover, if radiative feedback is included, the optical properties (i.e. the cross sections) of the medium may change due to the radiation field. Thus, the problem has in fact become a highly *nonlinear* problem.

We can use the transport equations of kinetic theory by considering the electromagnetic radiation to be composed of a 'photon gas' (Chandrasekhar 1950). Instead of using the phase space distribution function $f(\mathbf{x}, \mathbf{v}, t)$, it is customary in the field of radiative transfer to use the *radiation specific intensity*

$$I_\nu = h\nu cf, \quad (2.15)$$

in which c is the speed of light. The frequency ν has replaced energy $E = h\nu$ as an independent variable. More specifically, the specific $I_\nu(\mathbf{r}, \boldsymbol{\Omega}, t)$ intensity defines the amount of radiation of frequency ν , at a certain location \mathbf{r} , moving into a certain angular direction $\boldsymbol{\Omega}$, at time t .

We can define several moments of the specific intensity, for example the *average density*, the *flux vector*, and the *radiation pressure tensor*:

$$\begin{aligned} J_\nu(\mathbf{r}, t) &\equiv \frac{1}{4\pi} \int d\boldsymbol{\Omega} I_\nu(\mathbf{r}, \boldsymbol{\Omega}, t) \\ \mathbf{q}_\nu(\mathbf{r}, t) &\equiv \int d\boldsymbol{\Omega} \boldsymbol{\Omega} I_\nu(\mathbf{r}, \boldsymbol{\Omega}, t) \\ \mathbf{P}_\nu(\mathbf{r}, t) &\equiv \frac{1}{c} \int d\boldsymbol{\Omega} \boldsymbol{\Omega} \boldsymbol{\Omega} I_\nu(\mathbf{r}, \boldsymbol{\Omega}, t). \end{aligned} \quad (2.16)$$

Finally, the equation of photon transport can be written as

$$\frac{1}{c} \frac{\partial I_\nu}{\partial t} + \boldsymbol{\Omega} \cdot \nabla I_\nu = j_\nu(\mathbf{r}, \boldsymbol{\Omega}, t) - \alpha_\nu(\mathbf{r}, t) I_\nu(\mathbf{r}, \boldsymbol{\Omega}, t), \quad (2.17)$$

where $\alpha_\nu I_\nu$ denotes the absorption terms, for an *absorption coefficient* α_ν , and j_ν denotes the amount of *radiant energy* emitted (per unit phase space volume). Note that the equation becomes nonlinear, when the process includes re-emission and scattering, because then the j_ν depends on I_ν . The equation of radiation transport Eq.(2.17) can be made to incorporate as many detailed physics as one likes, from photoionization, to Rayleigh, Compton or Thomson scattering. It is not very difficult to write down an equation like that, given the relevant macroscopic cross section for each interaction, but solving it is a very different matter.

Finding analytic solutions to a linear transport equation as in Eq.(2.17) can already be a formidable task. In almost all realistic cases, one has to resort to numerical methods (cf. Sect. 2.6), sometimes partly based on known analytic solutions. The matter gets even more complex, when the collision operator under consideration is nonlinear by definition, as we will discuss in the next section.

2.4.2 The Boltzmann Collision Operator

We already mentioned in the previous section, that some linear collision operators, commonly associated with the transport of particles through a medium, quite easily become nonlinear, when more physics is included. If the collective transport of a system of particles is considered, the collision operator becomes nonlinear by construction.

One of the most well known nonlinear transport equations is the Boltzmann equation. Originally derived in the 19th century to describe the dynamical evolution of rarefied gases, it has proven its use in a wide variety of fields, from the flow of traffic (Bellomo et al. 2002), to the distribution of wealth (Dragulescu & Yakovenko 2000), to even the behavioral patterns of people confined in a room (Helbing et al. 2000). Boltzmann's original derivation was based on somewhat heuristic physical arguments. Finding a more rigorous derivation from the microscopic equations of motion has been the subject of a great deal of research, even today. A very successful attempt at obtaining a deeper understanding involves the use of stochastic processes, something which will be discussed to some extent in the next section. For now, we shall content with giving a derivation much along the lines of Boltzmann's phenomenological one.

Derivation of the Boltzmann Equation

Assuming that the gas is dilute, the mean free path between particle collisions is much bigger than the size of the particles a , i.e. $\lambda \gg a$. This condition ensures it is highly unlikely to have any encounters other than binary collisions. The collision operator will have two different terms, a gain term of particles with a certain initial speed ending up with velocity \mathbf{v} , and a loss term of particles with initial velocity \mathbf{v} ending up with a different velocity due to a collision. Suppose two beams of particles with number density n and n_1 , respectively, have initial velocity \mathbf{v} and \mathbf{v}_1 , and have velocities \mathbf{v}' and \mathbf{v}'_1 after collision. A particle in the second beam experiences a flux $I = n_1 |\mathbf{v} - \mathbf{v}_1|$ of particles from the first beam. The number of collisions δn_c per unit time per unit volume which deflect particles from the second beam through a solid angle Ω is

$$\delta n_c = \sigma(\mathbf{v}, \mathbf{v}_1 | \mathbf{v}', \mathbf{v}'_1) n |\mathbf{v} - \mathbf{v}_1| n_1 d\Omega, \quad (2.18)$$

in which we defined the differential cross section for the encounter $\sigma(\mathbf{v}, \mathbf{v}_1 | \mathbf{v}', \mathbf{v}'_1)$. Given the time-reversibility of molecular processes, we can assume

$$\sigma(\mathbf{v}', \mathbf{v}'_1 | \mathbf{v}, \mathbf{v}_1) = \sigma(\mathbf{v}, \mathbf{v}_1 | \mathbf{v}', \mathbf{v}'_1). \quad (2.19)$$

The total number of collisions (per unit phase space volume) in the loss term can be computed by integrating δn_c over all locations x , all solid angles Ω , and all particle collision velocities \mathbf{v}_1 . Using $n = f(\mathbf{x}, \mathbf{v}, t)d^3v$ and $n = f(\mathbf{x}, \mathbf{v}_1, t)d^3v_1$,

$$C_{\text{loss}} = d^3x d^3v \int d^3v_1 \int d^3\Omega \sigma(\mathbf{v}, \mathbf{v}_1 | \mathbf{v}', \mathbf{v}'_1) |\mathbf{v} - \mathbf{v}_1| f(\mathbf{x}, \mathbf{v}, t) f(\mathbf{x}, \mathbf{v}_1, t). \quad (2.20)$$

Using similar arguments and the reversibility criterion Eq.(2.19), we arrive at a similar expression for the gain term:

$$C_{\text{gain}} = d^3x d^3v \int d^3v_1 \int d^3\Omega \sigma(\mathbf{v}, \mathbf{v}_1 | \mathbf{v}', \mathbf{v}'_1) |\mathbf{v} - \mathbf{v}_1| f(\mathbf{x}, \mathbf{v}', t) f(\mathbf{x}, \mathbf{v}'_1, t). \quad (2.21)$$

Thus, we obtain the Boltzmann equation:

$$\begin{aligned} \frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} + \frac{\mathbf{F}}{m} \cdot \frac{\partial f}{\partial \mathbf{v}} &= \int d^3v_1 \int d^3\Omega \sigma(\mathbf{v}, \mathbf{v}_1 | \mathbf{v}', \mathbf{v}'_1) |\mathbf{v} - \mathbf{v}_1| \\ &\times [f(\mathbf{x}, \mathbf{v}', t) f(\mathbf{x}, \mathbf{v}'_1, t) - f(\mathbf{x}, \mathbf{v}, t) f(\mathbf{x}, \mathbf{v}_1, t)], \end{aligned} \quad (2.22)$$

For convenience's sake, this is mostly abbreviated as

$$\frac{Df}{Dt} = \int d^3v_1 \int d^3\Omega \sigma(\Omega) |\mathbf{v} - \mathbf{v}_1| (f'f'_1 - ff_1) \equiv J(f), \quad (2.23)$$

in which we defined the Boltzmann collision operator $J(f)$.

Properties of the Boltzmann Equation

It can be shown that the integral operator $\int d^3v J(f)$ has 1, \mathbf{v} and v^2 as eigenfunctions, with eigenvalues 0. Thus,

$$\int d^3v J(f) \begin{bmatrix} 1 \\ \mathbf{v} \\ v^2 \end{bmatrix} = 0. \quad (2.24)$$

This expresses nothing more than that we have assumed that the binary collisions are elastic, and therefore conserve mass (particle number), momentum and energy.

Note that the Boltzmann is clearly nonlinear, considering the fact that products of two distribution functions appear in the collision term. However, this is still an approximation, because multi-body collisions, which are not an exception but the rule in denser fluids, are excluded. Because of its nonlinearity, the Boltzmann equation is very difficult to solve, even in the most trivial situations. There is, however, one nontrivial result that is easily obtained from it. Consider an equilibrium solution, in which the phase space distribution function is homogeneous in space and time. That is $f(\mathbf{x}, \mathbf{v}, t) = f(\mathbf{v})$. Then, in the absence of external forces \mathbf{F} , Eq.(2.22) implies $J(f) = 0$, which is satisfied by the condition on $f(\mathbf{v})$ of

$$f(\mathbf{v}')f(\mathbf{v}'_1) - f(\mathbf{v})f(\mathbf{v}_1) = 0. \quad (2.25)$$

It can be demonstrated (Huang 1963) that the most general solution of this equation, given the conservation of momentum of energy, is of the form

$$f_0(\mathbf{v}) = A \exp[-B(\mathbf{v} - \mathbf{C})^2], \quad (2.26)$$

in which A , B , and \mathbf{C} are arbitrary constants. Of course, this is nothing more than the usual *Maxwell-Boltzmann distribution* for velocities in an ideal gas. One can use Boltzmann's *H-theorem* to show that any nonequilibrium state will evolve towards this equilibrium distribution.

We finish this section by noting that the low density approximation $\lambda \gg a$ led to the Boltzmann equation. Using similar steps (Montgomery 1967), one can use a weak-coupling expansion $V/kT \ll 1$ to obtain the *Fokker-Planck equation*, or an expansion in the plasma parameter to obtain the *Vlasov* or *Balescu-Lenard* equation.

The Continuum Limit

It is often desirable to coarse grain the Boltzmann equation Eq.(2.22) even further into macroscopic equations for hydrodynamic variables describing the continuum limit. There are several procedures for obtaining these equations from the Boltzmann equation, such as *Grad's 13-moment method* (Grad 1949) and *generalized polynomial expansions* (Gross et al. 1957).

The most popular scheme for generating hydrodynamics equations is still the *Chapman-Enskog method* (Chapman & Cowling 1991). In this method, one Taylor expands the phase space distribution function with respect to a parameter ζ :

$$f = \sum_{n=0}^{\infty} \zeta^n f^{(n)}, \quad (2.27)$$

in which the $f^{(n)}$ are functions of ζ themselves. The parameter ζ is the *Knudsen number*, defined as

$$\zeta = Kn = \frac{\lambda}{L}, \quad (2.28)$$

which is the ratio between the mean free path λ and the relevant physical length scale L . The Knudsen number is an unbiased parameter to decide whether to use kinetic-level or continuum-level equations to describe the behavior of a many particle system. If the Knudsen number is greater than or near one, kinetic-level equations must be used, and when the Knudsen number becomes very small, the continuum limit can be used.

It turns out that the zeroth order term in the expansion Eq.(2.27) is equivalent to the equilibrium Maxwell-Boltzmann distribution function. The first order expansion term gives the familiar Euler equations, and incorporating the second order term results in obtaining the Navier-Stokes equations. Here, we just give the general form of these hydrodynamics equations, which can easily be obtained by multiplying the Boltzmann equation successively by the conserved variables 1, \mathbf{v} and v^2 , and integrating over \mathbf{v} . The resulting equations are three conservation equations for mass, momentum and energy, and have the general form

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \nabla \cdot \rho \mathbf{u} &= 0 \\ \frac{\partial \rho \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \rho \mathbf{u} &= \mathbf{F}_{\text{ext}} - \nabla \cdot \mathbf{P} \\ \frac{\partial \rho \epsilon}{\partial t} + \mathbf{u} \cdot \nabla \rho \epsilon &= -\nabla \cdot \mathbf{q} - \mathbf{P} : \mathbf{\Lambda}, \end{aligned} \quad (2.29)$$

in which ρ is the fluid density, $\mathbf{u}(\mathbf{r}, t)$ the fluid velocity, and in which we have defined the *pressure tensor* \mathbf{P} , the *heat flux vector* \mathbf{q} and the *viscous stress tensor* $\mathbf{\Lambda}$. The energy ϵ can be related to the local kinetic temperature via $\epsilon(\mathbf{r}, t) = c_V kT(\mathbf{r}, t)$, in which c_V is the heat capacity at constant volume, equal to $2/3$ for an ideal monatomic gas, and k is Boltzmann's constant.

It depends on the problem to be solved whether to use the equation of motions for a many body system on the microscopic, kinetic or continuum level. Each level has its own validity, and its own methods of solution, either analytical, or numerical. However, it so happens that there is an alternative way of approaching many body systems, that has already been hinted at in the previous sections. The theory of stochastic processes, described in the next section, has proven to be a versatile alternative to the deterministic approach of Boltzmann and others.

2.5

A Stochastic Interpretation

Science in general revolves around trying to find systematics in the world around us. From these observations, one hopes to derive a model, preferably in the form of a theory, from which one can extract predictions that conform to these observations. Some theories cry out, from the very beginning, for a probability model. Without feeling the need to justify that approach, one simply makes the assumption that the process has a stochastic character, and can be treated as a *random process*. Examples abound, such as the noise in an electric circuit, the fluctuation of stock value in the exchange market, to the Brownian motion of minute particles immersed in a fluid. The mathematical theory of stochastic processes has been well established and widely used in most of these areas.

On the other hand, we have the older body of physical theory, that of kinetic theory and statistical mechanics, in which a probabilistic treatment is explicitly justified instead of assumed. Indeed, most physicists feel that one should be reluctant with the use of probabilistic methods, and should only resort to such methods when deterministic methods fail. However, in our desire to make the Liouville equation more tractable, we already introduced a (phase-space) probability density, or distribution function, $f(\mathbf{x}, \mathbf{v}, t)$. This function can be used to calculate ensemble averages Eq.(2.8), and, as such, the exact, deterministic microscopic state of the system has been replaced by a stochastic variable. Thus, each physical quantity has become a stochastic process, whose average and higher order moments are now related to observables.

We should of course not forget that quantum theory has an extra unpredictability inherent in its theory. This is not because of the complicated nature of a many body system, as is the case in statistical mechanics, but because, as far as we know, it is just Nature's way of behaving on those small scales.

2.5.1 The Master Equation

We can take a different point of view from the previous sections, and use this stochastic interpretation of many body systems, not as an intermediate step to make the dynamics of a many body system less complicated, but as a way of characterizing the system as a whole. To proceed with explaining how this point of view can be beneficial to understanding and solving transport processes, we need to introduce some preliminary basics of stochastic

processes, more elaborately worked out in the definitive book on the subject, Van Kampen (1981).

In this probabilistic perspective, one generally defines a general stochastic variable $X(t)$, which defines the state of the system. An example of this variable we have already encountered as the phase-space density distribution. The detailed mechanism of microscopic dynamics is then only used to guess, or derive, the formula for the transition probability of the current state to the next. If we assume that the time steps from one state of the system to the next are discrete, we can write

$$X(t+1) = RX(t), \quad (2.30)$$

in which R is an operator describing the transition probabilities. The matrix R is called a stochastic matrix, and has elements that are all non-negative. Moreover, the sum of the elements in each row is one.

Eq.(2.30) is known as the (discrete) *Master Equation*, a term originally invented by Uhlenbeck many years ago. The "Master" in his terminology refers to the fact that Eq.(2.30) gives us precisely, at each time, the probability of any given situation. In the derivation of the Master equation, one usually assumes the stochastic process has a *Markov* character. That is to say, the current state only depends on the previous state,

$$P(X(t+1)|X(t); X(t-1); \dots; X(0)) = P(X(t+1)|X(t)). \quad (2.31)$$

The Markov property is justified for most interaction scenarios. When describing the behavior of gases, for example, the Markov property is satisfied, when only binary collisions occur. This is not always satisfactory, however. The resulting Boltzmann equation, which only describes binary collisions, predicts that viscosity and pressure are independent, but we know empirically that viscosity is highly dependent on pressure. It is obvious that this behavior will be described by equations that do incorporate higher order collisions. The governing equations will then not be of a simple Markov form.

We can make Eq.(2.30) more specific. Suppose we want to know the Master Equation for the transition to a state η , given that the system is in some state at the previous time step. Then we should sum over the transitions from every possible prior state δ to the state η :

$$X(\eta, t+1) = \sum_{\delta} R(\eta|\delta)X(\delta, t). \quad (2.32)$$

In the mathematical literature, this equation is known as the *Chapman-Kolmogorov equation*.

For continuous time processes, the Master equation Eq.(2.30) can be written in its differential form

$$\frac{\partial X(t)}{\partial t} = \Omega X(t), \quad (2.33)$$

in which Ω is the continuous transition operator. Given a time-independent operator Ω , the formal solution of Eq.(2.33) is

$$X(t) = e^{t\Omega}X(0). \quad (2.34)$$

The transition operator Ω thus fully defines the time evolution of the system, given an initial state. As such, spectral analysis can be used to reveal what the operator's eigenvectors are. Most important is the eigenvector with eigenvalue zero, because this corresponds to the asymptotically reached equilibrium state.

Now that we have defined the basic terminology and equations concerning stochastic processes, we proceed by pointing out how these probabilistic models can be used to describe transport processes, in which we again make the distinction between linear and nonlinear transport.

2.5.2 Stochastic Transport

We have already pointed out that the linear transport of particles through a background medium is described and solved much more easily than that of the nonlinear collective transport of gas-like systems. It is therefore not very surprising to find that finding a stochastic process to accurately model the first is much easier than finding one for the latter. Indeed, as it turns out, a stochastic game that accurately describes the macroscopic behavior of gases has not yet been found. As before, we treat the linear and nonlinear transport cases separately, and we will try to give some insight into some of the stochastic models developed for both classes.

Linear Transport Models

Most linear transport problems are accurately modeled by using the analogy of a stochastic random walk of particles through the medium. An example of a simple isotropic random walk is depicted in Fig. 2.1. Because the collision terms are linear, the particles' trajectories can be treated independently, and we can define the random walk to consist of particles moving from one interaction event to the next, until they are destroyed. The walk is fully characterized by the path length of each step. Thus, all we have to do is derive a probability distribution function for the path length, based on the properties of the background medium.

For the sake of simplicity, we assume that the medium is locally homogeneous with density n and that the particles can only interact via one type of interaction with cross section σ . If we then consider an infinitesimal path ds , the probability dp of interaction is linearly proportional to this distance ds ,

$$dp = -\frac{1}{\lambda} p ds, \quad (2.35)$$

in which we introduced a certain proportionality constant λ . We can find the probability $p(s)$ for the first collision occurring at s , by solving Eq.(2.35) given the condition $p(0) = 1$ (the minimum path length is at least 0!). After normalizing, we obtain

$$p(s) = \frac{e^{-s/\lambda}}{\lambda}. \quad (2.36)$$

This is the probability distribution function for the free path length. It has as k -th order moments

$$\langle s^k \rangle = \int_0^\infty s^k p(s) ds = k! \lambda^k. \quad (2.37)$$

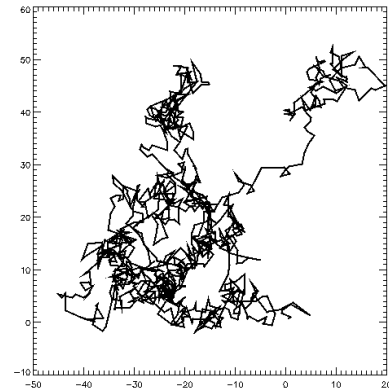


Fig. 2.1: The results of a random walk process in the plane. A thousand steps of equal length are taken in uniformly distributed random directions.

Thus, a typical path has average length equal to λ . This average length is the mean free path, already defined in Eq.(2.12), and is inversely proportional to the medium density n and the interaction cross section σ .

From this description of the path lengths, we can write down a Master Equation for the local number of particles, that describes the number of particles that are destroyed and possibly created, due to these interactions. In most cases, such as scattering or absorption processes, one can show (Chandrasekhar 1943; Duderstadt & Martin 1979; Van Kampen 1981) that these Master Equations transform into the relevant macroscopic transport equations (e.g. the diffusion equation) in the limit. For absorption processes, this derivation is trivial, because the expected analytic behavior of the number of particles along a trajectory, or ray, has exactly the form of Eq.(2.36).

Each interaction has its own cross section, and thus its own mean free path. Henceforth, the probability distribution for the path length can be made to incorporate as many particle-medium interactions as we like. The overall stochastic model remains the same: a particle moves from one interaction to the next, along a trajectory that has a length described by the distribution function Eq.(2.36). This perspective of linear transport will serve as a basis for the statistical numerical transport methods discussed in the next section.

Nonlinear Transport Models

As will be more elaborately explained in the next section, our understanding of the probabilistic nature of linear transport processes has benefited tremendously not only our understanding of the process itself, but also the development of versatile and powerful numerical techniques for these types of processes. In a similar effort, physicists have tried to find a stochastic process that conforms to the requirements imposed by Boltzmann or Boltzmann-like transport problems. This endeavor has been only partly successful; the stochastic processes used give insight into the physical processes behind the Boltzmann equation, but are too much a simplification of realistic transport problems to be used as a starting point for new stochastic numerical methods, except in the most simple problems, for which the solutions are trivial to start with. Thus, statistical methods used in the field of rarefied gas dynamics rely on the methods developed in linear transport theory, and try to incorporate nonlinear behavior using tricks, but more on that in the next section. For now, we shall briefly describe the current status in our understanding of the stochastic process behind the Boltzmann equation.

The stochastic approach to the Boltzmann equation can be traced back to the famous work of Kac (Kac 1956), who, under certain simplifying assumption of spatial homogeneity, formulated a linear Master Equation for an ensemble of interacting gas particles. In the limit of a large ensemble of particles, he showed that initially independent particles remain independent for all times. He called this the "propagation of chaos", because it is this *stoßzahlansatz* or *molecular chaos* assumption that Boltzmann used in his derivation of his famous equation. It was a remarkable feat, that he showed that, given this propagation of chaos, the nonlinear Boltzmann emerged from his linear Master equation. This, for the first time, gave insight into the fact that the Boltzmann equation is not nonlinear by necessity, but only because we chose it to be such, via the chaos assumption. It is important, however, to note that Kac's Master Equation was by no means more easy to solve than the nonlinear Boltzmann equation.

Deeper insight and a rigorous justification of the propagation of chaos was subsequently given in McKean (1966) and Grünbaum (1971). However, because of its simplifications (spatial homogeneity, for example), Kac's model was frowned upon and even mocked. Uh-

lenbeck even termed it a ‘caricature of gas’. Recently, the Kac approach has regained a lot of interest, and many new interesting results have been obtained that relieve the need for the aforementioned simplifications. The results for inhomogeneous systems have been summarized in Graham & Méléard (2001), and one even obtained a rate of convergence to equilibrium for these inhomogeneous cases (Desvillettes & Villani 2005). Without doubt, more results will be obtained in the future, which will hopefully result in a stochastic model for the flow of gas that fully incorporates all of the complexities involved. For now, we will have to do with the stochastic methods that have been developed for linear transport problems, and these will be the subject of the next section.

2.6

Monte Carlo Methods

We started this chapter by introducing a deterministic dynamical model for the microscopic behavior of many body systems. With the use of several wisely chosen approximations, this microscopic behavior can be contracted into more tractable (kinetic) transport equations. In almost all realistic cases, the solution of these equations cannot be found analytically, and one has to resort to numerical methods. Starting from these equations, the most natural choice would be to discretize them onto a uniform grid, and develop a finite differencing method that solves the equations exactly in the limit of infinite resolution. It turns out that there are some inherent difficulties associated with these methods, most notably with the grids used, but this will be discussed in more depth in the next chapter. The stochastic perspective discussed in the previous section, however, provides us with a new way of trying to numerically solve transport problems. Using this approach, we can solve the transport process on a computer without ever referring to the transport equations themselves.

Stochastic methods used as a numerical tool in physics and mathematics are most commonly referred to as *Monte Carlo* methods (Hammersley & Handscomb 1964), a name that was first used by one of the pioneers in the field, Stanislaw Ulam, in honor of his uncle who was a gambler, at the suggestion of Nicholas Metropolis. Originally motivated by the desire to solve the complicated diffusion problems encountered in the early work on atomic energy, Metropolis & Ulam (1949) were the first to develop a computer-oriented Monte Carlo method for transport problems. Since the transport of particles through a medium can be described stochastically as a random walk from one interaction event to the next, the full process can be parametrized by statistical variables only, such as the mean free path or cross sections (cf. Sect. 2.5.2). Therefore, it was a natural step to utilize statistical methods for simulating transport processes. Generally speaking, Monte Carlo methods determine the characteristics of the many body system as statistical averages over a large number of particle histories, which are computed with the use of random sampling techniques. We should note, that there is also a separate, dual, class of Monte Carlo solvers, in which one uses the same random sampling techniques not as a stochastic method to solve stochastic processes, but as a stochastic method to solve deterministic equations, such as the kinetic equations. This branch of Monte Carlo methods has shown various uses in solving complicated integro-differential equations, and the like.

As can be concluded from some of the previous sections, transport problems can, in general, be formulated as

$$\mathbf{D}f = \mathbf{C}f, \quad (2.38)$$

in which D is a drift operator, and C is a collision operator. Depending on one's point of view, Eq.(2.38) can be interpreted as a deterministic equation, or as a (stochastic) Master equation. Using a Monte Carlo method, one tries to solve this equation by trying to simulate a large number of different particle trajectories, or histories. To proceed, one usually decouples the drift and collision operators: one tries to model the relevant physics of each particle-medium interaction as closely as possible (the collision part), whereafter the particle is allowed to freely stream towards the next interaction event (the drift part). Effectively, this approach states that the change of particle number can be separated to be due to collision and streaming, and the result is the sum of the two. It is a strange thing that most people generally accept this approach without any objections: why a sum, and not a product? Surely, the motion of particles is influenced by the short range forces that dictate the collisional behavior? Indeed, Uhlenbeck, using the BBGKY hierarchy technique, derived a Boltzmann equation that does have coupling terms between streaming and collision, and he showed that it was satisfactory to use a sum, when the gradients determining the flow are small. One should always be aware of these caveats, when assuming a seemingly trivial thing like this.

Assuming that this separation is justified, we give an example of the use of Monte Carlo methods in transport problems, by considering the transport of photons through a homogeneous medium. Starting with the collision part, we define the interactions to consist of an absorption, with cross section σ_a , and a scattering event, with cross section σ_s . Now suppose that, starting from a certain point \mathbf{r} , an interaction has occurred. To determine the type of interaction, we compute the probabilities for absorption p_a , and scattering p_s :

$$p_a = \frac{\sigma_a}{\sigma_a + \sigma_s}, \quad p_s = \frac{\sigma_s}{\sigma_a + \sigma_s} \quad (2.39)$$

Since, $p_a + p_s = 1$, we can use a (pseudo-)random number ξ_i chosen on the interval $\xi_i \in [0, 1]$ to decide which event has taken place. If $0 \leq \xi_i < p_a$, then an absorption event has taken place, and if $p_a \leq \xi_i \leq 1$, a scattering event occurred. When the absorption event occurred, we terminate the history of the photon; when scattering occurred, the photon is redistributed into a new random direction, assuming the scattering is isotropic.

Proceeding with the drift, or streaming, part, we must compute the distance the photon will travel until the next interaction event. This distance is the free path length, as given in Eq.(2.36), and is distributed in this case as

$$p(s) = \frac{e^{-s/\lambda}}{\lambda}, \quad (2.40)$$

in which $\lambda^{-1} = n(\mathbf{r})(\sigma_a + \sigma_s)$. In order to randomly sample this distribution function, we may use a commonly used trick known as the *direct inversion method* (Press et al. 1992), which makes use of the fact that the cumulative distribution function of Eq.(2.40),

$$P(s) \equiv \int_0^s ds' p(s') = 1 - e^{-s/\lambda}, \quad (2.41)$$

is a monotonically increasing function of s , that ranges from 0 to 1, when s ranges from 0 to ∞ . As such, it can be inverted, such that

$$s = -\lambda \ln(1 - P(s)), \quad (2.42)$$

and one can use the same random number generator that produces the $\xi_i \in [0, 1]$ as input to Eq.(2.42). When ξ_i is uniformly distributed in its interval, we can substitute it for $P(s)$,

to obtain a path length s that is distributed according to Eq.(2.40). Note that $1 - P(s)$ is also distributed uniformly on the interval $[0, 1]$, so we can determine the path length until the next interaction using

$$s_j = -\lambda \ln \xi_j. \quad (2.43)$$

Thus, we have accounted for the drift part of the transport process, and that is all that there is to it.

There are a few things to note, however. First, as one can see from what we have discussed, Monte Carlo methods heavily rely on random sampling techniques. Therefore, we have to take care to use a pseudo-random number generator that is satisfactory, in that it for example has a large enough cycle (we refer the reader to Press et al. (1992) and Knuth (1997) for a more elaborate treatment). Second, there are certain problems, for which Monte Carlo methods such as we described just now are not suitable. This occurs, for example, when the probabilities for interaction become very small, and thus the mean free paths become very large. The simulation of such events would require a prohibitively large number of histories, but one circumvents this by employing certain *variance-reduction* techniques (Hammersley & Handscomb 1964), in which the original problem is modified so that the interaction event will occur more frequently. Finally, Monte Carlo methods are statistical by nature. This means that the results obtained will be correct on average, but that there will be an inherent noise on top of that. One can use the central limit theorem to analyze this statistical noise and show that, when the number of trials, or histories, N increases as $N \rightarrow \infty$, the average over the trials will asymptotically approach the first order expectation value, but that the standard deviation, or noise, will only decrease with $N^{-1/2}$. This is the famous *inverse square root law* for statistical methods, that dictate the use of very many trials to ensure a numerical prediction that is accurate enough. Fortunately, with the every increasing computer power of today, performing many different trials is, in most cases, no longer a bottleneck.

We conclude this section by a brief discussion of how these Monte Carlo methods can be used to solve nonlinear, collective transport problems. The simplest such applications in gas dynamics are the free streaming problems, in which the interactions can be ignored to a first approximation. For smaller Knudsen number (λ/L), however, one does have to take into account the particle interactions. In the most popular approach to this, the gas is divided into cells, after which the probability of collisions within each cell is sampled based on the particle distribution within each cell. Such methods have been dubbed *Direct Simulation Monte Carlo* (DSMC) methods (Bird 1994), and have become quite popular in simulating complicated flow problems, such as shock wave propagation, and high speed flow past complex geometrical shapes. Although Belotserkovskiy & Yanitskiy (1975), Deshpande (1978) and Nanbu (1980) provided alternative methods with a more solid theoretical basis, the first two based on Kac's Master equation, the Bird scheme is still the main one being used, because it has no rival in computing times.

2.7

Conclusions

The new method that is the topic of this thesis can solve the equations governing the transport of many particle systems, including those through a background medium. This whole matter is strongly rooted in the theory of nonequilibrium statistical mechanics that tries to systematically derive the kinetic transport equations from the microscopic dynamics.

As such, we have provided in this chapter the basics of transport theory, together with the relatively new stochastic perspective in which the dominant class of numerical transport methods, and also our new method is rooted. Building on this foundation, we proceed in the next chapter by delving into the particulars of carefully choosing an appropriate numerical grid on which one can solve the transport equations.

Acknowledgements We are much indebted to Jan Lub and Peter Denteneer for shedding some light on this sometimes somewhat obscure subject.

Bibliography

- Balescu, R. 1975, *Equilibrium and Nonequilibrium Statistical Mechanics* (New York: Wiley-Interscience)
- Barreira, L. 2005, Poincaré recurrence: old and new, XIV International Congress on Mathematical Physics (World Scientific), 415–422
- Bellomo, N., Coscia, V., & Delitala, M. 2002, *Math. Mod. Meth. App. Sc.*, Vol. 12, No. 12, 1801
- Belotserkovskiy, O. M. & Yanitskiy, V. Y. 1975, *Zhurn. Vych. Mat. i Mat. Fiz.*, 15, 1195
- Bird, G. 1994, *Molecular Gas Dynamics and the Direct Simulation of Gas Flows* (Clarendon)
- Chandrasekhar, S. 1943, *Reviews of Modern Physics*, 15, 1
- . 1950, *Radiative Transfer* (London: Oxford University Press)
- Chapman, S. & Cowling, T. G. 1991, *The Mathematical Theory of Non-uniform Gases* (Cambridge University Press)
- Coveney, P. V. 1988, *Nature*, 333, 409
- Deshpande, S. M. 1978, Report 78 FM 4, Dept. Aero. Engng., Indian Int. Science
- Desvillettes, L. & Villani, C. 2005, *Invent. Math.*, 159, 245
- Dragulescu, A. & Yakovenko, V. M. 2000, *European Physical Journal B*, 17, 723
- Duderstadt, J. J. & Martin, W. R. 1979, *Transport Theory* (New York: J. Wiley)
- Grad, H. 1949, *Commun. Pure Appl. Math.*, 2, 331
- Graham, C. & Méléard, S. 2001, *Présentation du CEMRACS 1999*, Vol. 10, Probabilistic tools and Monte-Carlo approximations for some Boltzmann equations (Les Ulis: EDP Sciences), 77–126
- Gross, E. P., Jackson, E. A., & Ziering, S. 1957, *Ann. Phys.*, 1, 141
- Grünbaum, F. A. 1971, *Arch. Rational Mech. Anal.*, 42, 323
- Hammersley, J. M. & Handscomb, D. C. 1964, *Monte Carlo Methods* (New York: Chapman and Hall)
- Helbing, D., Farkas, I., & Vicsek, T. 2000, *Nature*, 407, 487
- Huang, K. 1963, *Statistical Mechanics* (New York: Wiley)
- Kac, M. 1956, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. III, *Foundation of Kinetic Theory* (Berkeley: Univ. of California Press), 171–197
- Knuth, D. E. 1997, *Art of Computer Programming*, Volume 2: *Seminumerical Algorithms*, 3rd edn. (Addison-Wesley Professional)
- McKean, H. P. 1966, *Arch. Rational Mech. Anal.*, 21, 343
- Metropolis, N. & Ulam, S. 1949, *J. Am. Stat. Assoc.*, 44, 247
- Montgomery, D. 1967, *Lectures in Theoretical Physics*, Vol. 7 (New York: Wiley)
- Nambu, K. 1980, *J. Phys. Soc. Jpn.*, 49, 2042
- Noether, E. 1918, *Nachr. d. König. Gesellsch. d. Wiss. zu Göttingen, Math-phys. Klasse*, 235
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, *Numerical Recipes in C: The Art of Scientific Computing* (New York, NY, USA: Cambridge University Press)
- Reichl, L. E. 1998, *A Modern Course in Statistical Physics*, 2nd edn. (New York: Wiley-Interscience)
- Van Kampen, N. 1981, *Stochastic Processes in Physics and Chemistry* (North Holland)

Random Lattices

In this chapter, we will lay the final founding stone for our new transport method. We will argue that Nature has shown us that so-called regular lattices are far from optimal for use in numerical methods, because they break symmetry and introduce spurious invariants. The least restrictive alternative that *does* conserve motion-invariance is the random lattice based upon a (generalized) Poisson point process. The Voronoi diagram, and its geometrical dual, the Delaunay triangulation, have been used widely in many different sciences. Their geometrical properties, and efficient construction techniques are given. We conclude by giving a brief overview of their applications, in Nature and in the sciences.

3.1

Introduction

The numerical description of physical systems that obey a differential expression requires the use of a differencing technique on some sort of grid, mesh or lattice (except in rare cases where computational symbolic algebra can be applied). Accordingly, the question is not if errors are made, but rather what kind, and of what severity. In the past, computational lattices were almost always regular and rectangular, the Cartesian grid being the archetypal example. Analytic estimates of the corresponding errors are routinely made for ordinary differential equations on such grids (e.g. Press et al. 1992); in the case of partial differential equations, especially those of hydrodynamics, the error analysis may even be quite elegant (see e.g. Richtmyer & Morton 1967; Van Leer 1970; Icke 1988).

In almost all of these cases, the numerical method may be chosen in such a way that the errors become small with increasing grid resolution. In practice this is often less than helpful, due to the steep increase of computational effort with decreasing mesh size. Moreover, there is one type of error that does not automatically vanish, namely those effects that are due to the geometry of the grid. Any intrinsic regularity (such as is introduced by using a Cartesian grid) will, by Noether's Theorem (Noether 1918), produce spurious conservation laws. These are sometimes innocuous, but may be rather vicious under some circumstances. One of the simplest examples is that of the Short Characteristics method (Kunasz & Auer 1988) in radiation transport, in which the numerical diffusion along the axes is negligible compared to that along the diagonals, resulting in spiky features along the axes when modeling isotropic outflow (Steinacker et al. 2002).

It is this deficiency in particular that we wish to address in Sect. 3.2. The solution seems to be obvious: pick an irregular grid that is locally isotropic in the mean. The question then

immediately arises, how to construct such a grid, and especially how to design its properties (and the corresponding algorithm that describes the physical process). This will be the subject of Sect. 3.3, in which we introduce the Voronoi diagram and its dual, the Delaunay triangulation, which will serve as a basis for our new transport method. As it turns out, these random lattices have shown their versatility in a wide variety of subjects, in almost all branches of science, including computer visualization disciplines. As such, many powerful, sometimes open-source, packages have been developed that construct these tessellations¹ very efficiently. These construction algorithms will be discussed in Sect. 3.4. We will conclude this chapter by giving a small glimpse of the widespread use of random lattices in the sciences and in Nature, in Sect. 3.5.

3.2

Regular Meshes

Space-time is continuous as far as we know (see e.g. Ambjørn et al. 2006, for recent new insights). Thus, the laws of Nature can be written in the form of equations in which one uses continuous differential operators. A computer's CPU consists of a discrete number of logic gates, that have to be fed one instruction after the other. Thus, in order to utilize the wealth of computer available nowadays in our quest for understanding Nature, we need to discretize the differential operators, but also the state variables, that describe the world around us. The discretization can be done in a variety of ways, as long as one can prove that, in the limit of an infinitely fine grid, the error associated with this discretization of continuous operators and variables approaches zero, and that at the same time the numerical solution approaches the 'exact' one. Note that, in most cases, this exact solution is not known, and that one is therefore usually content with only the first condition (also known as the *consistency condition*) being satisfied. As long as one is careful with defining the corresponding discretized operators, there are, in principle, an infinite number of ways to subdivide space and time into discrete chunks.

Perhaps it is our misguided sense of order that has led us to the belief that Cartesian grids are the optimal solution. One glimpse of the road map of Manhattan might actually strengthen that assumption. However, one can prove that the distance to travel from one random point on a Manhattan road map to another is longer than for, for example, a triangular gridded road map. So, what is 'optimal'? Obviously, the criteria that led to the widespread use of Cartesian grids is that the cell size, morphology and distribution are the same, globally. Moreover, the number of directions (or axes) is minimal. All this makes the regular square, or cubical, grids optimal from an implementation point of view. Nature, however, is hardly ever cubical.

A regular mesh might seem the most obvious choice, but it is known to cause several problems, inherent to its structure. If we want, for example, to sample the background medium in a linear transport problem, and we know that its distribution is highly inhomogeneous, the global resolution of a stiff rectangular mesh will have to be high enough to be able to sample the highest Fourier component of the density spectrum (cf. Chapter 6). This, of course, results in a high redundancy of grid cells in fairly homogeneous regions. This problem has been partly solved by the introduction of Adaptive Mesh Refinement (AMR) (Berger & Olinger 1984; Berger & Colella 1989), in which the grid refinement follows some criterion,

¹A *tessellation* is a space-filling arrangement of polygons (\mathbb{R}^2), polyhedra (\mathbb{R}^3), or polytopes (\mathbb{R}^d).

such as the gradient in the density. Even refined grids, though, suffer from another problem all regular grids have: they are known to break physical symmetries (Nature is not cubical!). Because the cells are congruent, albeit not of the same size, the rotation group of the lattice in d -dimensional space is not isomorphic to $SO(d)$, but to some discrete subgroup thereof. Also, the fixed widths of the cells impose a non-physical constraint, and are known to break translational symmetry. Several different communities have described these shortcomings in different forms. DSMC methods (cf. Sect. 2.6, and Bird 1994) require the mesh size to be much smaller than the local mean free path, in order to accurately capture the flow features. When the density of the gas locally becomes very high, the computational cost will increase tremendously, imposing a difficulty for extending DSMC methods to the continuum limit, where the Knudsen number $\lambda/L \rightarrow 0$. The Lattice Boltzmann (LB) community has shown (Frisch et al. 1986) that the discrete rotation group of a rectangular lattice does not have enough symmetry to obtain Navier Stokes-like equations in the limit, and they had to resort to hexagonal lattices in 2D and multi-speed models in 3D, in the absence of Platonic solids that at the same time have a large enough symmetry group *and* tessellate space (cf. Chapter 5 and Gladrow 2000). Moreover, the regular lattices are known to break Galilean symmetry (Succi 2001). Moreover, the fixed cell-widths impose constraints on the Reynolds number the method can resolve. In the lattice gauge community, it has been known for quite some time that regular (Wilson) lattices break Poincaré symmetry (Christ et al. 1982). Supersymmetry closes on the Poincaré group by necessity, and therefore has difficulty being defined on regular lattices (Kaku 1983). Moreover, because the lattices are invariant under translations of one or more cell-widths, or rotations of $\pi/2$ (or $\pi/3$ when one uses triangular grids) with respect to one of the axes, spurious invariants are introduced.

Accordingly, the choice of a regular grid which has nothing to do with the underlying physical problem results in the introduction of unphysical conserved quantities, and the breaking of several very physical symmetries.

3.3

Random Lattices

Various different communities have independently found an answer to this mesh-related problem: dispense with the regular grids altogether, and introduce *random* lattices, which will be described in more detail later on in this section. In those fields of physics, where symmetries are most important, the lattices were used first. General relativity was discretised onto a simplicial lattice (Regge 1961), even resulting in quantum gravity theories, and lattice gauge theories were defined on similar random lattices (Christ et al. 1982). Their use was hinted at in cellular automaton fluids (Wolfram 1986), but was overlooked as a possibility to solve the dichotomy between symmetric and space-filling lattices in LB solvers (cf. Chapter 5). A review of the use of unstructured grids in LB methods is given in Ubertini & Succi (2005). Voronoi lattices were also recognized to be of use in the field of Dissipative Particle Dynamics (Español 1997; Flekkøy et al. 2000). The use of unstructured grids was also recognized in the field of geophysics, where this type of lattice is used to interpolate irregularly spaced geophysical data (Sambridge et al. 1995). Similar techniques were used to solve PDE's on these type of meshes (Braun & Sambridge 1995).

These random lattices still have a problem, though. In most cases, a Poisson point process lies at their basis, a result of which is that the average point-to-point distance is homogeneously of the order of $n_p^{-1/d}$, in which n_p is the density of points, and d is the

dimension. For an inhomogeneous medium distribution this length scale does not have an immediate correspondence to the length scales of the physical problem, and, as said, impose constraints on the physical parameters (i.e. the Reynolds number for LB methods and the Knudsen for DSMC methods) to be resolved.

In what follows, we shall give a brief mathematical introduction into the topic of random lattices, and, being pragmatic, we shall only point out those facts that shall prove useful for the rest of this thesis. A much more thorough account of random lattices and related topics of *stochastic geometry* can be found in Van de Weygaert (1991), Stoyan et al. (1995), and Okabe et al. (1999), and references therein (this list is far from exhaustive). All of what is discussed in this chapter is applicable in general d -dimensional space, although the examples we give may only be for two or three-dimensional space, for visualization purposes only. We shall delve deeper into the history and wide-spread applications of random lattices in Sect. 3.5.

3.3.1 Point Processes

Regular lattices are always assumed to be given. Adaptive grids are more flexible, in that they can adapt to the medium properties. The random lattices that we will be using for our new transport methods, however, are not *given* ab initio, but are *constructed* on the basis of a distribution of points. That distribution of points can have a wide variety of forms; indeed, we will see in the next chapter, that our new method uses a point distribution that adapts to the background medium properties, by which the unphysical length scales mentioned above are avoided.

In general, the lattices we will use in this thesis will always be constructed on the basis of a *spatial stochastic point process*, which is a process in which points are generated in a space according to a probability function defined over that same space. The term *process* might call to mind a process over time, but this dependence is only implicit. Many different types of point processes have been described in the mathematical literature (e.g. Stoyan et al. 1995), but there is one that is without doubt the most fundamental, mainly because of its simple definition on the one hand, and its widespread use on the other. The *Poisson point process* can be defined as follows: given a volume $S \subset \mathbb{R}^d$, and $N(A)$ as the number of points in any non-empty subset A of S , the probability that A contains x points is

$$\Pr(N(A) = x) = \frac{n_p |A| e^{-n_p |A| x}}{x!}, \quad x = 0, 1, 2, \dots \quad (3.1)$$

The Poisson point process has only one parameter, the *point intensity* n_p , which is defined to be a global constant. An example of a Poisson point process in the plane is given in Fig. 3.1, left.

It is often useful to examine stochastic point processes in terms of invariance properties. The Poisson point process has many symmetries, that will prove useful for various applications of this point process. First, of all, we have $\Pr(N(A_1) = x) = \Pr(N(A_2) = x)$, for $|A_1| = |A_2|$, by definition. Second, we can define a *translation* transformation ϕ_c of A in S by

$$\phi_c(A) = \{\mathbf{x} + \mathbf{c} | \mathbf{x} \in A\}, \quad A \subset S, \quad \phi_c(A) \subset S, \quad (3.2)$$

in which $\mathbf{c} \in \mathbb{R}^d$ is a constant vector. In observing Eq.(3.1), one readily notices that $\Pr(N(A) = x) = \Pr(N(\phi_c(A)) = x)$ holds, because $|A| = |\phi_c(A)|$, by which the Poisson point process is translation invariant, or homogeneous. Next, we consider the *rotation*

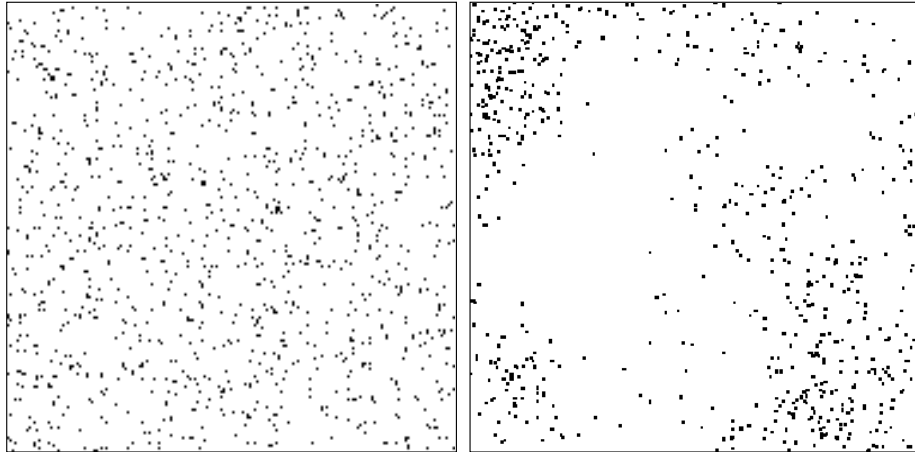


Fig. 3.1: Left: Poisson point process in the plane with constant point intensity n_p ; right: Generalized Poisson point process with variable $n_p(\mathbf{x})$.

transformation defined as

$$\phi_\theta(A) = \{B\mathbf{x} | \mathbf{x} \in A, |B| = 1, B^T B = 1\}, \quad A \subset S, \quad \phi_\theta(A) \subset S, \quad (3.3)$$

where B is an $m \times m$ matrix, and $|B|$ is its determinant. For this rotation operator B , it is true that $|A| = |\phi_\theta(A)|$, and, thus, $\Pr(N(A) = x) = \Pr(N(\phi_\theta(A)) = x)$. Henceforth, the Poisson point process is rotation invariant, or isotropic. If a point process is homogeneous and isotropic, we call the point process *motion-invariant*.

Another very important property of the Poisson point process is its *ergodicity*. In order to explain this property, we must redefine the Poisson point process, and treat it not as a counting process as before, but as a random set of points Θ that is determined by a probability function P . Then, we can state that a stationary point process is ergodic, if P is not a mixture of two distinct stationary distributions, i.e. there do not exist two distributions P' and P'' such that $P = pP' + (1-p)P''$ for any $0 < p < 1$. Thus, P is not decomposable. This ergodicity property enables us to find spatial averages using only individual realizations of the stochastic process, and, as such, it will be very important in our discussions of the average characteristics of random lattices, and the transport quantities computed on them.

We can extend the Poisson point process by relaxing the homogeneity property, and defining the process as

$$\Pr(N(A) = x) = \frac{n_p(A) |A| e^{-n_p(A)|A|x}}{x!}, \quad x = 0, 1, 2, \dots \quad (3.4)$$

where

$$n_p(A) = \int_A n_p(\mathbf{x}) d\mathbf{x}. \quad (3.5)$$

We call this point process the *generalized Poisson point process* with the intensity function $n_p(\mathbf{x})$. An example hereof can be seen in Fig. 3.1, right. The definition of this process in Eq.(3.4) ensures that *locally* the point process behaves as a normal Poisson point process with a *local* constant point intensity $n_p(\mathbf{x})$. This entails that this generalized Poisson point process is an ideal way of representing an inhomogeneous medium distribution, thereby avoiding

unphysical length scales, without throwing away the rotational invariance of the Poisson point process. We will use this generalized Poisson point process in the next chapter as a basis for our new transport method.

In order to be able to perform the transport, we need to create a lattice based on this point distribution, in such a way that we retain the rotational invariance associated with the point process. To our knowledge, the least restrictive, and most general way of doing this is to construct the so-called *Voronoi diagram*, and its dual graph, the *Delaunay triangulation*. We note that other type of graphs are also possible, as long as they are simple graphs, which preserve the causal structure of particle transport.

3.3.2 Voronoi Diagram

Given that a certain point process has generated a point distribution $P = \{p_1, \dots, p_n\} \subset \mathbb{R}^d$, where $2 \leq n < \infty$, and $\mathbf{x}_i \neq \mathbf{x}_j$ for $i \neq j$, $i, j \in I_n$, we can tessellate \mathbb{R}^d , by defining the region

$$V(p_i) = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_i\| \leq \|\mathbf{x} - \mathbf{x}_j\| \forall j \neq i, j \in I_n\}. \quad (3.6)$$

The region $V(p_i)$ is called the *Voronoi* polytope, or cell, associated with p_i , and can be described as the region that consists of all point closer to the point p_i than to all other points. In this thesis, we will always use for the norm $\|\cdot\|$ a ‘Pythagorean’ one, which is a measure for the Euclidean distance between points. This norm is also a freedom we may use, when the space in which the transport takes place has a metric which is not Euclidean.

Given the concept of *dominance regions* to define which part of space is closer to one point p_i than another point p_j

$$H(p_i, p_j) = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_i\| \leq \|\mathbf{x} - \mathbf{x}_j\| \forall j \neq i\}, \quad (3.7)$$

we can use this for an alternative definition of a Voronoi polytope:

$$V(p_i) = \bigcap_{j \in I_n \setminus \{i\}} H(p_i, p_j), \quad (3.8)$$

which clearly indicates that a Voronoi cell of a point p_i is the intersection of all its dominance regions. The boundaries of a Voronoi cell consist of *facets*. In \mathbb{R}^3 , for example, the Voronoi polyhedron, based on the Pythagorean norm $x^2 + y^2 + z^2$, has facets in the form of polygons; the boundaries of that Voronoi facet consists of line segments, half lines, or infinite lines, which are called *Voronoi edges*; the boundaries of a Voronoi edge consists of point, which are called *Voronoi vertices*. The set of all Voronoi cells $\mathcal{V}(P) = \{V(p_1), \dots, V(p_n)\}$ is known as the d -dimensional *Voronoi diagram* generated by P . An example of a Voronoi diagram based on a point process is depicted in Fig. 3.2, middle. For all purists in the field, we note that the Voronoi diagram is not a tessellation in the strict sense of the word, that is commonly reserved for regular tilings. The polytopes that make up the tiling have a number of facets that may change from one Voronoi cell to the next. Every cell in the Voronoi diagram’s dual, the Delaunay triangulation that is discussed in the next subsection, is a so-called *simplex*, which does have a constant number of facets, namely $d + 1$. As such, the Delaunay triangulation *is* a proper tessellation.

3.3.3 Delaunay Triangulation

It is known from graph theory that every planar graph has its dual graph. In the same way, we can construct from the Voronoi diagram its dual graph, the Delaunay tessellation. We

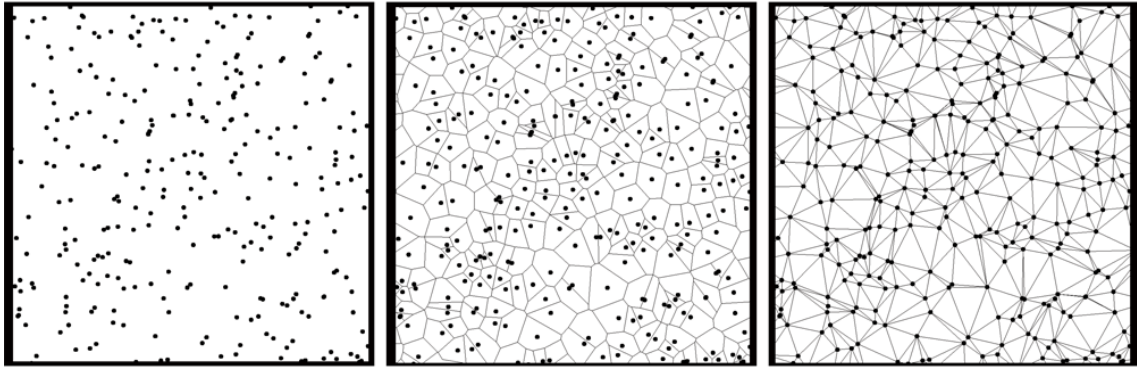


Fig. 3.2: Left: *Poisson point process in the plane*; middle: *the same point process, but now with the associated Voronoi diagram added*; right: *same, but now with the Voronoi diagram's dual, the Delaunay triangulation*. We have made use of open boundary conditions.

do this by joining all points, or generators, of the Voronoi diagram whose Voronoi polytopes share a common $(d - 1)$ -facet. The result is a *triangulation* of space, which consists of simplices². An example of a Delaunay triangulation in the plane is depicted in Fig. 3.2, right. It is always possible that some, often artificial, point distribution results in a tessellation in which one or more of the Delaunay cells have more than d facets. In general, this occurs, when four or more points lie on the same circle. The resultant Delaunay cell is generally termed *degenerate*, and the Delaunay tessellation can be converted into a triangulation by using the degenerate cell's diagonals to convert them into simplices.

There are several interesting properties of a Delaunay triangulation. First, by definition, an edge of a Delaunay simplex joins two points, say p_i and p_j , whose Voronoi cells are neighbors. This Delaunay edge crosses the Voronoi facet which is defined as the half-space $H(p_i, p_j) = H(p_j, p_i)$. Thus, we can conclude that every Voronoi facet intersects its associated Delaunay edge perpendicularly, and exactly halfway between points. Note that about $1/7$ of all Delaunay simplices has a circumcenter that lies outside of the simplex. In that case, the Delaunay edge only intersects the half-space, and not the Voronoi facet itself. Second, one can prove that the circumcircles³ of all Delaunay triangles are empty, cf. Fig. 3.3. Moreover, the circumcenter is a Voronoi vertex. This is all trivially extendible to \mathbb{R}^d . This property can also be used the other way around, as an alternative definition for the Delaunay triangulation: the triangles of some given triangulation of a point set P only satisfy the empty circumcircle criterion, if the triangulation is the Delaunay triangulation. Also, it can be shown that the

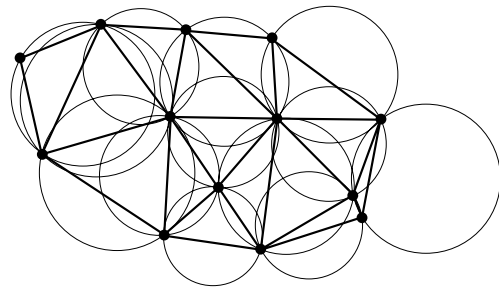


Fig. 3.3: *A Delaunay triangulation, and the circumcircle for each Delaunay triangulation*. For a Delaunay triangulation, it is true that every circumcircle is empty.

²A *simplex* is the generalization of a triangle in \mathbb{R}^d , i.e. a triangle in \mathbb{R}^2 and a tetrahedron in \mathbb{R}^3 .

³A *circumcircle* is a triangle's circumscribed circle, i.e. the unique circle that passes through each of the three triangle's vertices.

Delaunay triangulation in 2D satisfies the *local max-min angle theorem*, which states that the Delaunay triangulation is such that the minimum (acute) angle of each Delaunay triangle is as large as possible. Alternatively, the Delaunay triangulation tries to minimize deviations from equilateral simplices. Moreover, the Delaunay triangulation in the plane is known to minimize the largest circumcircle, and to minimize the largest minimum containment circle, which is the smallest circle containing the triangle. In 3D, the Delaunay triangulation is only known to minimize the largest containment sphere (Rajan 1991; Bern & Eppstein 1992).

These properties all relate to the Delaunay triangulation as a geometrical entity. Alternatively, we can take the point of view that the Delaunay tessellation is a graph of vertices and links between them. The Delaunay graph also has several remarkable properties, which can be studied within the context of *graph theory*. Most notably, it can be shown that the (Euclidean) *minimal spanning tree*⁴ of a set of points in the plane is a subset of the edges in every Delaunay triangulation of the points.

3.3.4 Poisson Delaunay

In what was said above, we have defined the Voronoi diagram and its dual the Delaunay triangulation, which are both constructed on the basis of a point process. What kind of process is used to construct the point distribution has been left unspecified, except for some secondary mathematical restrictions, such as that several points can not lie on one line, or on a circle. If we want to obtain more properties of the tessellations, we need to specify which point process we use.

We started this section by noting that for our new transport method we will use the versatile (generalized) Poisson point process as the generator for our point distribution. The resultant *Poisson Voronoi diagram*, and its dual, the *Poisson Delaunay triangulation*, are the most commonly encountered random tessellations, and it is this Poisson Delaunay graph that we will use as the grid in our new method. Henceforth, when we use *random lattice* in the rest of this thesis, we always mean to denote the Delaunay lattices constructed from a generalized Poisson point process, using the Pythagorean norm, unless explicitly stated otherwise. In the literature, random lattices may also refer to Poisson Voronoi diagrams, or other grids based on stochastic point processes.

The Poisson Voronoi and Poisson Delaunay lattice are ideal from a symmetry point of view. Because the construction recipe Eq.(3.6) only contains distances, the rotational and translational invariance of the Poisson point process is retained. As was already mentioned, there is still a spurious invariant, namely the width of a typical Voronoi cell, or the length of typical Delaunay edge, which both scale with $n_p^{-1/d}$. This can be resolved, by using a generalized Poisson point process with a variable $n_p(\mathbf{x})$ (cf. Chapter 4).

Fortunately, it turns out that it is feasible for these Poisson Voronoi diagrams and Poisson Delaunay triangulations to exactly derive moments and other distributional properties of the geometrical properties of these graphs. We can derive moment measures by using the *ergodic theorem* (Miles 1970, 1974), or using the associated *Palm distribution* (Møller 1989, 1994). The ergodicity of the Poisson point process (cf. Sect. 3.3.1) implies that averages along an individual realization of, for example, a Poisson Delaunay tessellation converge to the theoretical mean, which is, in principle, the average of an infinite number of realizations. Thus, it is meaningful to take averages over only N Voronoi cells in one realization. To

⁴An Euclidean minimal spanning tree connects a set of dots using lines such that the total length of all the lines is minimized, and any point can be reached from any other by following the lines.

accomplish this, one needs to make precise the notion of a ‘typical’ point, for which the characteristics can be evaluated using the Palm distribution theory. In intuitive terms, the Palm distribution probabilities are the conditional probabilities of point process events, given that a point (the typical point) is observed at a specific location. In what follows, we shall only point out those characteristics of the typical Delaunay cell that shall prove useful as background material for our new transport method.

Volume

Given the Poisson point intensity n_p , one can show that the volume of a typical Poisson Voronoi cell in \mathbb{R}^d is

$$E(V_{\text{vor}}) = n_p^{-1}, \quad (3.9)$$

which might seem somewhat trivial, but the similar expectation value for the volume of the average Delaunay cell is

$$E(V_{\text{del}}) = \zeta_V(d)n_p^{-1}, \quad (3.10)$$

in which $\zeta_V(d)$ is now a geometrical constant, that can be different for every dimension d . For \mathbb{R}^2 and \mathbb{R}^3 , this constant is known (Okabe et al. 1999), and is $\zeta_V(2) = 1/2$ and $\zeta_V(3) = 35/24\pi^2$, respectively. This is surely not very trivial.

Number of $(d - 1)$ -facets

Of much importance for our method is the expected number of $(d - 1)$ -facets of a Poisson Voronoi cell. Given the Voronoi-Delaunay duality, every $(d - 1)$ -facet corresponds to one Delaunay edge, and, thus, this expectation value corresponds to the expectation value for the number of edges at a typical Delaunay vertex (which is, of course, just a point out of the Poisson point distribution). This expectation value differs for each dimension, and is known for \mathbb{R}^2 and \mathbb{R}^3 (Okabe et al. 1999),

$$E(E)_{2D} = 6 \quad (3.11)$$

$$E(E)_{3D} = \frac{48\pi^2}{35} + 2. \quad (3.12)$$

This entails that the typical 2D Voronoi cell is a regular hexagon, which is why the 2D Voronoi diagram resembles a honeycomb-like structure. In 3D, however, the typical Voronoi cell has 15.535 faces. The fact that the value Eq.(3.11) is a perfect integer in 2D, but a real number in 3D, is a consequence of the fact that there do exist regular polygons that tessellate the plane that are isotropic enough, to retain enough rotational symmetry, compared to there being no Platonic solid that satisfies that criterion. We will return to this subject in Chapter 5, when we discuss this in the setting of Lattice Gas and Lattice Boltzmann solvers. Even more interesting is the fact that $E(E)_{2D} = 6$, even when the point distribution is *not* Poissonian. In 3D, however, the $E(E)_{3D}$ varies for different point distribution. So, not only is the expectation value of the number of edges not an integer in 2D, it is also no longer a constant!

As an alternative point of view, we can use Eq.(3.11) as an estimate for the (solid) angle subtended by each Delaunay line emanating from a typical vertex:

$$E(\Omega)_{2D} = \frac{\pi}{3} \quad (3.13)$$

$$E(\Omega)_{3D} = \frac{140\pi}{48\pi^2 + 70} \approx \frac{\pi}{3.88}. \quad (3.14)$$

This will be important in estimating how accurate the angular sampling of our transport method is.

Edge Lengths

One of the most important features of our transport method is, that we will use the lines of our random lattices, that is the edges of the Delaunay triangulations, as the channels along which transport (e.g. of photons) occurs. As such, it of the utmost importance to derive not only the expectation value of the Delaunay edge length, but also the distributional properties thereof.

The major results of the distributional properties of typical Poisson Delaunay cells have been derived using Miles' formula (Miles 1970, 1974). Let c and r be the circumcenter and circumradius, respectively, of a $(d+1)$ -dimensional Delaunay simplex in \mathbb{R}^d . Then the $(d+1)$ vertices of the cell can be defined as $\{c + r\mathbf{u}_i\}$, in which $\{\mathbf{u}_i\}$ are unit vectors for $i = 0, \dots, d$. The ergodic probability distribution for a typical Delaunay cell is completely specified by the joint probability density function (Miles' formula)

$$f(r, \mathbf{u}_0, \dots, \mathbf{u}_d) = a(n_p, d) \Delta_d r^{d^2-1} e^{-n_p \omega_d r^d}, \quad (3.15)$$

in which Δ_d is the volume of the $(d+1)$ -simplex, $\omega_d = \pi^{d/2}/\Gamma(d/2+1)$ is the volume of the unit ball, and

$$a(n_p, d) = \frac{\pi^{(d^2+1)/2} \Gamma(d^2/2) \{2n_p \Gamma[(d+1)/2]\}^d}{d^{d-2} \Gamma(d/2)^{2d+1} \Gamma[(d^2+1)/2]}. \quad (3.16)$$

From Eq.(3.15), one can derive the probability distributions $f_L(l)$ for the edge lengths. Knowledge of the exact form of this function is limited to $d = 2$ and $d = 3$. For \mathbb{R}^2 , Muche (1996) derived

$$f_L(l) = \frac{n_p \pi l}{3} \left(n_p^{1/2} l e^{-n_p \pi l^2/4} + \frac{2}{\sqrt{\pi}} \int_{(n_p \pi)^{1/2} l/2}^{\infty} e^{-x^2} dx \right), \quad l \geq 0. \quad (3.17)$$

This distribution function is shown for $n_p = 1$ in Fig. 3.4, left. Visual inspection readily shows that the variance is of the order of the mean. The distribution in Eq.(3.17) has k -th order moments

$$E_{2D}(L^k) = \frac{2^{k+1}(k+1)(k+3)\Gamma(\frac{k+1}{2})}{3(k+2)\pi^{k+1/2} n_p^{k/2}}. \quad (3.18)$$

Thus, the expectation value for the edge length of a typical Delaunay cell is

$$E_{2D}(L) = \frac{32}{9\pi} n_p^{-1/2} \approx 1.132 n_p^{-1/2}. \quad (3.19)$$

An important fact is that Mecke & Muche (1995) showed that the length of an edge emanating from a typical Delaunay vertex has the same distribution as L . This is a subtle and important theorem, because for our transport method we need analysis of the Delaunay lines as sole entities, that as a group make up our lattice, and not as part of one single typical Delaunay simplex.

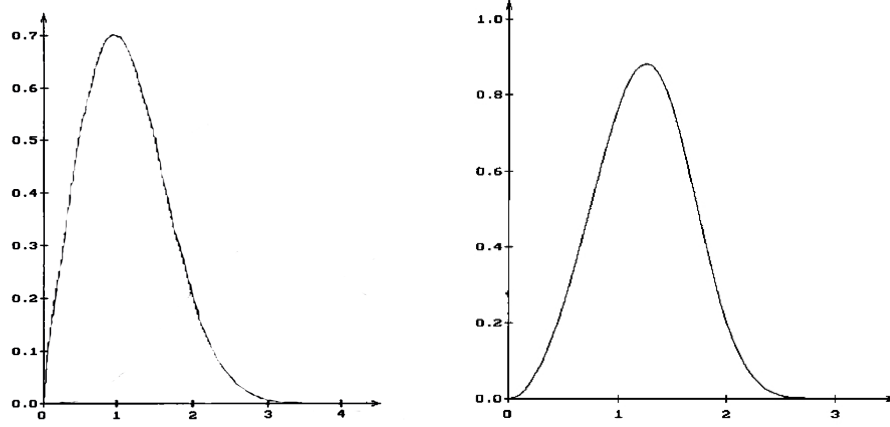


Fig. 3.4: Probability distribution function of the length L of a typical edge of a two-dimensional (left) and three-dimensional (right) Poisson Delaunay tessellation ($n_p = 1$). Source: Muche (1996).

For \mathbb{R}^3 , Muche (1998) derived

$$f_L(l) = \frac{35n_p\pi l^2}{2^8} \left[\left(\frac{\pi n_p l^3}{4} + \frac{5}{2} \right) e^{-\pi n_p l^3/6} - \frac{\pi n_p l^2}{3} \int_{1/2}^{\infty} \left(1 + \frac{\pi n_p l^2 x}{6} \right) e^{-4\pi n_p x^3/3} dx \right], \quad l \geq 0. \quad (3.20)$$

This distribution function is plotted for $n_p = 1$ in Fig. 3.4, right. Eq.(3.20) has k -th order moments

$$E_{3D}(L^k) = \frac{35}{32} \frac{(k+8)(k+6)}{(k+7)(k+5)(k+3)} \Gamma\left(3 + \frac{k}{3}\right) \left(\frac{6}{\pi n_p}\right)^{k/3}. \quad (3.21)$$

The expectation value for a typical Delaunay edge in \mathbb{R}^3 is therefore

$$E_{3D}(L) = \frac{1715}{2304} \left(\frac{3}{4\pi}\right)^{1/3} \Gamma(1/3) n_p^{-1/3} \approx 1.237 n_p^{-1/3}. \quad (3.22)$$

For \mathbb{R}^4 , which is of interest when solving transport problems in space-time, the probability distribution function for the typical edge length is not known, but Christ et al. (1982) give an expression for its expectation value:

$$E_{4D}(L) = \left(\frac{2}{\pi^6}\right)^{1/4} \left(\frac{27}{77}\right) \left(\frac{16!!}{15!!}\right) \Gamma(1/4) n_p^{-1/4} \approx 1.3825 n_p^{-1/4}, \quad (3.23)$$

in which we have used the double factorial, commonly defined as

$$n!! \equiv \begin{cases} n \cdot (n-2) \dots 5 \cdot 3 \cdot 1 & n > 0 \text{ odd} \\ n \cdot (n-2) \dots 6 \cdot 4 \cdot 2 & n > 0 \text{ even} \\ 1 & n = -1, 0. \end{cases} \quad (3.24)$$

Although the probability distribution function for the length of a typical edge can in principle be derived in \mathbb{R}^d from Miles' formula Eq.(3.15), they are only known for \mathbb{R}^2 and \mathbb{R}^3 . What is known, is that in \mathbb{R}^d , the k -th order moments of this probability distribution function will be of the form

$$E_{\mathbb{R}^d}(L^k) = \zeta_L(k, d)n_p^{-k/d}, \quad (3.25)$$

in which $\zeta_L(k, d)$ is a certain geometrical constant for a given k and d that can be derived exactly, once the probability distribution function is given.

We conclude by noting that, although the edge length distribution is well-behaved for a Poisson point distribution, this distribution may be much more erratic for very inhomogeneous point distributions. This matter has been studied thoroughly in Schaap (2006).

Shortest Path

Somewhat related to the distribution of edge lengths is the length of the shortest path over the Poisson Delaunay graph, from one vertex s to another t . Ideally, this length should equal the Euclidean distance $\|s - t\|$, but the associated straight line connecting these two points does not coincide with the lines of the graph. One can derive expressions for the deviation of this path from the Euclidean distance in \mathbb{R}^d , but here we shall only give the results for planar graphs.

Since each realization of the Poisson point process is locally finite, we know that there exists a shortest path $p(s, t)$ between any two points s and t . Keil & Gutwin (1992) showed that, for an arbitrary vertex set, the ratio of the Delaunay distance $|p(s, t)|$, and the Euclidean distance does not exceed $2\pi/(3\cos(\pi/6)) \approx 2.42$. Earlier on, Chew (1986) had already shown that this ratio can be arbitrarily close to $\pi/2 \approx 1.57$, and it is still a standing conjecture that $\pi/2$ is the worst case. For more 'regular' Delaunay graphs, this Delaunay distance can be much less. Indeed, for the Poisson Delaunay graph, Baccelli et al. (1998) showed that this ratio is asymptotically less than $4/\pi \approx 1.27$. Using simulations, however, Baccelli et al. (1998) showed that this ratio gets a value of approximately 1.05, which is very close to unity.

We refrained here from specifying exactly the recipe for choosing the shortest path along a graph. We will describe one in the next chapter, when we implement the ballistic transport of particles. What does matter here, is that one can evaluate for a random lattice what the Delaunay-Euclidean distance ratio is, and that one can even incorporate this ratio to rescale the lengths along the graph in order to match the Delaunay distance to the Euclidean one. *This will ensure that the expected transport velocities are recovered.*

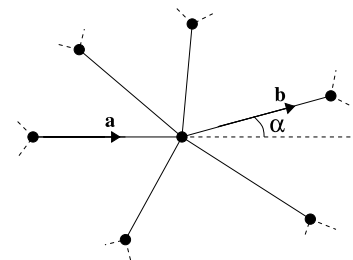


Fig. 3.5: A typical Delaunay vertex, and the definition of the deflection angle α .

Deflection Angle

One very important characteristic of the Poisson Delaunay lattice that we need for our transport method is the *deflection angle* when moving from one edge to the next. Let us be a bit more specific: take a typical vertex of our Delaunay graph (cf. Fig. 3.5, for an example in the plane), at which a certain number of Delaunay lines meet (on average six, in the plane). Then we can consider for each of these lines, what its most 'straightforward' continuation

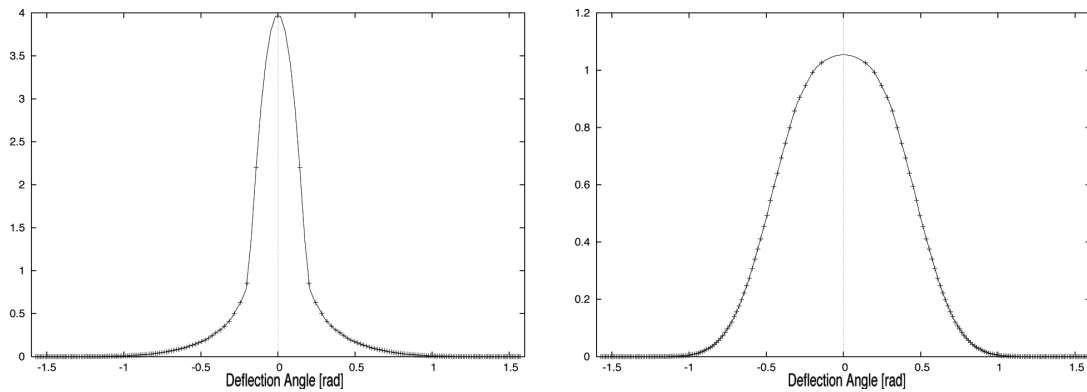


Fig. 3.6: Normalized histograms of the deflection angle α in radians for a Delaunay triangulation of a Poisson point distribution of a million points in \mathbb{R}^2 (left), and \mathbb{R}^3 (right).

is. That is, we determine which of the outgoing lines minimizes the deflection angle α , calculated using $\mathbf{a} \cdot \mathbf{b} = ab \cos \alpha$, in which \mathbf{a} and \mathbf{b} are the vectors of the two lines (note that this recipe is valid in general d -dimensional space, for $d \geq 2$). This deflection angle is important, because in transport methods we want to minimize any artificial scattering, and it is therefore of the utmost importance to know the distributional properties of this angle, in order to evaluate its moments.

Unfortunately, the probability distribution function for α is not known. This is probably because in finding this distribution, one cannot just consider one Delaunay simplex (for the incoming line), but also another (for the outgoing) must be involved (Chiu 2004). Therefore, Miles' formula Eq.(3.15), and the machinery of subsequent probabilistic analysis, cannot be used. On the other hand, we *can* make some educated guesses as to the expectation values and forms of these distribution functions. Moreover, we can do Monte Carlo experiments to obtain an indication as to the shape of these functions. To accomplish this, we constructed the Delaunay triangulation of a Poisson point process of a million points in \mathbb{R}^2 and \mathbb{R}^3 . Subsequently, we calculated the deflection angle α for each pair of incoming and most straightforward lines. From this, we constructed normalized histograms, which are plotted in Fig. 3.6. Note that the x-axes have the same scale, by which the widths of the function can be directly compared.

As was to be expected, both histograms are symmetric around $\alpha = 0$, which is a direct consequence of the fact that a typical Voronoi cell is, on average, (axi-)symmetric around each incoming Delaunay line. Except for this similarity, the histograms differ from each other in the sense that the distribution function for the deflection angle in \mathbb{R}^2 is much more strongly peaked than its equivalent in \mathbb{R}^3 . This is probably due to the fact that the average Poisson Voronoi cell in the plane is a perfect hexagon, whilst the typical Voronoi cell in 3D is a 15.535-hedron (cf. Sect. 3.3.4). Thus, the typical Poisson Voronoi cell in \mathbb{R}^2 ensures a most straightforward path that lies exactly along the incoming one, while in \mathbb{R}^3 the typical Voronoi cell is a strange many-sided polyhedron that does not have a most straightforward path that is exactly along the incoming one. Therefore, the distribution function appears to be wider.

Thus, we can conclude this section by noting that, although the exact form of the distribution function for the deflection angle is not known, we do know that it is symmetric around $\alpha = 0$, which will turn out to be important for our transport method, in which we want

to conserve momentum. Apart from the characteristics of the random lattices mentioned in this section, there are many others that will not be used in the analysis of our transport method. We will refrain from listing them here, and refer the interested reader to the very complete review Okabe et al. (1999).

3.4

Construction Algorithms

Because of the optimal properties of random lattices, from both a physical and mathematical point of view (not to mention the diversity in other uses, briefly covered in the next section), there has been considerable interest from the field of *computational geometry* in designing algorithms that optimize the operation count associated with the construction of the Delaunay triangulation from a given point set. This section briefly reviews the random lattice construction algorithm *Quickhull* we use as a plugin for our transport code.

Because of the dimensionality of observable space, most algorithms developed within the computational geometry community focus on two- and three-dimensional space. Fortunately, most of the algorithms are trivially (mathematically, not necessarily computationally) extended to \mathbb{R}^d . Given an initial distribution of N points, it can be shown that optimal algorithms triangulate \mathbb{R}^d in $O(N \log N)$ time for $d \leq 3$ (Yao 1981), and $O(N^{\lfloor d/2 \rfloor} / \lfloor d/2 \rfloor!)$ otherwise. This expected operation count depends strongly on the use of the clever implementation of sorting algorithms, and the use of special data structures. Still, this is the worst case scenario; randomized point sets can lower the operation count to, for example, linear time in 2D. A detailed comparison between the most well-known Delaunay triangulation algorithms can be found in Bern & Eppstein (1992), Fortune (1992), and Su & Drysdale (1995).

3.4.1 Quickhull

There are several reasons why we chose, from that whole arsenal of triangulation algorithms, the Quickhull algorithm (Barber et al. 1995). First, it has been shown to perform in optimal time, even in \mathbb{R}^3 and above. Second, many computational geometry algorithms traditionally assume that the point set is well behaved; when an algorithm is implemented, however, floating-point arithmetic is prone to roundoff errors, that can lead to erratic behavior of the resultant triangulation. This is very important for our transport method, because, for realistic cosmological simulations, billions of particles are needed (cf. Chapter 10). Inside potential wells, these particles can get so close that their coordinates, expressed in floating point numbers, can start to overlap. Quickhull was one of the first algorithms that solve this problem, and ensures a stable triangulation, even in these extreme cases. Finally, because we did not want to invent the wheel ourselves, we compared several open-source packages, available online, that we could use as a modular plugin routine. Quickhull has a very solid, fast and flexible C implementation, `QHull`⁵, developed by Brad Barber himself, that can compute the Delaunay triangulation in any space up to \mathbb{R}^8 , while in the meantime also computing volumes and surface areas. It also handles roundoff errors from floating point operations. Moreover, except for the usual options for outputting the Delaunay and Voronoi

⁵<http://www.qhull.org>

graph connectivity schemes, it also has the option to output VTK⁶ and GeomView⁷ files that can be used directly to give a 3D visualization of the tessellation.

The Algorithm

In order to explain how the algorithm works, we commence by introducing the concept of a *convex polytope*. This can be defined in two ways: algebraically, a convex polytope in \mathbb{R}^d can be defined as the set of solutions to a system of inequalities

$$m\mathbf{x} \leq \mathbf{b}, \quad (3.26)$$

in which m is a real $s \times d$ matrix, and \mathbf{b} is a real s -vector. A stricter definition is often given, in which the polytope is required to be bounded. Geometrically, a convex polytope can be defined as a polytope for which a line connecting any two (noncoplanar) points on the surface always lies in the interior of the polyhedron. As an example in \mathbb{R}^3 , there are 92 convex polyhedra having only regular polygons as faces; these are called *Johnson solids*, and they include the Platonic solids, but also the *Archimedean solids*⁸. Moreover, convex polyhedra have the nice property that, given the number of vertices V , edges E , and faces F ,

$$V + F - E = 2. \quad (3.27)$$

A more general concept is that of a *convex hull*. Loosely speaking, the convex hull of a set of points is the smallest convex set that contains the points. Given a set of points S , with N points p_1, \dots, p_N in \mathbb{R}^d , the convex hull C is mathematically defined as

$$C \equiv \left\{ \sum_{j=1}^N \lambda_j p_j : \lambda_j \geq 0 \forall j \wedge \sum_{j=1}^N \lambda_j = 1 \right\}. \quad (3.28)$$

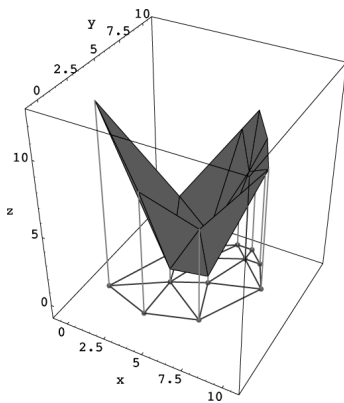


Fig. 3.7: Points are lifted to a paraboloid of revolution. The downward projected convex hull of those points is the sought after Delaunay triangulation.

A convex polytope is obtained, when computing the convex hull of a set of points. A Voronoi cell is a convex polyhedron, by its definition (compare Eq.(3.26) with Eq.(3.8)). One can show that the construction of halfspace intersections, Voronoi diagrams, and Delaunay triangulations can be reduced to the convex hull problem. The remarkable connection between the convex hull and the Delaunay triangulation was discovered by Brown (1979). He proved that we can construct the Delaunay triangulation from a set of points P in \mathbb{R}^d by lifting them to a paraboloid in \mathbb{R}^{d+1} , computing the convex hull, and projecting that down. See, for example, Fig. 3.7. Considering \mathbb{R}^2 , we lift the point $p = (x, y)$ in the plane to the point $l(p) = (X, Y, Z)$, in which $X = x$, $Y = y$, and $Z = x^2 + y^2$. Brown (1979) proved that the projection of the part of the convex hull of

⁶<http://public.kitware.com/VTK/>

⁷<http://www.geomview.org/>

⁸The 13 Archimedean solids are the convex polyhedra that have a similar arrangement of nonintersecting regular convex polygons of two or more different types arranged in the same way about each vertex with all sides the same length (Cromwell 1997).

the lifted set $l(P)$, consisting of the downward-facing faces, is the Delaunay triangulation of P . The Voronoi diagram can be obtained in a similar fashion, but now considering the planes tangent to the paraboloid. The Quickhull algorithm performs its calculations in the space of points and convex hulls, constructing the Voronoi diagram and/or the Delaunay triangulation from that. Points are added incrementally in a random order as a variant to the original randomized incremental algorithm proposed by Clarkson & Shor (1989).

We end this section by noting that there are other open-source tessellation packages (e.g. the one described in van de Weygaert 1994, or the CGAL package⁹), some of which are implementations of algorithms that not only add their points incrementally, but also store a hierarchical tree of these operations. This ensures, that if one point is deleted one can go one step back in the tree, without the need to compute the whole tessellation again. This could prove handy when wanting to incorporate dynamic meshes, in which the movement of the point can be implemented as a deletion and creation operation. Another use for this property, which we leave for future consideration, is the construction of Delaunay triangulations for radiative transfer through non-grey atmospheres.

3.5

Applications

Now that we have given an introduction to the topic of random lattices, and some of their construction algorithms, we conclude this chapter by giving just a small glimpse of their rich history and wide variety of applications.

Many natural structures on Earth and beyond resemble the structure of Voronoi diagrams. Indeed, it so seems that, as soon as something is not continuous, the Voronoi cell is the always recurring basic structure. Several examples are depicted in Fig. 3.9, ranging from the smallest to the largest scales. We see soap bubbles forming a Voronoi-like structure which, as can be shown, minimizes the foam's free energy that is associated with its surface tension (Cox & Graner 2004). We see geological plates in the Icelandic landscape, next to the all too familiar honeycomb, and in all we recognize the Voronoi diagram. Last, but not least, we see a simulation output of the large scale structure, the distribution of galaxies in our Universe, on scales of tens of megaparsec, which can be shown to kinematically evolve towards the Voronoi diagram in the limit (van de Weygaert 2002; Sheth & van de Weygaert 2004). This is just a small sample of what one would find, when searching for Voronoi-like structures in Nature.

As such, it would not seem all too surprising, should we make an archeological find that shows some ancient culture being aware of this omnipresent pattern. The earliest published work depicting Voronoi-like diagrams is by Descartes (1644), who used this structure in an explanation of the disposition of matter in the Solar system (cf. Fig. 3.8). As Okabe et al. (1999) aptly remark, the fact that the construction of this diagram is explained nowhere

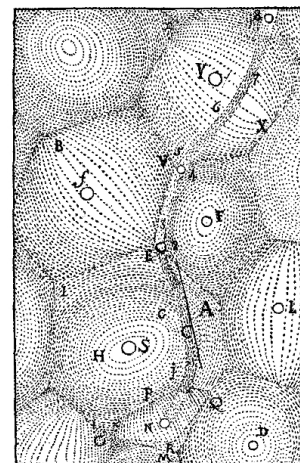


Fig. 3.8: Descartes claims that the Solar system consist of vortices of matter revolving around fixed stars, while comets follow an erratic path in between.

⁹<http://www.cgal.org>

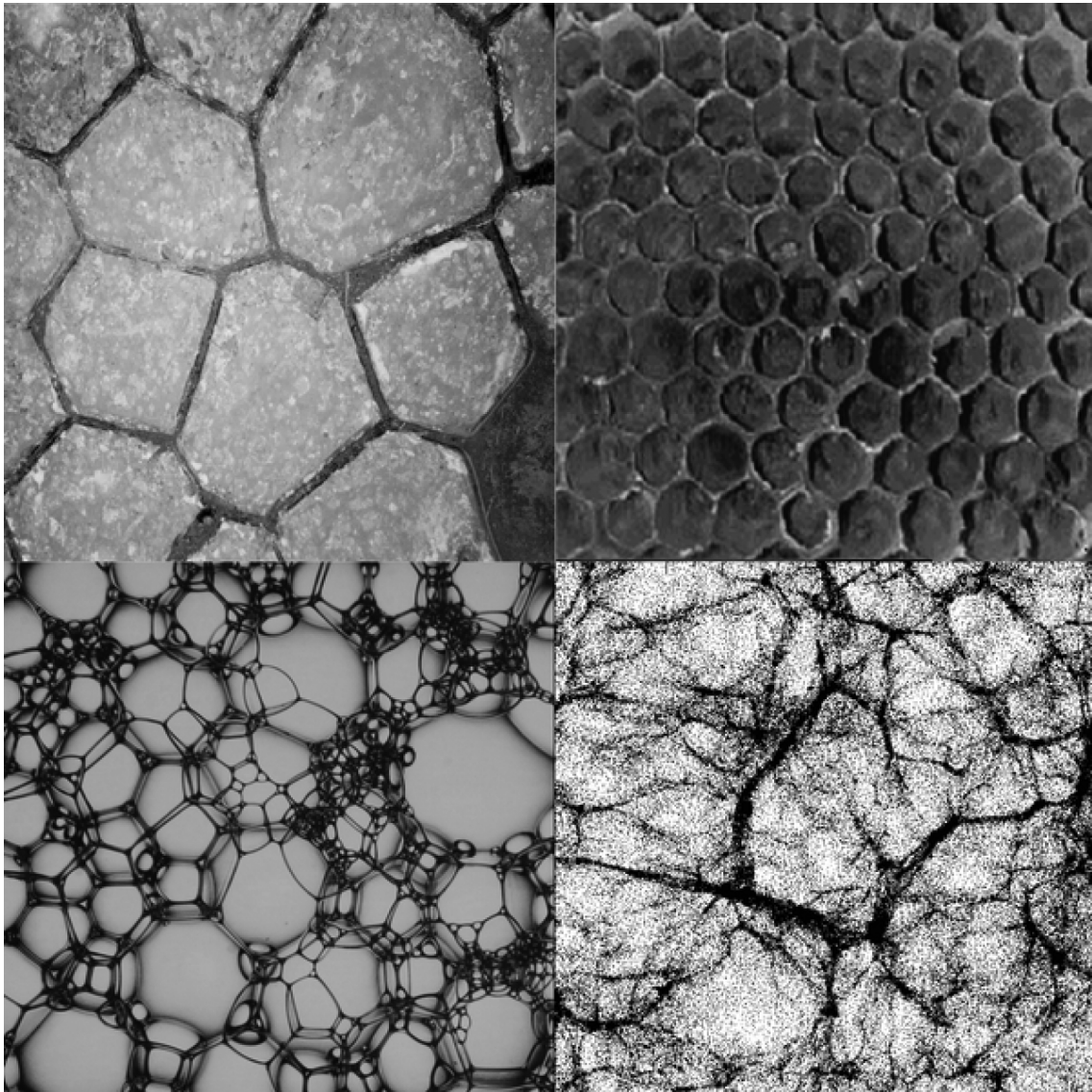


Fig. 3.9: Several examples of Voronoi-like structure in Nature. Top: Geological plates in the Icelandic landscape, and the familiar honeycomb; bottom: soap bubbles in equilibrium having a variant of the Voronoi foam structure (Johnson-Mehl), next to a simulation of the large scale structure of the Universe.

in the book might indicate that their use was quite common, even in that age. Peter Gustav Lejeune Dirichlet (1805-1859) and Georgy Fedoseevich Voronoy (1868-1908) were to first to systematically describe Voronoi diagrams in their studies on positive quadratic terms (Dirichlet 1850; Voronoi 1908). It was the Russian number theorist Boris Nikolaevitch Delone (1890-1980) who associated the names of Dirichlet ("Dirichlet domain") and Voronoy ("Voronoi cell") to such structures. The dual concept of the Delaunay tessellation, that we use as a random lattice, was already described by Voronoy (Voronoi 1908), but it was Delone, using the French version of his name, Delaunay, who first described the structure in terms of the empty sphere method (Delaunay 1934).

It would make a very interesting research project in both sociology and communication

theory, to study the history of the reinvention of the Voronoi diagram and its dual. Throughout the previous century, both graphs have been reinvented many times within many different branches of science, and even art. The lack of communication between disciplines led to the multiple reinvention of the Voronoi diagram. In physics and geography, for example, the Voronoi cells are known as Thiessen-polygons, and Wigner-Seitz cells, respectively, after the 'inventors' in both fields. This miscommunication has led to much redundant work, in which different authors reinvent the wheel with respect to, for example, the average properties of the Delaunay graph. Fact is, that scientists from every field that needs to describe something that is discretized in space and/or time eventually stumble upon the optimal way to divide space, one way or another. One would believe that nowadays, with the advent of mass communication channels, such as the Internet, and the very good sales numbers of standard works such as Okabe et al. (1999), this lack of communication would diminish. Unfortunately, this still seems not to be the case. Therefore, we end this chapter by giving a small, albeit incomplete, overview of the modern-day applications of random lattices in the sciences.

- *Anthropology and archaeology*: Identify the parts of a region under the influence of different neolithic clans, chiefdoms, ceremonial centers, or hill forts.
- *Biology, ecology, forestry*: Model and analyze plant competition.
- *Cartography*: Piece together satellite photographs into large "mosaic" maps.
- *Crystallography and chemistry*: Study chemical properties of metallic sodium ("Wigner-Seitz regions"); modeling alloy structures as sphere packings ("Domain of an atom").
- *Finite element analysis*: Generating finite element meshes which avoid small angles.
- *Geography*: Analyzing patterns of urban settlements.
- *Geology*: Estimation of ore reserves in a deposit using information obtained from bore holes; modeling crack patterns in basalt due to contraction on cooling.
- *Geometric modeling*: Finding well-behaved triangulations of 3D surfaces.
- *Marketing*: Model market of US metropolitan areas; market area extending down to individual retail stores.
- *Mathematics*: Study of positive definite quadratic forms ("Dirichlet tessellation" and "Voronoi diagram").
- *Metallurgy*: Modeling grain growth in metal films.
- *Meteorology*: Estimate regional rainfall averages, given data at discrete rain gauges ("Thiessen polygons").
- *Pattern recognition*: Find simple descriptors for shapes that extract 1D characterizations from 2D shapes ("Medial axis" or "skeleton" of a contour).
- *Physiology*: Analysis of capillary distribution in cross-sections of muscle tissue to compute oxygen transport ("capillary domains").
- *Physics*: Regge calculus; simplicial SUSY; sphere packings.
- *Robotics*: Path planning in the presence of obstacles.
- *Statistics and data analysis*: Analyze statistical clustering ("Natural neighbors" interpolation).
- *Telecommunications*: Determine which antenna is closest to a mobile phone.
- *Zoology*: Model and analyze the territories of animals.

In astrophysics, the advantageous properties of adaptive random lattices have been used extensively to study and sample continuous fields. Application range from astrophysical hydrodynamics and N-body simulation (e.g. Gnedin 1995; Whitehurst 1995; Pen 1998; Pelupessy et al. 2003) to field reconstruction techniques (Schaap & van de Weygaert 2000;

Schaap 2006).

3.6

Conclusions

Examples abound: Nature has known all along that the Voronoi diagram is the optimal way to partition space; what can we do but follow? In this chapter, we have introduced the concept of a random lattice, which is a graph based on a (generalized) Poisson point process. This Delaunay graph retains the motion-invariant and ergodic properties of the Poisson point process, and many of its mathematical and geometrical properties are known analytically. Its uses are abundant, covering a wide range of sciences. It is this lattice that we will use as a basis for our transport method, as we shall explain in the next chapter.

Bibliography

- Ambjørn, J., Jurkiewicz, J., & Loll, R. 2006, *Contemporary Physics*, 47, 103
- Baccelli, F., Tchoumatchenko, K., & Zuyev, S. 1998, *Markov Paths on the Poisson-Delaunay Graph*, Tech. Rep. 3420, INRIA
- Barber, B. C., Dobkin, D. P., & Huhdanpaa, H. T. 1995, *ACM Trans. on Mathematical Software*, 22, 469
- Berger, M. J. & Colella, P. 1989, *J. Comput. Phys.*, 82, 64
- Berger, M. J. & Olinger, J. 1984, *J. Comput. Phys.*, 53, 484
- Bern, M. & Eppstein, D. 1992, *Mesh Generation and Optimal Triangulation, Computing in Euclidean Geometry* (World Scientific)
- Bird, G. 1994, *Molecular Gas Dynamics and the Direct Simulation of Gas Flows* (Clarendon)
- Braun, J. & Sambridge, M. 1995, *Nature*, 376, 655
- Brown, D. 1979, *Inf. Process. Lett.*, 9, 223
- Chew, P. 1986, There is a planar graph almost as good as the complete graph, *Proc. of the 2nd Symposium on Computational Geometry* (New York: Yorktown Heights), 169–177
- Chiu, S. N. 2004, private communication
- Christ, N. H., Friedberg, R., & Lee, T. D. 1982, *Nuclear Physics B*, 202, 89
- Clarkson, K. & Shor, P. 1989, *Discr. Comput. Geom.*, 4, 387
- Cox, S. J. & Graner, F. 2004, *Phys. Rev. E*, 69, 031409
- Cromwell, P. R. 1997, *Polyhedra* (New York: Cambridge University Press), 79–86
- Delaunay, B. 1934, *Classe des Sciences Mathématiques et Naturelles*, 7, 793
- Descartes, R. 1644, *Principia Philosophiae* (Amsterdam: Ludovicus Elzevirius)
- Dirichlet, G. L. 1850, *Journal für die Reine und Angewandte Mathematik*, 40, 209
- Español, P. 1997, *Europhysics Letters*, 39, 605
- Flekkøy, E. G., Coveney, P. V., & de Fabritiis, G. 2000, *Phys. Rev. E*, 62, 2140
- Fortune, S. 1992, *Voronoi Diagrams and Delaunay Triangulations, Computing in Euclidean Geometry* (Singapore: World Scientific), 193–233
- Frisch, U., Hasslacher, B., & Pomeau, Y. 1986, *Phys. Rev. Lett.*, 56, 1505
- Gladrow, D. W. 2000, *Lattice Gas and Lattice Boltzmann Methods* (Springer-Verlag, Berlin)
- Gnedin, N. Y. 1995, *ApJS*, 97, 231
- Icke, V. 1988, *A&A*, 202, 177
- Kaku, M. 1983, *Phys. Rev. Lett.*, 50, 1893
- Keil, M. & Gutwin, C. 1992, *Discrete Comput. Geom.*, 13
- Kunasz, P. B. & Auer, L. 1988, *JQSRT*, 39, 67

- Mecke, J. & Muche, L. 1995, *Mathematische Nachrichten*, 176, 199
- Miles, R. E. 1970, *Izvestilia Akademii Nauk Armlianskoi SSR Matematika*, 5, 263
- . 1974, A Synopsis of 'Poisson flats in Euclidean space', *Stochastic Geometry* (New York: John Wiley), 202–227
- Møller, J. 1989, *Advances in Applied Probability*, 21, 37
- . 1994, *Lecture Notes in Statistics*, Vol. 87, *Lectures on Random Voronoi Tessellations* (New York: Springer-Verlag)
- Muche, L. 1996, *Journal of Statistical Physics*, 84, 147
- . 1998, *Mathematische Nachrichten*, 191, 247
- Noether, E. 1918, *Nachr. d. Königl. Gesellsch. d. Wiss. zu Göttingen, Math-phys. Klasse*, 235
- Okabe, A., Boots, B., Sugihara, K., & Chiu, S. 1999, *Spatial Tessellations, Concepts and Applications of Voronoi Diagrams*, 2nd edn. (John Wiley & Sons)
- Pelupessy, F. I., Schaap, W. E., & van de Weygaert, R. 2003, *A&A*, 403, 389
- Pen, U.-L. 1998, *ApJS*, 115, 19
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, *Numerical Recipes in C: The Art of Scientific Computing* (New York, NY, USA: Cambridge University Press)
- Rajan, V. T. 1991, Optimality of the Delaunay Triangulation in \mathbb{R}^d , *Proc. 7th Annual Symposium on Computation Geometry* (Vancouver: ACM)
- Regge, T. 1961, *Nuovo Cimento A*, 19, 558
- Richtmyer, R. D. & Morton, K. W. 1967, *Difference Methods for Initial Value Problems* (New York: Interscience Publishers)
- Sambridge, M., Braun, J., & McQueen, H. 1995, *Geophysical Journal International*, 122, 837
- Schaap, W. E. 2006, DTFE: The Delaunay Tessellation Field Estimator, PhD thesis, Groningen
- Schaap, W. E. & van de Weygaert, R. 2000, *A&A*, 363, L29
- Sheth, R. K. & van de Weygaert, R. 2004, *MNRAS*, 350, 517
- Steinacker, J., Hackert, R., Steinacker, A., & Bacmann, A. 2002, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 73, 557
- Stoyan, D., Kendall, W., & Mecke, J. 1995, *Stochastic Geometry and its Applications* (Chichester: Wiley)
- Su, P. & Drysdale, L. S. 1995, A Comparison of Sequential Delaunay Triangulation Algorithms, *Proc. 11th Annual Symposium on Computational Geometry* (Vancouver: ACM), 61–70
- Succi, S. 2001, *The Lattice Boltzmann Equation for Fluid Dynamics and Beyond* (Oxford University Press, Oxford)
- Ubertini, S. & Succi, S. 2005, *Progress in Computational Fluid Dynamics*, 5, 85
- Van de Weygaert, R. 1991, Voids and the Geometry of Large Scale Structure, PhD thesis, Leiden
- van de Weygaert, R. 1994, *A&A*, 283, 361
- van de Weygaert, R. 2002, in *ASSL Vol. 276: Modern Theoretical and Observational Cosmology*, ed. M. Plionis & S. Cotsakis, 119–+
- Van Leer, B. 1970, A choice of different schemes for ideal compressible flow, PhD thesis, Leiden
- Voronoi, G. 1908, *Journal für die Reine und Angewandte Mathematik*, 134, 198
- Whitehurst, R. 1995, *MNRAS*, 277, 655
- Wolfram, S. 1986, *J. Stat. Phys.*, 45, 471
- Yao, A. C.-C. 1981, *J. ACM*, 28, 780

Transport on Adaptive Random Lattices

In this chapter, we will discuss the specifics of a new method that is, in principle, able to solve all transport equations that can be described by a general Master Equation. We will focus, however, on linear transport problems of particles moving through a background medium. It resembles Monte Carlo methods in the sense that it samples the free path space of the particles. This sampling, however, is done in a drastically different way: it uses a random lattice, shown to be locally motion-invariant, that adapts to the medium properties in such a way that the lattice itself is a sampling of the path space. The resultant transport process can be implemented as a simple random walk along this adaptive random lattice. Our implementation is shown to be extremely efficient.

Transport on Adaptive Random Lattices
J. Ritzerveld and V. Icke
Physical Review E (2006), **74**, 026704

4.1

Introduction

In the previous two chapters, we have explained the essentials that are necessary for laying out the specifics of a new transport method. This method is an amalgamation of the Monte Carlo stochastic transport solvers, covered in Sect. 2.6, and the use of the versatile random lattices of the previous chapter. In effect, it combines the best of both worlds, adapting the versatility of the Monte Carlo methods to the symmetry preserving properties of the Poisson Delaunay graphs. We do not use a standard Monte Carlo transport method that works on some arbitrary (possibly adaptive) mesh, but take it one step further, and give physical meaning to the mesh itself. *We will choose the properties of the random lattice, and the point process that lies at its basis, in such a way that the lengths of the Delaunay edges correspond to the physical length scales of the transport process.* We accomplish this by letting the point process adapt to the background medium properties. Like we described in Sect. 2.5.2, the transport process then reduces to the random walk of particles along a graph from one interaction event to the next. In this case, the lattice vertices represent the interaction centers, and the lines channel the movement from one center to the next.

The transport process that we wish to solve with our new method can be written down in the general form of the Master Equation, cf. Eq.(2.38),

$$\mathbf{D}f = \mathbf{C}f, \tag{4.1}$$

where \mathbf{D} is a drift and \mathbf{C} is a collision operator, defined on some abstract space, that can take the form of the familiar phase space, but also that of a data space, a money space, or even a gene space. Several of these more unusual applications will be discussed in the next chapter. The method we will describe is general enough to solve almost all Master Equations of type Eq.(4.1), and can be used as such, but in this chapter we will specifically focus on the linear transport processes of particles being transported through a (possibly dynamically evolving) background medium, with which they interact. Nonlinear particle-particle interactions are ignored, except through feedback from the background medium itself. Thus, the transport method described in this chapter can be used to model transport processes such as electron, neutron or photon transport. We refer the reader to Chapter 5 for more diverse applications of this transport method.

We will commence by discussing how to construct the random lattices that have line lengths corresponding to the relevant physical transport length scale in Sect. 4.2, after which we delve deeper into the mathematical and physical properties of the resultant lattice in Sect. 4.3. Thereafter, we are ready to explain in Sect. 4.4 how the transport can be defined, and implemented, on such lattices, giving some examples along the way. We shall explicitly point out in Sect. 4.5 how our transport can be trivially extended into \mathbb{R}^d , and we will end this chapter by making some remarks about the efficiency of the new method in Sect. 4.6.

4.2

Adaptive Random Lattices

The goal of this section is to demonstrate how to construct a random lattice that adapts to the background medium properties, according to a specific criterion. When introducing random lattices in the previous chapter, we already briefly pointed out that, although spurious invariance and symmetry breaking associated with rotational and translational invariance are prevented by the use of random lattices, one drawback still remains: the introduction of an unphysical length scale, determined by the average point to point distance, or the expectation value of the Delaunay line length $\langle L \rangle$. For Delaunay lattices based on a Poisson point process Eq.(3.1), we specifically derived the k -th order expectation value for the line length L as, cf. Eq.(3.25),

$$\langle L^k \rangle = \zeta(k, d) n_p^{-k/d}, \quad (4.2)$$

in which $\zeta(k, d)$ is some geometrical constant for each pair of the value of k and the dimension d . The average Delaunay line length in 3D, for example, is

$$\langle L \rangle = \zeta(1, 3) n_p^{-1/3} \approx 1.237 n_p^{-1/3}. \quad (4.3)$$

Although this line length does not have a delta function as a probability function, as in the case of a regular mesh in which the delta function peaks at a length of one cell width, but a certain spread $\sigma^2 = \langle L^2 \rangle - \langle L \rangle^2 \propto n_p^{-2/3}$, it does still have a global first order expectation value that scales with $n_p^{-1/d}$, which is a constant once the globally fixed Poisson point density n_p has been chosen.

As already discussed in Sect. 3.2, these fixed values cause problems when the medium distribution itself is not homogeneous, as they might underresolve high density regions, and impose constraints on parameters like the Knudsen and Reynolds numbers. More generally

speaking, it causes problems in regions where the mean free path is shorter than this expectation value. Regular meshes have resolved this problem somewhat by introducing AMR, in which the mesh refines itself according to some predefined criterion, often based on the gradient of the pressure or density. Because here we are considering a statistical method, it is justifiably better to choose as an adaptation parameter (a function of) the mass density of the medium. Similar efforts have been made with respect to structured grids, which have one basic congruent cell shape as a basis, in the DSMC community (Garcia et al. 1999; Wu et al. 2002), in which the authors wanted to make sure that the local cell sizes at least resolve the local mean free path, which is the most relevant physical length scale of transport processes. In this case, we do not have a regular mesh as a basis and proceed by trying to define a point process which does refine based on the local mass density.

To accomplish this, we discard the regular Poisson point process Eq.(3.1), which is the usual basis for random lattices. Instead, we define a point distribution, which is a convolution of a homogeneous Poisson point process Φ and a function of the possibly inhomogeneous density distribution $n(\mathbf{x})$, symbolically written as:

$$n_p(\mathbf{x}) = \Phi * f(n(\mathbf{x})). \quad (4.4)$$

The only constraint is the maximum number of points, or resolution, N available for the simulation. Eq.(4.4) is an alternative, more physics oriented, recipe for constructing the generalized Poisson point process Eq.(3.4). It amounts to randomly sampling the function $f(n(\mathbf{x}))$ using a Direct Inversion of Rejection method. If the distribution $n(\mathbf{x})$ is inhomogeneous, one expects the point distribution $n_p(\mathbf{x})$ to be inhomogeneous too, mimicking the medium. But, as long as our number of points N is high enough, we can always zoom in so far that locally the medium distribution is homogeneous, and the point distribution Poissonian. Thus, locally, the point distribution defined by Eq.(4.4) retains the rotational and translational symmetries associated with Poisson point processes.

Until now, we have not specified the exact form of the correlation function $f(x)$ in Eq.(4.4). We will discuss its details in the next subsection. For now, we will give an example of a very inhomogeneous point distribution in Fig. 4.1, left. It was constructed using a fractal point process, which has been shown to mimic the distributional behavior of galaxies and galaxy clusters (cf. Appendix 4.A). The resultant random lattice is shown next to it. It is apparent that the Delaunay line lengths are not of the same order, but cover a wide range of scales.

4.2.1 The correlation function

What do we choose for the correlation function $f(x)$? Because we introduced this function to ensure adaptation of the point distribution to the medium distribution, the obvious conclusion is that we need $f(x)$ to be a monotonically increasing function of x , by which the average point-to-point distances will actually be shorter in denser regions. In light of what we discussed about Monte Carlo methods transporting particles along trajectories which have as an average length one mean free path, we may choose one particular form of $f(x)$ which makes the resultant Delaunay line lengths have a very physical meaning.

From basic transfer theory, we know that the local mean free path relates to the local particle density in the following way (valid for every dimension $d \geq 1$), cf. Eq.(2.12):

$$\lambda^{-1}(\mathbf{x}) = n(\mathbf{x})\sigma, \quad (4.5)$$

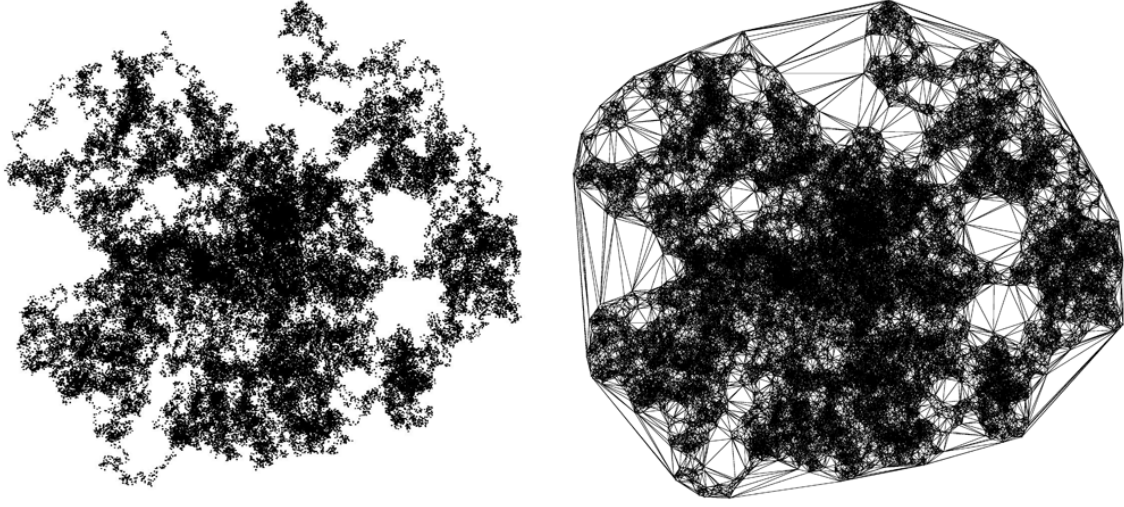


Fig. 4.1: Left: *Inhomogeneous point distribution for 10^5 points. It is the result of a fractal point process with fractal dimension $D = 0.5$, as described in Appendix 4.A.* Right: *The resultant Delaunay graph, with line lengths covering a wide range of scales, in this case 1 : 100, which is in fact far too large to be expressed in print.*

in which $\sigma = \sum_j \sigma_j$ is the total cross section, possibly consisting of many different interaction cross sections σ_j , each having its own mean free path $\lambda_j = 1/n(\mathbf{x})\sigma_j$. Because the mean free path is a statistical length, it scales in a different way with the medium density than the average Delaunay line length, which has an extra dimension dependence [cf. Eq.(4.2)]. From that, we conclude that, if we choose our point distribution to sample the d -th power of the density, i.e. $f(x) \propto x^d$, or, more specifically,

$$n_p(\mathbf{x}) = \Phi * n^d(\mathbf{x}), \quad (4.6)$$

the length of a Delaunay line $\langle L \rangle(\mathbf{x})$ between two points, will scale linearly with the local mean free path of the medium $\lambda(\mathbf{x})$ via a constant c . That is

$$\langle L \rangle(\mathbf{x}) = c\lambda(\mathbf{x}). \quad (4.7)$$

Thus, because we choose the point distribution to conform to the density profile of the medium according to Eq.(4.6), the average Delaunay line length and the mean free path have the same n^{-1} dependence, by which Eq.(4.7) is a *global* relation with a *global* constant c . In other words, a lattice composed according to Eq.(4.6) may look very inhomogeneous to us, but from the point of view of the transported entity (e.g. a photon) the grid is homogeneous!

There are two things to note. First, the medium, and thus the medium density distribution, might evolve. In that case the lattice, which is Lagrangian by definition of Eq.(4.6), will evolve with it. In most relevant cases, the transport of particles through a medium is studied with respect to static media, but for several cases, such as radiation hydrodynamics, it is worthwhile to keep in mind that the medium density $n(\mathbf{x})$, and thus the point density $n_p(\mathbf{x})$, can change with time. This is also how we see a possible extension to the nonlinear regime of, for example, hydrodynamics. As in the DSMC method, we could define a linear

Monte Carlo-like transport process through a background medium, now consisting of the particles themselves; in this case, however, the background medium, i.e. the particles, can not be considered static. As such, the lattice needs to be updated very frequently, which is a costly, and intricate task.

Second, one can choose the correlation function to be any monotonically increasing function different from the one in Eq.(4.6), but in that case the global constant c would change into a locally varying function $c(\mathbf{x})$. For example, if we choose $f(x) \propto x^e$, where $e \geq 0$, that varying function would be $c(\mathbf{x}) \propto n^{(d-e)/d}$. This stratagem bears some resemblance to the "variance reduction" techniques known from Monte Carlo algorithms.

4.3

Lattice Properties

In the previous section, we described how we construct the adaptive random lattices based on the medium density distribution. Before we set out to define the way one can transport particles on this lattice, we need to discuss the exact statistical properties of the lattice, and especially the errors associated with its stochastic nature.

4.3.1 Distributional Equivalence

The linear correlation of Eq.(4.7) can be taken one step further, by relating the full k -th order moment of the free path length Eq.(2.37) to Eq.(4.2), and recognizing that, by choosing a correlation function $f(x) \propto x^d$, not only the first order moment, but *all* k -th moments of the exponential distribution in Eq.(2.36) will scale linearly with the k -th order expectation of the Delaunay line length, i.e.

$$\langle L^k \rangle(\mathbf{x}) = c(k)\lambda^k(\mathbf{x}), \quad (4.8)$$

in which $c(k)$ is still a global constant, but one that now depends on the order k . This similarity in statistical properties is, of course, not very surprising, because the interval between two events, or the distance between two points, in the Poisson point process in Eq.(3.1) has an exponential distribution $p(x) = n_p |A| e^{-n_p |A|x}$ equivalent to that of the path length between two events in Eq.(2.36). It is to be expected that the distribution function for the average distance, or Delaunay line lengths between two of those points follows a similar distribution, with some modifications because of the dimension of space. The mean free path statistical length is one-dimensional always, not depending on dimension, so by choosing the correlation Eq.(4.6) we remove the dimensional dependence of the expectation values Eq.(4.2), by which the path lengths of particles *and* the average lattice line lengths are distributed similarly.

4.3.2 Length Sampling

We described how to construct a lattice, which is homogeneous and isotropic locally, and which adapts to the medium properties. Moreover, we showed that, by choosing a smart correlation function, the lattice becomes a direct representation of the 'free path space' of the particles. That is, locally, the Delaunay line lengths have the same distributional properties, i.e. the same k -th order moments, as the path lengths. Thus, the Delaunay line originating from one point in the medium has distributional properties (k -th order moments) which all scale linearly with the distributional properties of the path lengths originating from

that point. Stated differently, the variance of the Poisson process is now not associated with noise, but is exactly the variance of the free path of the particle.

Given a number of points, or resolution N , we construct one instance of the ensemble of the point distribution, and the free paths are accurately sampled at each one of those points. This could cause inaccuracies, once the medium between two points is not locally homogeneous, so that it does not have enough sampling points. Thus, we need to impose as a sampling condition, that in the vicinity of each point (or, more accurately, within each Voronoi cell), the medium can be considered homogeneous, a condition needed in almost every numerical method. Symbolically, we need $\frac{1}{n(x)} \frac{\partial n(x)}{\partial x} < \frac{1}{\langle L \rangle}$. This can be dealt with in two ways. First, one can increase the number of points N , by which the global parameters $c(k)$ decrease correspondingly, until the condition is satisfied. It is obvious that, when $N \rightarrow \infty$, the density field is sampled continuously, and the result is exact. Second, when the available resolution is less than needed, we can construct, by the ergodic property of Poisson point processes (cf. Sect. 3.3.1), several instances of the same point distribution, and overlay them afterwards. This is allowed, because the exponential distribution, on which the Poisson point process is based, is memoryless, so that the individual instances are independent. Note that the sampling condition is reached faster by choosing a correlation function as in Eq.(4.6). A more in-depth discussion on sampling theorems for spatial point processes will be given in Chapter 6.

4.3.3 Angular Sampling

Another variable which has to be sampled accurately, is the number of directions into which a particle can propagate at each point. As said, the average number of directions at each point is fixed (6 and 15.54, in 2D and 3D, respectively), and will not increase when the resolution or the number of instances increases. We have to take this into account when we design a transport algorithm, in which we want to conserve momentum at each point.

If we construct many instances of the ensemble of point distributions, the directions of the lines will differ for each instance, and, eventually, the continuous rotation group $SO(d)$ will be sampled continuously. We can, however, apply the ergodic principle, and state that it is equivalent to replace the ensemble average by a volume average. The number of directions within a volume, containing N points, scales as $O(N)$. Thus, within a certain locally homogeneous medium, the number of points, and henceforth the number of lines has to be large enough to accurately sample the unit sphere. This amounts to stating that a random lattice is isotropic. Thus, we can conclude that, when the number of points within a certain small region increases to infinity, the number of directions will follow that trend, and the angular sampling will become infinitely precise.

There is an additional effect, however. The angular sampling resolution Eq.(3.13) for the Poisson Delaunay graph is basically a worst case result. One expects, for inhomogeneous point distributions, that lines emanating from one point can be closer together, so that the angular resolution can become higher. In fact, we conjecture that the angle between longer Delaunay lines (large mean free paths), which lie in tenuous parts of the medium, is much smaller than the angle between short lines (short mean free paths), which lie in dense parts. This is very advantageous, because the longer the lines, the higher the angular resolution needed. One way to assert this claim is to examine the fractal point process depicted in Fig. 4.1, left. We refer the reader to Appendix 4.A for details on the construction of this point process.

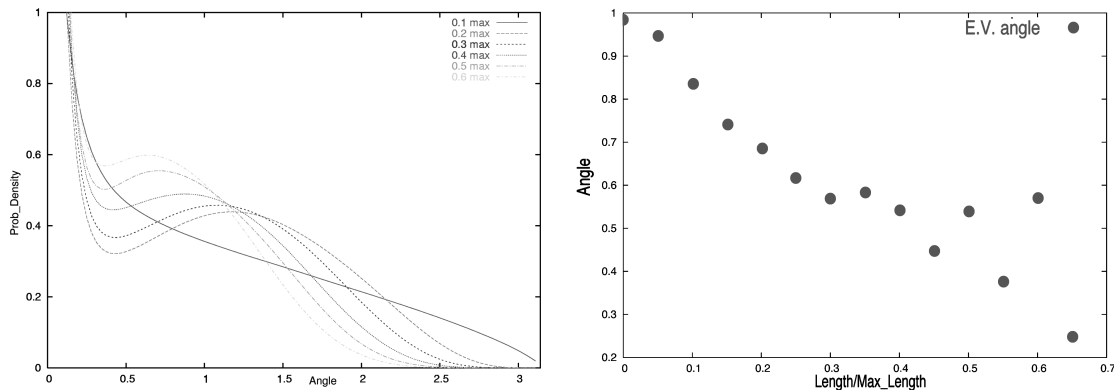


Fig. 4.2: Left: Normalized distribution functions for the angle spanned by two Delaunay lines in a triangulation of a fractal point process in \mathbb{R}^2 . Plotted are the results for several length bins. Right: First order moments of each distribution function.

For the Delaunay graph constructed on the basis of this fractal point process, we can order clockwise the Delaunay lines connected at each vertex. Evaluating the average length of two neighboring lines, as well as the angle between them, we make a statistic of the length-angle correlation. We sample the lengths using 20 bins, going from length 0 up to the maximum edge length (within the triangulation), and we sample the angles by using 50 bins in the range $[0, 2\pi]$. Thus, we can plot a normalized distribution function for the angle for each edge length bin. The result for several length bins is depicted in Fig. 4.2, left. The lines were made using Bézier curves (see Bartels et al. 1998) so as to approximate the trend of the data points. Use of these Bézier curves is justified, because we are only concerned with the trend, or overall behavior, and not with accurate quantitative results. We remark that the distribution functions do go to zero, as is required, when we put more angle bins near zero. It is still interesting that the distribution function has high values just above zero. For a normal Poisson distribution of points, the distribution function for the angle between two Delaunay edges can be derived analytically as $f(\theta) = \frac{4 \sin \theta}{3\pi} [(\pi - \theta) \cos \theta + \sin \theta]$ (e.g. Icke & van de Weygaert 1987), which decreases monotonically when approaching $\theta = 0$. Apparently, the introduction of an inhomogeneous, in this case fractal, point process introduces a large number of small angles. One can readily see that, when the average edge length increases, the average angle between the two lines decreases. This is even more apparent when plotting the expectation value of the angle versus the edge length, as is depicted in Fig. 4.2, right. There is a clear downward trend, with more scatter towards larger values of the length because of the increasing statistical noise.

Thus, we can conclude that the angle between two long Delaunay lines will be smaller than the angle between two short lines. Thus, the adaptive random lattice automatically ensures that the angular resolution increases when the lines get longer. This is another property of the Delaunay graph based on inhomogeneous point distributions, that demonstrates that it is highly advantageous to use *adaptive* random lattices. Although we have only demonstrated this angle-length correlation for the planar graph, a similar trend can also be seen in higher dimensional space.

4.4

Transport

The Lagrangian random lattice, described in the previous subsection, is a direct representation of the ‘free path space’ of the medium. As such, the stochastic element of the transport of particles on a fixed, deterministic grid in regular Monte Carlo methods is moved to the simple deterministic transport of particles on a lattice, which now has the stochastic properties. Moreover, we can dispense with the regular grid or the underlying medium altogether, because all the information needed for the transport of particles has been given to the lattice, so that the transport of particles through a medium has been translated to the transport, or percolation, of particles on the graph consisting of the lattice lines.

In the following, we will describe how we can use this lattice as a basis for transporting particles, depending on what kind of process we would like to model. As an example, we will use the two elementary transport processes, ballistic transport (possibly with an additional absorption term) and transport through a scattering medium, and combinations hereof.

4.4.1 Ballistic Transport

We will begin by describing how we can transport particles along the lattice for a medium in which the scattering cross section is negligible. Because, in this case, we transport particles which have a predetermined momentum, it is of utmost importance to define what we do at each grid point to ensure momentum conservation, and prevent numerical diffusion. Because our method works in such a way that the line lengths correlate linearly with the mean free paths, we assume that the homogeneously distributed medium is absorptive, and that the associated cross section σ_{abs} is the only contribution to the total cross section. In this case, we obtain a lattice which is similar to the Delaunay graph in Fig. 3.1, right.

We define one point as a source of particles, sending them out along one of the lines. The particles move along the line, until they come upon the next point. They have moved along a line, the length of which correlates linearly with the mean free path via a constant factor c_{abs} . We can exactly evaluate the value of this factor, given Eq.(4.2) and Eq.(4.6), as

$$c_{\text{abs}} = \xi(d, N, D)\sigma_{\text{abs}}, \quad (4.9)$$

in which $\xi(d, N, D)$ is some constant depending on the choice of dimension d , the number of points N , and the size of the domain D . Note that each other type of particle-medium interaction j would have a coefficient c_j determined via a relation similar to that in Eq.(4.9). Ideally, one would want all $c_j = 1$, but, almost always, the resolution N will result in a coefficient larger than unity, in which case the line length is larger than the local mean free path, or smaller than unity, in which case it is the other way around. This can be taken care of by defining how the interaction, in this case absorption, is accounted for at each point. To accomplish this, we define the incoming number of particles, or intensity, as I_{in} , and determine the outgoing intensity as

$$I_{\text{out}} = I_{\text{in}}e^{-c_j}, \quad (4.10)$$

which is equivalent to the familiar $I(x) = I_0e^{-x/\lambda}$, which can be derived from Eq.(2.36). Note that, with c_j being a global constant, e^{-c_j} is a global constant too, which can be

determined a priori, via Eq.(4.9). In some cases, when the resolution N is high, and thus the coefficients c_j small, it might be useful to approximate $e^{-c_j} \approx (1 - c_j)$.

Locally the number of particles absorbed can be exactly evaluated as

$$I_{\text{abs}} = I_{\text{in}} (1 - e^{-c_j}), \quad (4.11)$$

which ensures that this method conserves particles exactly. The remaining particles I_{out} have to be sent out along one of the lines emanating from this point. The directions we choose depend on whether we want to conserve momentum, or if the interaction was with a scattering medium, in which case we want to distribute the particles isotropically. We can even incorporate anisotropic scattering, such as Rayleigh scattering. In the present case, we need to ensure momentum conservation, and we have to choose an outgoing line which is in the same direction as the incoming one. As we already discussed in the previous section, the lines only sample the unit sphere on average, and although the Voronoi cells are cylindrically symmetric, on average, with respect to every incoming Delaunay line, every one particular cell will deviate from that (cf. Sect. 3.3.4 for a discussion on this deflection angle). This has as a result, that almost always, there is no outgoing line in the same direction as the incoming one. Thus, particles are deflected from their original direction by the irregularity of the grid. This can be viewed as the introduction of space-dependent inertia forces, and it is very important to keep track of these, especially in the nonlinear regime (Benzi et al. 1992; Karlin et al. 1999), as we shall point out in the next chapter.

We can resolve this by doing the following. We refer to Fig. 4.3, which is an example in 2D space. In this figure, the dashed line is the line along which momentum would be conserved. Instead, if we now choose the ‘most straightforward’ line II as the outgoing line, in which nearest can be defined in several ways, the most intuitive being that line for which the inner product with the dashed line is largest, it can be shown that momentum is conserved on average, which immediately follows from the fact that the mean Voronoi cell is axisymmetric with respect to every incoming Delaunay line. From a different perspective, we can say that the distribution function for the deflection angle in Fig. 3.6 is symmetric around $\alpha = 0$. However, from a numerical implementational point of view, it is more elegant to split up the two particles and distribute one of them along line II and the other along line III, being the next to most straightforward path. More generally, if the transported quantity is a continuous entity, we have found that it is most efficient to split this quantity into d equal parts, in which d is the dimension of our computational domain, and distribute each of these parts along the d most straightforward paths. Because by definition, the Poisson-Voronoi cells are axisymmetric with respect to every associated Delaunay line in every dimension $d \geq 2$, we know that a similarly modified set of rules will ensure conservation of momentum. From the same alternative perspective as before, we can say that, if we had not chosen to split up the packages, our sole outgoing line would only sample one side of the deflection angle distribution function in Fig. 3.6. Choosing more lines ensures that this distribution is sampled more accurately. We note that it is possible to introduce a functionality $f(\alpha)$ in this splitting procedure, for

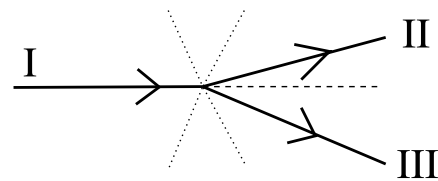


Fig. 4.3: Example of a number of Delaunay lines meeting at a node. Incoming particles (along line I) interact with the medium at the node, and the remaining particles should continue along the dashed line to conserve momentum. Choosing lines II and III as outgoing ensures conservation of momentum, on average.

example based on the value of the deflection angles (lines with larger deflection angles get less). However, we found experimentally that the overall numerical result is more accurate, when we omit this. Introducing any function of the angle was shown to enhance local asymmetries between the d most straightforward lines. The distribution function is symmetric anyway, which favors our choice for splitting the particles in equal parts.

To demonstrate conservation of momentum, we constructed a 2D random lattice with a relatively low resolution of $N = 5 \cdot 10^4$ points, which is sufficient to show the trends and is still coarse enough to clearly show noise. Next, we define one point as source of particles, each emitted with the same momentum vector. If the particles were defined to be photons, this source might be a laser. The rules at each site are chosen such that the two most straightforward paths with respect to this momentum vector are chosen, and that the incident package of particles is split in two and continues along these two lines. We follow the particles until they hit the absorbing boundary. The result is plotted in Fig. 4.4, left, in which we plotted the logarithm of the number of particles at each vertex, given that the source emits a high number of particles. One immediately sees that, on average, the resultant momentum vector is in the original direction, but that there is some inherent noise associated with the use of these random lattices. Fortunately, one can prove that this noise vanishes in the limit of $N \rightarrow \infty$, by recognizing that in d -dimensional space the propagation of a particle on a lattice with this set of rules is equivalent to the process of an anisotropic random walk on a graph. An exact mathematical derivation is given in Appendix 4.B, and we give a simplified version here.

Defining a signal $\mathcal{S} \propto \langle L \rangle n$ as the (average) distance travelled after n steps, and realizing that for an anisotropic random walk we have a noise $\mathcal{N} \propto \langle L \rangle \sqrt{n}$, we obtain a signal to noise ratio $\mathcal{S}/\mathcal{N} \propto n^{1/2}$. This is why the beam in Fig. 4.4 does not diverge. Given a particle location x along the momentum vector at a distance s from the source, we know that the average number of steps n for a particle to reach s scales as $n \propto s/\langle L \rangle$ in which $\langle L \rangle \propto N^{-1/d}$. From this, we can conclude that the signal to noise ratio at a distance s from the source scales as

$$\frac{\mathcal{S}}{\mathcal{N}} \propto N^{1/2d}. \quad (4.12)$$

Thus, in the limit of $N \rightarrow \infty$, momentum is conserved exactly, not just on average. In other words, the width of the beam will shrink to zero. Note that, when N increases, the interaction coefficients $\{c_j\}$ will decrease.

Thus, we have shown that choosing this transport recipe for ballistic transport will not only conserve momentum on average, but it does also deal with the coarse angular sampling at each point of the lattice, ensuring that any numerical diffusion will be minimal. Moreover, we have shown that this artificial widening of the beam will go to zero for infinite resolution.

4.4.2 Transport through Scattering Media

In the previous section we discussed how to do transport of particles through a medium for which the interaction conserves momentum, and does not change the original momentum vectors of the particles. It is also possible that one of the interactions $\{c_j\}$ influences the vectorial properties of the particles, for example when the transport is through a scattering medium.

Without loss of generality, we assume a similar homogeneous medium as before, without absorptive properties but *with* an extra interaction coefficient, σ_{scat} , which accounts for the scattering properties of the medium. In this case, the method for transportation is rather

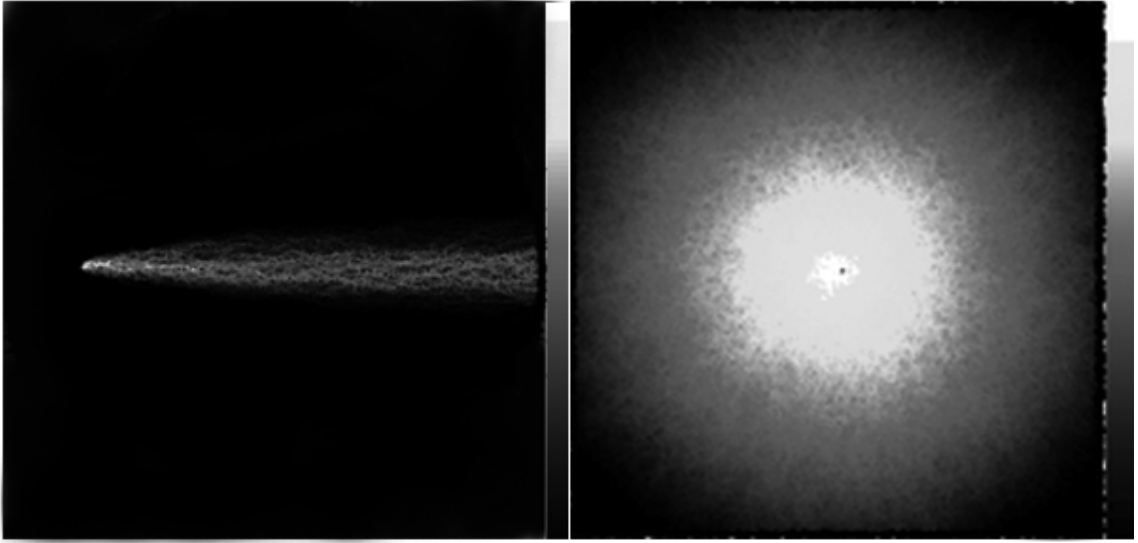


Fig. 4.4: The result of two simple test on a 2D Poisson-Delaunay random lattice with $N = 5 \cdot 10^4$ points. Both are logarithmic plots of the number of particles at each site. Left: Illustration of the conservation of momentum by means of the transport of particles with constant momentum vectors. Right: Illustration of a scattering transport process.

similar to the previous section, in that we propagate the scattered particles $I_{\text{out}} = I_{\text{in}}e^{-C_{\text{scat}}}$ along the d most straightforward paths. What differs is that the retained particles $I_{\text{ret}} = I_{\text{in}}(1 - e^{-C_{\text{scat}}})$ are now redistributed isotropically. To accomplish this, one can choose to propagate an equal fraction along every Delaunay line. The angular sampling is not perfect, but similar considerations as made in Sect. 4.3.3 will ensure angular resolution.

An example of one such experiment is depicted in Fig. 4.4, right, in which we define several points in a small central region as sources, and for which we let the particles propagate along the lattice according to the rules described above. We take several points as a source in this case, because this will ensure that we have many different original directions. Note that the result is very isotropic, as is to be expected of this random lattice. Moreover, the result exactly depicts the noise associated with the finite angular sampling, and thus the finite amount of points.

4.4.3 General Interactions

The particle-medium interaction may consist of many different types, each one having its own contribution to the right hand side of Eq.(4.1). Each of these interactions has its own cross section σ_j , and its own corresponding linear correlation coefficient c_j . Note, however, that the conversion factor $\xi(d, N, D)$ in Eq.(4.9) is the same for all interactions, given one choice of d , N and D . The total interaction factor $C = \sum_j c_j$ determines what factor of the incoming particles interacts, in one way or another, $I_{\text{out}} = I_{\text{in}}e^{-C}$. Every single interaction in the set $\{c_j\}$ has its own contribution $\frac{c_j}{C}e^{-C}$ (cf. Sect. 2.6). When our resolution is high enough, and thus our factors $\{c_j\}$ correspondingly small, we can approximate the total factor of the particles that interact as $(1 - e^{-C}) \approx C$, from which one can conclude that every interaction retains a factor c_j of the incoming particles. This last approximation can be implemented more efficiently.

The behavior of particles depends on the type of interaction, and the type of transport

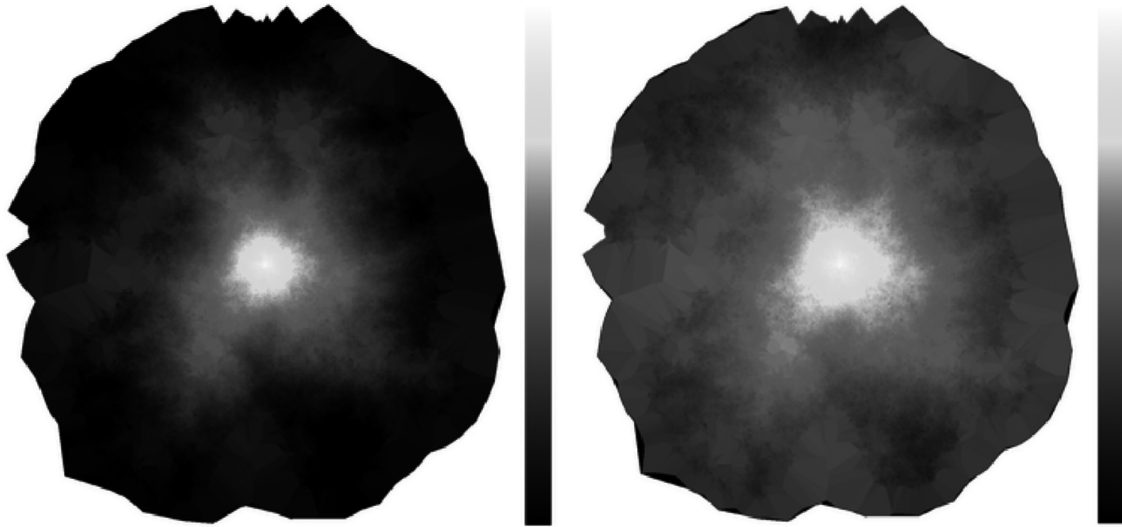


Fig. 4.5: Plot of the logarithm of the number of particles for the result of a transport process in which the interactions consist of scattering and absorption. The scattering interaction coefficient $c_{\text{scat}} = 1.0$, on the grid of Fig. 4.1, is the same in both, whilst the absorption interaction coefficient c_{abs} is on the left is twice the value used on the right. Note that the scales are the same for both plots.

process in general. Particles may be redistributed isotropically, and continue on through the medium, or they may influence the medium by heating, ionization, or something similar. They might even be re-emitted as different particles. All these types of interactions can be implemented very easily, and the feedback of the particles on the medium is defined straightforwardly. One might even have multiple species of particles, in which case there would be a cross section σ_j for each type of particle, and the set of factors c_j would change into a matrix, which could include the transformation of one particle into the other. An example of the result of our method for the transport of particles through a medium in which several interactions (in this case, scattering and absorption) play a role is given in Fig. 4.5. As a mesh, we have used the fractal random lattice in Fig. 4.1, and we defined a small central region of points as an isotropically emitting source. Each point was assigned a constant interaction coefficient $c_{\text{scat}} = 1.0$ for scattering, but also a certain constant interaction coefficient c_{abs} for absorption. Left is the result for $c_{\text{abs}} = 0.050$, and right is one for $c_{\text{abs}} = 0.025$. Remember that $c_{\text{abs}} = 0.050$ means that the average Delaunay edge length is twenty times as small as the mean free path for absorption, cf. Eq.(4.7).

One should aim at having enough resolution to ensure $c_j \leq 1$. In cases, where scattering can be neglected, the solution based on Eq.(4.10) would still be exact, as we could not resolve the space between the points anyway, but when wanting to incorporate scattering, one might not accurately resolve the diffusion coefficient, when $c_{\text{scat}} \gg 1$. We refer to Chapter 6 for a more elaborate discussion on sampling criteria.

4.4.4 Time Stepping

So far, we have not elaborated on what we define as a time step. As in linear Monte Carlo methods, this depends largely on the transport problem at hand. We may distinguish two cases: A) the interaction dominated limit ($\text{Kn} = \lambda/L \ll 1$), and B) the free streaming limit ($\text{Kn} = \lambda/L \geq 1$).

The first case is what is mostly encountered in linear transport problems. Here, the mean free paths for the particles are so small compared to the size of the domain, that we know the particle will be absorbed somewhere within the domain. Given that the speed of the particles is very high, one can, for each time step, let the particles that are emitted in that interval move along the lattice, moving from one interaction point to the next, until they are annihilated. The time steps do have to be chosen such that one can accurately sample time dependent source functions, or such that, when absorption is followed by ionization (as is the case for cosmological radiative transfer, cf. Part III of this thesis, one can accurately follow the ionization front. Equilibrium solutions can be found, if they exist, by having a constant number of particles being emitted at each time step, that will exactly compensate for the number of particles that are absorbed, or leave the computational domain.

The second case is not very interesting from a particle-medium interaction point of view, because the densities and cross sections are of such a form that no interactions are expected to occur. Thus, the particles can stream freely through space. In this case, one may be interested in following the particles' trajectories, and the time steps are then dictated by the size of the domain divided by the speed of the particles.

An interesting, albeit a bit artificial, intermediate case is the one described in Sect. 4.4.2. In this case of pure scattering, the particles are not expected to be annihilated. If we would not stop the particles at one point, they would eventually leave the domain. Thus, if we want, for example, to accurately follow the spherical wavefront around a point source in a homogeneous scattering medium, as it expands with time according to the familiar Gaussian diffusion profiles, we need to keep a clock for each particle, such that it does not take more mean free path steps than are allowed by its maximum speed.

It is clear that the relevant time step criteria depend greatly on what one is trying to solve. In each case, however, one may define the relevant time stepping unambiguously. Most linear transport problems we are interested in, including the one discussed in Part III of this thesis, will be of type A.

4.5

3D and Beyond

Because it may not seem obvious how our method is trivially extensible to \mathbb{R}^d , given all the 2D examples we gave, we will explicitly state how this extension is automatically achieved by the lattice construction procedures we gave.

First, the recipe for constructing our adaptive point process using Eq.(4.6) is applicable and valid in general dimensional space. Second, the procedure for constructing the Delaunay triangulation, and the corresponding recipe for the Voronoi tessellation, can be used and implemented efficiently in every \mathbb{R}^d (Okabe et al. 1999). Because we chose a correlation function that includes the dimension of space $n_p \propto n^d$, we made sure that the linear relationship between all line lengths of the resultant lattice and the local mean free paths is valid in every dimension, cf. Eq.(4.8). Thus, we have obtained an adaptive random lattice that adapts to the medium in exactly the same way for every \mathbb{R}^d .

As we have already pointed out, the resultant transport process is a walk from one interaction event to the next (as it is in Monte Carlo methods), or, alternatively, a walk from one vertex to the next along a d -dimensional Delaunay graph that has mean free path-like line lengths. The recipes for the different types of processes (ballistic, scattering, and combinations) can be extended trivially to \mathbb{R}^d . For example, for ballistic transport we do not

choose the two most straightforward paths, but the d most straightforward. The analysis concerning the conservation of momentum in the appendix is valid for every dimension, because it only assumes that the mean Voronoi cell is cylindrically symmetric around every associated Delaunay edge. This symmetry is a natural consequence of the motion invariance property of the underlying Poisson point process.

The only things that change from dimension to dimension are the average number of lines emanating from a typical point, or the number of walls of a typical Voronoi cell (6 in \mathbb{R}^2 , 15.54 in \mathbb{R}^3 , etc.), and the geometrical constant $\zeta(k, d)$ in the expressions for the k -th order moments of the Delaunay line lengths Eq.(4.2). These are mere constants that one has to incorporate when implementing the numerical method.

When using a regular grid, and the often associated finite-differencing, or some nontrivial form of interpolation, it is often very nontrivial to extend these operations to higher dimensional space. Because we chose to use an adaptive random lattice on which a (random) walk is performed, these difficulties are resolved, because both the lattice construction techniques and the (random) walks are trivially defined in every \mathbb{R}^d .

4.6

Efficiency

It is straightforward to implement our method using a high-level programming language. We have already done this, using C++, for the transport of radiation. We will give an in-depth description of this `SimpleX` package in Chapter 8.

For now, we will describe the basic steps of the algorithm. First of all, there have to be several pre-processing steps: 1) create a point process matching Eq.(4.6), in which the medium density function can be an analytic function or some data array from some other simulation; 2) construct the Delaunay triangulation; 3) determine the properties of the resultant lattice, namely the global interaction coefficients $\{c_j\}$, and the d most straightforward paths with respect to the other lines. All these values are fixed during the rest of the simulation. As an illustration, given an efficient tessellation code, such as `QHull` (cf. Sect. 3.4), these preprocessing steps can be completed within one minute wall-clock time on a simple desktop computer, for a resolution of $N = 10^6$ points.

Once the lattice and all its properties are known, the transport can commence. If we define one iteration of the algorithm as advancing particles from one point to the next, and subsequently performing the interactions at each point, and that for each point, we can make an estimate of the operation count of the algorithm. Each point can be dealt with independently of all the others (cf. the Markov property of Monte Carlo methods). Given that the particles can be redistributed along other angular directions at each point, each line has to be treated separately, but this is just a geometrical constant, given the dimension d , which does not scale with any resolution. Linear multiplications as in Eq.(4.10) have to be performed for each interaction j , and for each different particle species. The redistribution along the lines is just a pointer operation. Thus, the total operation count is of the form

$$N_{\text{ops}} = N_{\text{int}} N_{\text{spec}} N_p, \quad (4.13)$$

in which N_{int} , is the number of different interactions, or non-negligible cross sections σ_j , N_{spec} is the number of species, and N_p is the resolution, or number of points. Thus, if we just focus on the resolution, the operation count of this method scales as $O(N)$.

It should be noted that this is *independent of the number of sources*. As it turns out, most other transport methods scale with the number of sources, by which it is extremely time-consuming to do realistic calculations for a large number of inhomogeneously distributed sources. With our method, this is now feasible, even on a simple desktop computer. Apart from the fact that we now use a lattice that directly corresponds to the physical problem, this scaling behavior is the most important characteristic of our transport method. It enables us to do unprecedented simulations of the first light in the Universe. More on this in Part III of this thesis. Another thing we should point out is that our method bears much resemblance to cellular automata methods (discussed in the next chapter), in the sense that we have a global set of interaction coefficients, and a global set of rules, applied locally. Each point is influenced only by its neighbors (via the Markov criterion). Thus, like most cellular automata methods, our method can be parallelized trivially. Of course, this would be much less trivial, if nonlinear terms were included.

We conclude this section by noting that this new method is a stochastic method, just as the more usual Monte Carlo methods. As discussed in Sect. 2.6, transport methods of this kind obey the $N^{-1/2}$ -law, which says that the stochastic noise goes down with this functionality. Indeed, we explicitly derived this form in the derivation of the signal-to-noise ratio Eq.(4.12).

4.7

Conclusion

In this chapter, we have described how our new method works. We have shown how to construct a Lagrangian random lattice, which mimics the medium properties in such a way that locally all line lengths have the same distributional properties as the particles' path lengths. As such, the method can handle any geometry of the medium. The grid retains the translational and rotational symmetries inherent in most physical problems, whilst at the same time adapting to the medium properties, so that small mean free paths and rapidly fluctuating regions of the medium will not be undersampled. The resultant lattice is a direct representation of the free path distribution space of the medium. The stochastic character of the the particle trajectories, as in regular Monte Carlo, has been transferred to that of the grid itself, which can be shown to sample space, and all angular directions exactly when the resolution goes to infinity. Particles can be easily transported along the lattice lines, with the interactions taking place at each grid point, via an interaction coefficient c_j , which is directly proportional to the interaction cross section. Moreover, we have shown that the operation count of an implementation of the method is $O(N)$, which makes the method fast, even when increasing resolution, and suitable for performing 3D calculations.

It is important to stress that it is meaningful to extract observables from the simulations performed with this method. By this we mean that there are two complementary ways to determine certain transport variables from the transport method, and that it can be shown that the two give equivalent results. First, the mathematically exact approach would be to construct infinitely many instances of the (generalized or regular) Poisson point process, and determine via simulations on each of the resultant random lattices what, for example, the number of particles is at one particular point. The observable would then be an average over all these instances. Second, one can take one instance, and, after doing one simulation on that one random lattice, extract the number of particles at one point by taking an average of points within a suitably small volume around that point. The equivalence of these two

approaches is apparent by the ergodic properties of the Poisson point process, discussed in Sect. 3.3.1. This is very essential, because this shows that we get meaningful results by doing simulations on only one such lattice, provided that the resolution of points is high enough to be able to extract local averages. When the available resolution is insufficient, the alternative approach then makes it possible to decrease the stochastic noise of the results by doing more simulations using different instances. More on this in Chapter 6.

We conclude by noting that although in this chapter we focused on the transport of particles through a static medium, the method can be used much more generally. The space in which the path lengths are defined does need to be defined as a subset of phase space, but can be much more abstract, and diverse. Moreover, the interactions do not only have to be of the type particle-medium, but can also be between the particles themselves, as in hydrodynamics, or both. In each case, the method will consist of constructing a Lagrangian random lattice, as a direct representation of what the transported quantity would encounter, when it is transported through that abstract space.

Acknowledgements We thank Yuri Levin for carefully reading the original manuscript, and Garrelt Mellema and Erik-Jan Rijkhorst for fruitful discussions.

Bibliography

- Bartels, R. H., Beatty, J. C., & Barsky, B. A. 1998, *An Introduction to Splines for Use in Computer Graphics and Geometric Modeling* (San Fransisco, CA: Morgan Kaufmann), 211–245
- Benzi, R., Succi, S., & Vergassola, M. 1992, *Physics Reports*, 222, 145
- Garcia, A. L., Bell, J. B., Crutchfield, W. Y., & Alder, B. J. 1999, *J. Comput. Phys.*, 154, 134
- Icke, V. & van de Weygaert, R. 1987, *A&A*, 184, 16
- Karlin, I. V., Succi, S., & Orszag, S. 1999, *Phys. Rev. Lett.*, 82, 5245
- Lowen, S. B. & Teich, M. C. 1995, *Fractals*, 3, No. 1
- Mandelbrot, B. B. 1982, *The Fractal Geometry of Nature* (W.H. Freeman & Co.)
- Okabe, A., Boots, B., Sugihara, K., & Chiu, S. 1999, *Spatial Tessellations, Concepts and Applications of Voronoi Diagrams*, 2nd edn. (John Wiley & Sons)
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, *Numerical Recipes in C: The Art of Scientific Computing* (New York, NY, USA: Cambridge University Press)
- Wu, J.-S., Tseng, K.-C., & Kuo, C.-H. 2002, *Int. J. Numer. Meth. Fluids*, 38, 351

4.A

Fractal Point Processes

Fractal (stochastic) point processes have been widely used as a modeling tool (for a review, see Lowen & Teich 1995). In particular, Mandelbrot (1982) used it to model a non-standard random walk resulting in a linear Lévy dust, which, as he showed, could even be used to model clusters of galaxies. The statistical properties of fractal point processes are scale-free by definition. Conclusions drawn from the resultant tessellations will be very general with respect to the length scales involved. We used a modified version of his recipe, namely a Lévy flight, which is a sequence of flights separated by stopovers.

It is constructed by choosing the first stopover randomly, and starting the flight from that point. The (straight-line) flights have the following properties: their direction is random and

isotropic, the different angles are statistically independent (thus, the Lévy flight is a Markov process), and their lengths follow a probability distribution function

$$f(r) = kr^{d-1-D} = kr^{1-D}, \tag{4.14}$$

in which d is the dimension of the space in which the flights occur (in our case, $d = 2$), k is a normalization constant, and D is the *fractal dimension* as defined in Mandelbrot (1982). Eq.(4.14) is a modification of the distribution in Mandelbrot (1982) to the effect that the clustering of points near $r = 0$ is avoided. Thus, if $D = 0$, we obtain the regular Poisson point process, and if $D > 0$, we have an exponentially decaying distribution function, which is scale-free as required for a fractal distribution. Clustering increases when D is increased. It is only the stopovers we are interested in, because these will be the points of the resultant fractal point process. As said, the process is scale-free and Markovian, so that it is permitted to rescale and translate the resultant point distribution, so as to center and fit it in our $[0.0, 1.0]^2$ domain, without altering its statistical properties.

We have to choose a lower and upper bound (A and B , respectively) for r . Thus, we obtain for the flight-lengths the cumulative distribution function

$$F(r) = \frac{r^{2-D} - A^{2-D}}{B^{2-D} - A^{2-D}}. \tag{4.15}$$

Given a cumulative distribution function, we can use the Direct Inversion method (Press et al. 1992) to randomly sample it. To accomplish this, we fix the upper bound B (the maximum flight length) as half the width of our domain, and choose the lower boundary A to be two orders of magnitude smaller. Thus,

$$r \in [0.01, 0.5]. \tag{4.16}$$

An example of this process, satisfying Eqs.(4.15) and (4.16), with $D = 0.5$ and $N = 10^5$, is shown in Fig. 4.1.

4.B

Conservation of Momentum

The symmetry of the distribution of the deflection angle in Fig. 3.6 enforces that momentum is conserved on average, given the ballistic transport recipe in Sect. 4.4.1. What is more interesting, though, from an implementation point of view, is the width of the distribution around the original direction. That is, how wide would an otherwise infinitely thin laser beam become as function of the distance from the source. For an exact mathematical analysis hereof for the Poissonian random lattice, we proceed as follows.

An example of a path of a particle performing a walk in two dimensions is given in Fig. 4.6.

The following analysis, however, will be valid in d -dimensional space. Because of cylindrical

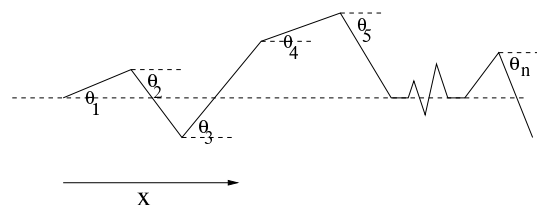


Fig. 4.6: One possible path of particles performing a walk of n steps on the Delaunay graph. The i -th step is parametrized by an angle θ_i , with respect to the original direction x .

symmetry around the original direction \mathbf{x} (the distribution function of the deflection angle is symmetric), we can parametrize the i -th step of the particle's walk by one angle θ_i , which is the angle between the i -th Delaunay edge and the original direction \mathbf{x} . Thus, the expectation value of the total displacement $\mathbf{R}_n = \mathbf{r}_1 + \dots + \mathbf{r}_n$ is

$$\begin{aligned}\langle \mathbf{R}_n \rangle &= \langle \mathbf{r}_1 \rangle + \dots + \langle \mathbf{r}_n \rangle \\ &= n \langle L \rangle_I \langle \cos \theta \rangle_\theta \frac{\mathbf{x}}{|\mathbf{x}|} \\ &= n \langle L \rangle_I \chi \frac{\mathbf{x}}{|\mathbf{x}|},\end{aligned}\quad (4.17)$$

in which $\langle L \rangle_I$ is the average Delaunay line length, defined in Eq.(4.2), and

$$\chi = \int_{-\pi}^{\pi} h(\theta) \cos \theta d\theta. \quad (4.18)$$

Note that in this appendix, we explicitly distinguish averaging over the length distribution function, denoting this as $\langle \rangle_I$, and averaging over the deflection angle distribution function, denoted as $\langle \rangle_\theta$. These can be explicitly shown to factorize for a Poissonian point distribution (Okabe et al. 1999). Here, we have used $h(\theta)$ as a certain symmetric function, which characterises the probability distribution of the angle θ and which, in most cases, cannot be evaluated analytically. Note that this distribution can be different from the one for the deflection angle as depicted in Fig. 3.6, because θ characterizes an angle with respect to a fixed, original direction \mathbf{x} , and not with respect to the previous Delaunay edge. As such, this walk is, strictly speaking, not Markovian.

The second-order expectation value can be evaluated as follows:

$$\begin{aligned}\langle \mathbf{R}_n^2 \rangle &= \langle \mathbf{r}_1^2 \rangle + \langle \mathbf{r}_1 \cdot \mathbf{r}_2 \rangle + \dots + \langle \mathbf{r}_n^2 \rangle \\ &= \langle L^2 \rangle_I [n + n(n-1) \langle \cos(\theta_i + \theta_j) \rangle_\theta],\end{aligned}\quad (4.19)$$

in which we may choose i and j randomly from the set $\{1, \dots, n\}$, as long as $i \neq j$, because the distribution function $h(\theta)$ has the same form for each angle θ_i . Using the cosine addition formula, Eq.(4.19) reduces to

$$\langle \mathbf{R}_n^2 \rangle = (n + n(n-1)\chi^2) \langle L^2 \rangle_I. \quad (4.20)$$

Thus, the variance of the displacement is

$$\sigma_{\mathbf{R}_n}^2 = n \langle L^2 \rangle_I (1 - \chi^2), \quad (4.21)$$

in which we made the simplifying assumption that $\langle L^2 \rangle_I = \langle L \rangle_I^2$. Of course, these differ by a constant factor $\zeta(1, d)/\zeta(2, d)$, cf. Eq.(4.2), but this only introduces a certain constant into this derivation, the conclusions of which are unaltered. When $h(\theta) \propto \delta(\theta)$, then $\chi = 1$, by which $\langle \mathbf{R}_n \rangle = \langle L \rangle_I n \mathbf{x} / |\mathbf{x}|$ and $\sigma_{\mathbf{R}_n}^2 = 0$ as should be expected. The exact form of the distribution function $h(\theta)$ is at present unknown, even for this well-studied Poisson case, but we can use a step function as an approximation. Thus, given that in \mathbb{R}^2 the average number of Delaunay lines meeting at a grid point is 6 (cf. Sect. 3.3.4), we use as a step function $h(\theta) = 3/\pi$ on the domain $\theta \in [-\pi/6, \pi/6]$. This results in $\chi = 3/\pi$, wherefore

$$\langle \mathbf{R}_n \rangle = \frac{3n \langle L \rangle_I}{\pi} \frac{\mathbf{x}}{|\mathbf{x}|}, \quad (4.22)$$

which is very close (difference of less than 5%) to the distance along a straight line, which would be $n\langle L \rangle_I$. We can always, of course, rescale the lengths so as to make sure that the distance traversed equals the exact physical one. For more on this, we refer the reader to the discussion on the shortest paths on random lattices in Sect. 3.3.4.

More importantly, the variance of the displacement, in this case, is

$$\sigma_{R_n}^2 = \frac{\pi^2 - 9}{\pi^2} \langle L^2 \rangle_I n. \quad (4.23)$$

We know that the results of using a step-function as distribution function gives upper bounds on the values of Eqs.(4.17) and (4.21), because the actual distribution function would peak around $\theta = 0$ and decreases as $|\theta|$ increases, so we expect the actual value of $\sigma_{R_n}^2$ to be smaller. Thus, we can simulate a straight line trajectory with this method, because $\mathbf{R}_n \propto n\mathbf{x}$, but with a standard deviation that increases with \sqrt{n} .

A crucial aspect is the behavior of the standard deviation, when the number of grid points N increases. Let us therefore examine a line segment in the simulation domain of length \mathcal{L} ($\leq \sqrt{d}$, if we have a $[0.0 : 1.0]^d$ domain). Because the point distribution is homogeneous, we can conclude that the number of steps to cover the line is

$$n = \xi N^{1/d}, \quad (4.24)$$

in which we have, for \mathbb{R}^2 , for example, $\xi \leq \frac{\pi}{3}\zeta(1,2)\sqrt{2}$, which can be found by using the upper bound Eq.(4.22) and the Eq.(4.2) for the length $\langle L \rangle_I$ of a Delaunay line. If we combine Eq.(4.23) with Eq.(4.24), again using Eq.(4.2), we obtain

$$\sigma \propto \langle L \rangle_I \sqrt{n} \propto N^{-1/2d}. \quad (4.25)$$

Thus, we conclude that the amount of widening of the beam will go to zero, when we increase the amount of grid points N .

Even if we do not have a large amount of points to suppress the widening of the beam, we have another effect that compensates for the widening. Namely, at each intersection the number of particles are split up into d parts. This means that the particle number at points farther away from the straight line trajectory is much less than at points close by, simply because of the fact that more paths cross each other at points close to the line.

Fields of Application

Having described our new transport method, we will in this chapter elaborate on the relationship of various aspects of our method with other fields of expertise. We show how our method can be linked to various other methods of (nonequilibrium) statistical physics, but also percolation and graph theory, and even the theory of cellular automata and Lattice Boltzmann methods. Useful analytic tools from those fields, and possible applications will be pointed out.

5.1

Introduction

Our original motivation for designing the method described in the previous chapter was to develop a fast algorithm to solve for the linear transport of photons. Parts II and III of this thesis will be devoted solely to that. This method's uses, however, are much more diverse. So, before we delve into the transport of radiation, we shall in this chapter make a brief excursion into several different fields of application.

Our new transport method is an amalgamation of different techniques from several distinct parts of science. As such, it touches upon, and brings together, many different fields of expertise, each having its own specific jargon, and its own agenda as to what part of the Universe it is trying to describe. A rough indication:

- We use Monte Carlo techniques from the field of *nonequilibrium statistical mechanics*.
- The transport takes place on the omnipresent random lattices, mainly studied in *computational geometry*.
- The resultant transport process is defined as a random walk on a graph, which is the object of study in *graph theory*.
- Interactions are incorporated via global collision rules at each vertex; similar considerations apply in the elegant *cellular automata theory*.
- The overall process solves for the (collective) behavior of particles interacting with and moving through a medium. *Transport theory* is applicable in a vast variety of situations.

It is therefore not surprising that there are many ways to elaborate on our transport method, depending on which of these fields one wishes to emphasize. Each brings its own mathematical machinery, and its own field of applications. Thus, we have tried in the previous chapter to be as general as possible in our exposition, not trying to conflict with any of these fields

of expertise. In this chapter, we will single out each of these subjects, consecutively, trying to relate our method to each. We will point out similarities, useful diagnostic tools, and even applications to which our method can contribute, in that it adds useful elements from one or more of the other areas. We shall omit the fields of transport theory and computational geometry, the connection to our method being so essential, that we have covered them already in Chapters 2 and 3.

5.2

Nonequilibrium Statistical Physics

The exposition of our method in the previous chapter focused on the linear transport of particles through a background medium. These transport processes have a much more widespread use than might seem at first glance. The general transport equation Eq.(4.1) is defined in some abstract space, which can take a wide variety of forms. Although stochastic methods for transport processes have originally been developed to solve for the transport of neutrons through a nuclear reactor, they have also proven useful in quantum chemistry, radiative transfer, radiotherapy, even population growth and genetics. As long as something is being transported through some space, whether these are photons through the atmosphere, or quantum phases along a Feynman diagram, stochastic transport methods like ours have a potential use. Although it is much more involved, nonlinear transport problems, solving for the collective behavior of particles, can be incorporated, via similar procedures as in DSMC methods (cf. Sect. 2.6), or using Lattice Boltzmann techniques, as briefly discussed in Sect. 5.4. A modified version of our transport method could even be used for fluid dynamics.

A more recent, somewhat exotic, field of possible applications for our transport method consists of those subject areas that are not normally considered to be physics at all, but for which people from that field have recognized that it is very worthwhile to apply the principles of statistical physics to tackle a particular problem from a different angle. In some cases, physicists themselves have entered these arenas by introducing certain models that present the 'physicist's view'. One of the core assumptions here, is that what one is trying to model is a *complex system*, in the sense that one is unable to follow the trajectories, or behavior, of each one of its constituents. Even without understanding the exact microscopic behavior, one is interested in making predictions, or even understanding the global behavior of the system as a whole. In recent years, people have come to understand that the principles of statistical physics can be used in wider contexts, trying to understand a variety of complex systems.

These non-physics applications take many very different forms. In biology, nonequilibrium statistical physics was used to study self-propelled particles in biological systems, such as schools of fish (Czirok & Vicsek 2000). Seen as *the* most complex system, it has become somewhat fashionable to apply statistical physics principles to understand sociological phenomena. *Sociophysics*, as it has been called, now studies how opinions are changed by contact between individuals, it predicts the dynamics of voting, and even makes claims as to understanding the spread of crime (for a nontechnical review, see Ball 2003). It has even been used to model the rise and fall of languages (Schulze & Stauffer 2005). One of the oldest non-physics applications is that of the analysis of traffic flow. Depending on the parameters and models used, it has been shown to predict free-flowing traffic, traffic jams, and combinations of the two (Chowdhury et al. 1997). Similar principles have been used to predict panic behavior of people in a confined room (Helbing et al. 2000). As it turns out, two

of the hallmarks of statistical physics, power laws and universal behavior, can also be identified in the analysis of existing financial data. In the preceding two decades many physicists have left their field, pursuing a career in *econophysics*. Many different aspects of economic systems can be reproduced, from the distribution of wealth in a closed system (Dragulescu & Yakovenko 2000), to trader-based models for stock market dynamics (Hammel & Paul 2002). Moreover, it has been accepted for quite some time in the trading business, that standard econometrics is no longer satisfactory, and that one needs the analysis of stochastic processes, such as Brownian motion and Lévy flights, to get a better grip on the market (for the definitive review, see Hull 2005). The number of non-physics uses is ever increasing, and we refer to the recent perspective of the role of statistical physics outside of physics by Stauffer (2004).

As said, in this thesis we focus mainly on applying our new transport method to solve linear transport problems. We will use it to solve the radiative transfer of photons emitted by the first stars in the Universe. It is worthwhile, however, to keep in mind these alternative applications of nonequilibrium statistical mechanics, because our transport method may prove its use in these fields too.

5.3

Graph Theory

Our method differs from Monte Carlo methods in the sense that the latter solve for random trajectories of particles on a fixed grid, whilst in our method the particles are forced to perform a walk on the edges of a random graph that is now a representation of the free path space, or collision space, of the medium. Thus, the stochastic element of the process that used to be only in the trajectories of the particles on a fixed grid is now also part of the graph itself. This is an essential difference between the usual Monte Carlo transport methods, and ours.

In this sense, our transport method bears some similarity to the analysis of *percolation* through random media (Broadbent & Hammersley 1957; Shante & Kirkpatrick 1971; Isichenko 1992; Sahimi 1993). In chemistry and material sciences, the movement and filtering of fluids through porous materials has been extensively studied. The resulting theory helped predict the percolation of petroleum through semi-porous rock, but also that of water through thin tissues. Percolation is different from, say, diffusion, in the sense that it is not the fluid itself that contains for the random motions, but it is the medium that enforces the random behavior, because it provides for only a limited number of transport channels. In the preceding three decades, the mathematical theory of percolation has brought new understanding to a broad range of topics in physics. It tackles the problem of percolation by representing the medium by a random graph that represents the possible transport channels through the medium. Hence, one tries to answer standard questions like if it is possible for fluid elements to leak from one point along the graph to any other. Principles from graph theory are used to answer these questions.

In these methods, as in ours, a suitable random graph is constructed, and the relevant motions are defined as a random walk on the resultant graph. Others recognized the potential of this approach too. In radiative transfer, people realized that one can use percolation theory to analyze the transport of photons through clumpy and fractal media (Varosi & Dwek 1999). A random graph representation of communication channels is widely used in network theory, used to model the Internet, airline routes, and electric power grids (see e.g. Callaway et al.

2000), but also human languages (Masucci & Rodgers 2006). Closer to home is the work by Schnakenberg (1976) who formulated a network theory for Markovian stochastic processes. Each Master Equation (cf. Sect. 2.5.1) has an associated graph, and cycles on the graph represent different possible transitions. Except for the last example, the random graphs constructed have channels that are some sort of representation of the separation between two points. Via our criterion Eq.(4.6), this distance now has a different meaning, Eq.(4.8). The random walk along the graph, normally only representing a movement from one point to the next, can now be incorporated to solve for interaction terms. Thus, using this extra physical aspect of the edges of the graph, the ordinary random walks along networks can easily be extended to solve for linear transport processes. This extension, incorporated in our new method, may have its use in the aforementioned cases. For example, when a random graph is used to model the Internet, the random walk, or percolation, along the graph will then not only represent the movement of data packets from one server to the next, but it could also automatically incorporate interaction terms, such as data loss, via recipes similar to the ones described in Chapter 4.

Percolation theory has advanced greatly by the use of available theorems from graph theory. Because, in essence, our method is closely similar, we briefly point out here some interesting results from graph theory that might prove useful for our transport method. Most essential is the theory of random walks on graphs. These are so basic that there is a wealth of mathematical machinery available, some of which are nicely bundled in the compact review by Lovasz (1996). Some practical results are that one can analyze a Markov random walk using matrices, in which every entry represents the probability of transition from one node to another. Spectral analysis gives possible equilibrium solutions as eigenvectors. Note the similarities with using spectral theory to analyze Master Equations (cf. Sect. 2.5.1).

These results are valid for the general concept of random graphs. As explained in previous chapters, we use a special version hereof, the Delaunay graph. Delaunay graphs have several extra properties that can be analyzed using tools such as Palm distribution theory (cf. Sect. 3.3.4), and the like. One nontrivial result, already covered in Sect. 3.3.3, is that a subset of the Delaunay graph coincides with the minimal spanning tree of the set of vertices. Another interesting result, relevant for our transport method, concerns recurrence of random walks. For deterministic lattices, such as (infinite) square grids, it was proven quite some time ago that the probability that a simple random walk in \mathbb{R}^d returns to the origin is unity for $d = 1$ and $d = 2$, but smaller than unity for¹ $d \geq 3$. Moreover, Polya (1921) proved that for $d = 1$ and $d = 2$, we have *recurrence*, in that we return to the starting point an infinite number of times, whilst for $d \geq 3$, we have *transience*, in that we return only a finite number of times. Other things studied include the number of steps needed to return to the origin, the number of steps needed to reach a certain node, and the number of steps needed to have reached all nodes (for a review, see e.g. Thomassen 1990; Doyle & Snell 2000). Given that we defined a scattering process as a random walk on our Delaunay graph (cf. Sect. 4.4.2), it would be very worthwhile to have similar results for nondeterministic graphs. It has turned out not to be very trivial to extend these results to random graphs, however. A noteworthy result was obtained by Addario-Berry & Sarkar (2006), who showed that the simple random walk on a Poisson Delaunay graph is recurrent for \mathbb{R} and \mathbb{R}^2 , and transient for higher dimensional space.

We conclude this section by noting that the relation between our method to percolation

¹The famous Polya's constants, $p(d)$, are defined as the recurrence probability in \mathbb{R}^d . As said $p(1) = p(2) = 1$, whilst $p(3) = 0.341$, $p(4) = 0.193$, etc.

and graph theory is indeed very intimate, and that it may be worthwhile to further deepen our understanding of our method by further analysis from those two different perspectives. It would not be surprising that, in that process, one might find further use of our method in either fields.

5.4

Cellular Automata

When one observes, describes and studies Nature, one of the striking things one will find is that the most complex things in the Universe have simple rules at their basis. A culmination hereof is found in statistical physics, in which intricately complex turbulent fluid flow can in principle be described on the microscopic level by the relatively simple laws of Newtonian dynamics. This is a somewhat pervasive dogma of modern day science: no matter how complex the emergent behavior, the system can always be reduced to more tractable, and simpler components, that often seem easier to stomach. As is the case with, for example, statistical physics, it is however almost always practically impossible to resolve the actual microscopic dynamics. As explained in Chapter 2, one can circumvent this problem by coarse graining, and finding a mesoscopic, or kinetic, description of the problem, which hopefully becomes much more tractable. One can also be a bit more bold, and take the point of view that one should not care about the exact microscopic details, and should just assume some simple mesoscopic model, so long as the emergent macroscopic behavior is satisfactory.

The simplest such approach is the one of *cellular automata*. A cellular automaton is a discrete model, consisting of an infinite, regular grid of cells, each one of which can have a finite number of states. Time is discrete, and the state of a cell at time t is a function of the states of a finite number of neighbors of the cell at time $t - 1$. In the rule, neighbors do not change. Most important is that every cell has the same global rule for updating. All this makes cellular automata methods ideal for parallelization. These rules can take a wide variety of forms, depending on the kind of problem one is trying to solve. First used by John von Neumann to study self-replicating systems, and widely popularized by John Conway's *Game of Life* (Gardner 1970), it has had numerous of uses in computability theory, mathematics, theoretical social sciences, theoretical biology and, of course, physics (for a review, see Farmer et al. 1984; Toffoli & Margolus 1987; Gutowitz 1990). In hydrodynamics, for example, the governing equations are expressions for the conservation of mass, momentum and energy (cf. Sect. 2.4.2). Hence, one can choose to describe the fluid at a mesoscopic level using cellular automata that have interaction rules that reflect these conservation laws. No reference to the actual underlying microscopic process is made. The emergent global behavior of the automata, however, turns out to be very fluid-like, which is not surprising given the local input rules. This approach has been the basis for the popular *Lattice Boltzmann Methods*, which we will discuss in the next subsection. Except for doing hydro, it has also been used to simulate, for example, magnetic spin systems (Vichniac 1984; Creutz 1986), crystal growth



Fig. 5.1: Patterns of the *Conus Textile* seashell are generated by natural cellular automata. Pigment is exchanged between neighboring cells via local rules. Picture: Richard Ling

(Mackay 1976; Packard 1986), and even galaxy formation (Gerola & Seiden 1978). Common to all these modeling applications is the belief that cellular automata can capture essential features of physical systems in which large scale behavior arises from the collective effect of large numbers of locally interacting simple components.

The connection between these cellular automata and our transport method is apparent. Our method consists of a transport process, in which the state of each point, or Voronoi cell, is updated via a set of global collision rules, with global interaction coefficients $\{c_i\}$. Those rules only take into account the state of the neighbors, i.e. the number of particles coming along each of the Delaunay edges. But the random lattice we use is not a valid basis for cellular automata in a strict sense. Standard definitions of the lattices used in these automata state that the number of neighbors of each cell should be a fixed global number. Random lattices are therefore excluded. Similar considerations concerning complications introduced by the use of only regular meshes, as given in Sect. 3.2, apply here. Therefore, the fixed-neighbor-number rule has been relaxed somewhat lately, linking random graphs to cellular automata, perturbing the original regular topology (Serra & Villani 2002). It was also explicitly shown that when one uses, instead of the usual rectangular lattices, the irregular Voronoi grid as a natural basis for the cellular automata modeling (Hegselmann et al. 2000), the global tendencies are unaltered, but that new, previously unresolved, features emerge (Flache & Hegselmann 2001). Our random lattices are now an accepted cellular automaton mesh.

Considering these striking similarities, it is worthwhile to further delve into the subject, identifying areas of overlap from which both our method, and the field of cellular automata might benefit. We immediately find something beneficial for our method: considering the global rules of our transport process, our process can be considered similar to a cellular automaton. It should be trivial to parallelize the method, just as it is for the regular automata. It might also be the other way around. We found that the Lattice Boltzmann community had been facing a problem, and could benefit from our transport method.

5.4.1 Lattice Boltzmann Methods

Isotropic Random Tessellations for Lattice Methods of Simulating Complex Flows

J. Ritzerveld and V. Icke

Physical Review Letters (2005), submitted

The preceding two decades have shown rapid advances in the development of mesoscopic particle methods for nonlinear fluid flow. Essential was the realization that, in order to simulate complex flows, one does not have to resort to the discretization of macroscopic partial differential equations. Instead the behavior of the fluid can be captured by following the trajectories of mesoscopic pseudo-particles. This work has culminated in the development of the Lattice Boltzmann (LB) methods (Benzi et al. 1992; Gladrow 2000; Succi 2001), which historically originated out of Lattice Gas Cellular Automata (LGCA) methods. These methods are didactically pleasing, because they simulate the fluid's microscopic behavior by coarse-graining it to a mesoscopic level, by which one can use standard procedures from kinetic theory (e.g. Chapman-Enskog expansions, Sect. 2.4.2) to establish that in the macroscopic limit Navier-Stokes-like equations emerge.

In these lattice methods, the pseudo-particles are confined to a regular lattice, which represents a fully discrete phase space, along which they propagate from one point to the

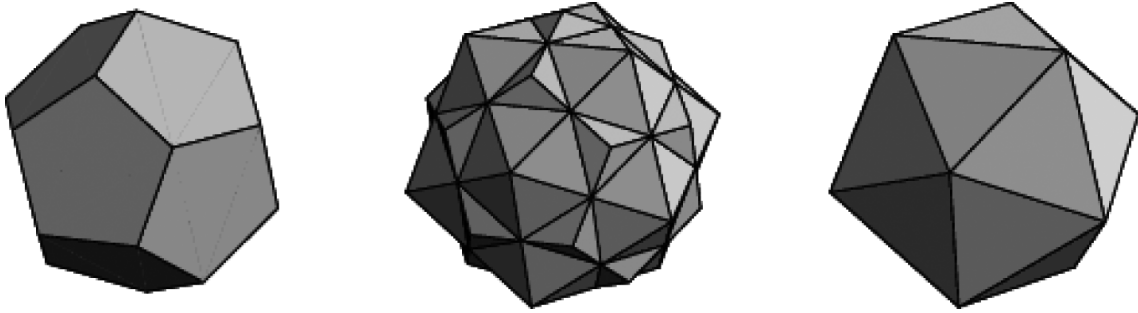


Fig. 5.3: The two Platonic solids with enough rotational symmetry to obtain Navier-Stokes-like equations in the limit, the icosahedron (left), and the dodecahedron (right), are each others dual (center).

next. The collisional terms on the right hand side of the Boltzmann equation are incorporated by imposing local collision rules at each lattice point. The detailed form of these collision rules depends strongly on the type of interactions represented by the collision operator, but in general they are chosen to obey conservation of mass, momentum and energy, to ensure the conservative nature of the emergent macroscopic picture.

Isotropy

In the absence of external potentials which superimpose a preferential direction, macroscopic differential equations describing the flow of some continuum are invariant under rotations, by which angular momentum is conserved. This condition should also constrain the mesoscopic pseudo-particle flow. The original lattice methods used uniform, regular lattices, which pose a severe restriction on the suitable regular lattices. In order to conserve isotropy, the discrete rotation group of the lattice should be a subgroup of the full continuous rotation group and have a large enough symmetry. More specifically, hydrodynamic behavior imposes that the lattice tensor of rank n , defined as (Wolfram 1986; Gladrow 2000)

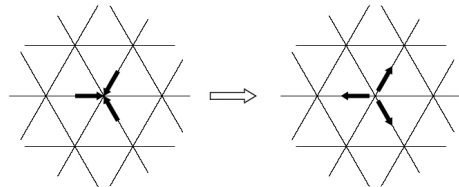


Fig. 5.2: FHP lattice gas model. Particles are redistributed at each node according to momentum conserving collision rules.

$$L_{\alpha_1 \alpha_2 \dots \alpha_n} = \sum_i c_{i\alpha_1} c_{i\alpha_2} \dots c_{i\alpha_n}, \quad (5.1)$$

in which $c_{i\alpha_\nu}$ are the cartesian components of the lattice velocities c_i , should be isotropic for $n \leq 4$. The simplest tessellating polygon satisfying these conditions in \mathbb{R}^2 is the hexagon, used extensively in the FHP lattice method (Frisch et al. 1986), depicted in Fig. 5.2.

Unfortunately, when trying to extend this to \mathbb{R}^3 , one finds that the only Platonic solids satisfying the tensor conditions, the icosahedron and the dodecahedron, shown in Fig. 5.3, do not tessellate three-dimensional space. This problem has been circumvented by using 4D face-centered hypercube (FCHC) methods (Frisch et al. 1987), or by introducing multi-speed models (Qian et al. 1992), in which different directions may have different speeds and/or different weights.

Recent developments in lattice methods for fluid flow (e.g Peng et al. 1998; Ubertini & Succi 2005) show a shift from stiff, regular lattices to the use of unstructured grids, which may adapt to a set of physical variables. Each of these unstructured cells are, in general, not (point-)symmetric. Thus, in order to conserve momentum locally, interpolation schemes or an extra set of local constraints (Karlin et al. 1999) are introduced into the initial ‘simple’ set of collision rules, thus departing from the elegance of a globally invariant set of rules.

Random Lattice

As an alternative and partial remedy to what has been mentioned before, we can immediately point out that this problem can be solved by the use of the random lattice used in our transport method. This lattice can just as easily be used within the original LGCA methods as in the more contemporary LB methods. Most importantly, it is isotropic by definition in any dimension, removing the necessity to move to \mathbb{R}^4 in single-speed models. As mentioned in Sect. 3.3.4, the average number of edges of a Voronoi cell in \mathbb{R}^2 , and thus the expected number of Delaunay lines meeting at one vertex, is equal to 6. Thus, on average, the Poisson-Delaunay lattice is equivalent to the hexagonal lattice used in FHP models. In \mathbb{R}^3 , however, the number of Delaunay lines meeting at one vertex can be derived to be $(48\pi^2/35) + 2 \approx 15.54$. The fact that this number is not an even integer is a direct consequence of the fact that there is no tessellating Platonic solid in \mathbb{R}^3 which retains enough rotational symmetry. Of course, every individual Voronoi cell will almost always be asymmetric, but on average it is symmetric, ensuring a conservation of momentum in the average. Thus, we have a lattice which tessellates every dimension $d \geq 1$, and is rotationally symmetric (for $d \geq 2$), which makes it ideal for use within these fluid lattice methods.

Modified Collision Rules

As already mentioned, a Voronoi cell in \mathbb{R}^2 will almost surely not be an exact hexagon, but will show asymmetries. Without loss of generality, we take as an illustration an FHP model, and see how the collision rules should be modified. What follows is very similar to our discussion of ballistic transport in Sect. 4.4.1. See, for example, Fig. 5.4, which depicts one vertex at which several Delaunay lines meet. In a simple FHP model, collision rules would dictate that two particles moving along lines I and II would continue along the dashed line. Because of the asymmetry of this particular Voronoi cell, there is no line along this direction. Instead, if we now choose the ‘nearest’ line III as the outgoing line, in which nearest can be defined in several ways, the most intuitive being that line for which the inner product with the dashed line is smallest, it can be shown that momentum is conserved on average (cf. Sect. 4.B), which immediately follows from the fact that the Voronoi cell is axisymmetric with respect to every incoming Delaunay line. However, as described in the previous chapter, from a numerical implementational point of view, it is more elegant to split up the two particles and distribute one of them along line III and the other along line IV, being the next to most straightforward path. More generally, if the transported quantity is a continuous entity, we have found it is

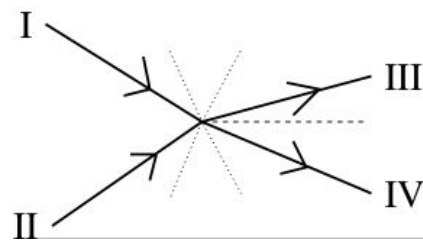


Fig. 5.4: Modified FHP rule. Two incoming particles (packets) along lines I and II collide, and should continue along dashed line to conserve momentum. Choosing lines III and IV as outgoing ensures conservation of momentum.

most efficient to split this quantity into d equal parts, in which d is the dimension of our computational domain, and distribute each of these parts along the d most straightforward paths. Because by definition, the Poisson-Voronoi cells are axisymmetric with respect to every associated Delaunay line in every dimension $d \geq 2$, we know that a similarly modified set of collision rules will ensure conservation of momentum.

Advantages of Random Lattices

We would like to stress that the only thing we try to demonstrate is the fact that for lattice methods for fluid flow (LGCA, LB, LBGK, etc.) the commonly used regular lattices can be replaced by the aforementioned Poisson Delaunay lattice, which has all the desired properties needed for lattice methods. Having shown that the lattice is isotropic in the mean by definition and that it conserves momentum (mass is conserved by definition of the collision rules), we proved that there is no need to change the collision rules used in a normal lattice method, when moving from a regular grid to this one. The only change introduced is the pre-processing step of determining which are the outgoing lines per site for each of the incoming ones. The lattice will ‘behave’ the same, on average, as the original regular one. Moreover, the use of such a lattice is better than using a regular lattice, in several ways. Because the lattice is isotropic and inherently stochastic, no spurious symmetries (and, therefore, no artificial conservation laws) occur. No unphysical features are introduced associated with a finite angular sampling of phase space: the unit direction vectors of the Delaunay lines will continuously fill the unit sphere, in the limit of the number of points $N \rightarrow \infty$, in contrast with, for example, the six angular components in an FHP model. This fact has as a result that the statistical noise inherent to particle based methods is smeared out over a continuous spectrum of directions instead of accumulated on a discrete set. In other words, the statistical noise of the system is directly visible in the plot. Last but not least, our Poisson Delaunay recipe produces a lattice which conserves isotropy on a macroscopic level, and fills the whole space without any overlap at the same time. This is even true for \mathbb{R}^3 , in which there is no sufficiently isotropic tessellating Platonic solid, removing the need to move to the 4D FCHC lattice or introduce multi-speed models. And all of this, without the need for modifications (interpolation scheme or an extra set of constraints) of the collision rules for a regular lattice.

Suggestions

Another quite severe problem for LB methods, is that the regular meshes not only break rotational symmetry, but the fixed cell widths also break Galilean symmetry (Succi 2001). The random lattices described in Chapter 3 are motion-invariant, and so retain both rotation, and Galilean invariance.

The fixed widths of the cells also impose constraints on the Reynolds number² the method can resolve. The introduction of random lattices that adapt to the fluid properties, as used in our transport method, might resolve this issue, because the adaptation parameter is chosen in such a way that even very small mean free paths can be resolved. Larger Reynolds numbers ($Re \propto \lambda^{-1}$) can still be resolved. This might be considered a huge improvement over the non-adaptive LB methods, that only have a very limited Reynolds number range. The adaptation of the lattice might even be so extreme, that the lattice points almost overlap. This could

²The *Reynolds number* characterizes the relative importance of inertial and viscous forces in a flow, and is commonly defined as $Re \equiv nUL/\mu$, in which n is the density, U is the flow velocity, L is the length scale of the flow, and μ is the dynamic viscosity. Therefore, $Re^{-1} = Kn \propto \lambda/L$.

entail that it is possible for LB methods to resolve discontinuities and shocks, by which the solvers can possibly be extended into the supersonic regime. More on the accuracy of resolving sharp discontinuities with the use of point processes in the next chapter.

If the incorporation of adaptive random lattices, as used in our transport method, into existing LB methods proves successful, we have a mean to extend our linear transport method, as described in the previous chapter, into the nonlinear regime. An extra complexity is introduced with respect to the linear transport method of Chapter 4. The background medium has now become an actively participating entity, that will evolve according to the fluid flow. Accordingly, the adaptive random lattice constructed from the medium properties will now be constantly transforming, in contrast with the static Delaunay lattice used in linear transport methods. The continuous updating of our random graph is the computational price we have to pay for the incorporation of nonlinear collision terms.

5.5

Conclusions

Although our method was originally developed for doing radiative transfer, it has much more potential. In this chapter, we have discussed how our new transport method is, or could be, linked to subjects from statistical physics, percolation and graph theory, and cellular automata methods. We have given several possible applications in each of those fields, and a number of alternative techniques for analyzing our method from those different perspectives.

This chapter is far from complete, however. There are several possible applications that we will refrain from explaining in-depth here. Some of those could be linked to the fields covered in this chapter. As an example, we point out that, when our method could be linked to LGCA and LB methods, it could also possibly be used to define cryptographic schemes based on these methods (Chopard & Marconi 2006). Others could be of a different class altogether. In *Quantum Electrodynamics*, for example, the vertices of a *Feynman diagram* always have degree 3, in the sense that at a vertex always two fermions (e.g. electrons) and one boson (e.g. a photon) meet. Complicated Feynman diagrams, depicting many different, possible concurrent, interactions, can be interpreted as a connected graph that always has vertices of degree 3. Voronoi diagrams in \mathbb{R}^2 have the remarkable property that, except for degenerate cases, the number of Voronoi edges meeting at one vertex is always 3. Could we use this Voronoi graph as a means to effectively draw, or calculate, Feynman diagrams, of many different orders? Or, could it, or possibly its dual, be used to effectively compute path integrals?

Bibliography

- Addario-Berry, L. & Sarkar, A. 2006, The simple random walk on a Voronoi tiling, submitted
 Ball, P. 2003, *Complexus*, 1, 190
 Benzi, R., Succi, S., & Vergassola, M. 1992, *Physics Reports*, 222, 145
 Broadbent, S. R. & Hammersley, J. M. 1957, *Proc. Cambridge Philos. Soc.*, 53, 629
 Callaway, D. S., Newman, M. E. J., Strogatz, S. H., & Watts, D. J. 2000, *Physical Review Letters*, 85, 5468
 Chopard, B. & Marconi, S. 2006, ArXiv Physics e-prints, [nlin.CG/0504059](https://arxiv.org/abs/nlin.CG/0504059)

- Chowdhury, D., Ghosh, K., Majumdar, A., Sinha, S., & Stinchcombe, R. B. 1997, *Physica A Statistical Mechanics and its Applications*, 246, 471
- Creutz, M. 1986, *Ann. Phys.*, 167, 62
- Czirok, A. & Vicsek, T. 2000, *Physica A Statistical Mechanics and its Applications*, 281, 17
- Doyle, P. G. & Snell, J. L. 2000, *ArXiv Mathematics e-prints*, math/0001057
- Dragulescu, A. & Yakovenko, V. M. 2000, *European Physical Journal B*, 17, 723
- Farmer, D., Toffoli, T., & Wolfram, S. 1984, *Cellular Automata: Proceedings of an Interdisciplinary Workshop* (Amsterdam: North Holland)
- Flache, A. & Hegselmann, R. 2001, *Journal of Artificial Societies and Social Simulation*, 4
- Frisch, U., D'Humières, D., Hasslacher, B., et al. 1987, *Complex Syst.*, 1, 648
- Frisch, U., Hasslacher, B., & Pomeau, Y. 1986, *Phys. Rev. Lett.*, 56, 1505
- Gardner, M. 1970, *Sci. Am.*, 223, 120
- Gerola, H. & Seiden, P. 1978, *Astrophysical Journal*, 223, 129
- Gladrow, D. W. 2000, *Lattice Gas and Lattice Boltzmann Methods* (Springer-Verlag, Berlin)
- Gutowitz, H. A. 1990, *Cellular Automata* (Cambridge, MA: MIT Press)
- Hammel, C. & Paul, W. B. 2002, *Physica A Statistical Mechanics and its Applications*, 313, 640
- Hegselmann, R., Flache, A., & Möller, V. 2000, *Cellular Automata as a Modelling Tool: Solidarity and Opinion Formation, Tools and Techniques for Social Science Simulation* (Heidelberg: Physica), 151–178
- Helbing, D., Farkas, I., & Vicsek, T. 2000, *Nature*, 407, 487
- Hull, J. C. 2005, *Options, Futures and Other Derivatives*, 6th edn. (Upper Saddle River, N.J.: Prentice-Hall)
- Isichenko, M. B. 1992, *Reviews of Modern Physics*, 64, 961
- Karlin, I. V., Succi, S., & Orszag, S. 1999, *Phys. Rev. Lett.*, 82, 5245
- Lovasz, L. 1996, *Combinatorics, Paul Erdős is Eighty, Vol. 2, Random walks on graphs: A survey* (Budapest), 353–398
- Mackay, A. 1976, *Phys. Bul. I*, 27, 495
- Masucci, A. P. & Rodgers, G. J. 2006, *Phys. Rev. E*, 74, 026102
- Packard, N. H. 1986, *Lattice models for solidification and aggregation, Proceedings of the First International Symposium for Science on Form* (KTK Scientific Publishers)
- Peng, G., Xi, H., Duncan, C., & Chou, S. 1998, *Phys. Rev. E*, 58, 4124
- Polya, G. 1921, *Math. Ann.*, 84, 149
- Qian, Y. H., D'Humières, D., & Lallemand, P. 1992, *Europhysics Letters*, 17, 479
- Sahimi, M. 1993, *Reviews of Modern Physics*, 65, 1393
- Schnakenberg, J. 1976, *Reviews of Modern Physics*, 48, 571
- Schulze, C. & Stauffer, D. 2005, *International Journal of Modern Physics C*, 16, 781
- Serra, R. & Villani, M. 2002, *Lect. Notes Comput. Sci.*, 2493, 168
- Shante, V. K. S. & Kirkpatrick, S. 1971, *Advances in Physics*, 20, 325
- Stauffer, D. 2004, *Physica A Statistical Mechanics and its Applications*, 336, 1
- Succi, S. 2001, *The Lattice Boltzmann Equation for Fluid Dynamics and Beyond* (Oxford University Press, Oxford)
- Thomassen, C. 1990, *J. Comb. Theory*, 49, 87
- Toffoli, T. & Margolus, N. 1987, *Cellular Automata Machines: A new environment for modeling* (Cambridge, MA: MIT Press)
- Ubertini, S. & Succi, S. 2005, *Progress in Computational Fluid Dynamics*, 5, 85
- Varosi, F. & Dwek, E. 1999, *ArXiv Astrophysics e-prints*, astro-ph/9905291
- Vichniac, G. Y. 1984, *Physica D*, 10, 96
- Wolfram, S. 1986, *J. Stat. Phys.*, 45, 471

Sampling Theorems for Spatial Point Processes

In this chapter, we discuss a final ingredient of our transport method, namely the sampling accuracy of the adaptive random lattice. Standard Nyquist sampling theorems apply when used in conjunction with regular meshes. We set up similar criteria for spatial point processes, and we derive a quantitative measure for the total number of points needed to correctly sample the local fluctuating behavior of the to be sampled functions. We will show that stochastic point processes do not show aliasing effects, when functions are undersampled. It might well be that the required resolution is beyond the limitations of our desktop computer, which is why we introduce an alternative way of increasing the method's sampling resolution. The (generalized) Poisson point process is ergodic, by which several instances of the same point process can be used to increase the overall resolution.

6.1

Introduction

This chapter concludes the first part of this thesis, devoted to the description of a new transport method. As is obvious from the somewhat lengthy Chapter 3, our method is rather unusual in the sense that it does not use the familiar regular meshes, but is defined to operate on adaptive random lattices. Random lattices stay much closer to the actual physical problem, and are, as such, much less prone to grid-based errors. It is therefore not surprising that random lattices are increasingly becoming *the* natural first choice for computational meshes. Still, these lattices are not as established as the archetypical rectangular meshes, whence it is necessary to point out several mesh related issues, that would have been considered trivial for regular lattices.

The keyword there is *discretization*. Otherwise continuous operators, state variables, and time evolution, have to be pixelated in order to feed to the CPU's logical gates the necessarily discrete instructions. Given the continuous nature of the physical problem at hand, the discretization procedure introduces errors ab initio, and analysis in the form of error control is needed. Standard considerations for the discretization of partial differential equations on regular grids lead to several restrictions imposed on the discretization scheme which states that the original differential operators should be retrieved in the limit of infinite resolution (*consistency condition*), that noise from round-off errors should not propagate or grow, (*stability condition*), and that the solution of the method should converge towards the real solution for infinite resolution (*convergence condition*). There are several theorems that

state that the first two conditions automatically impose the latter. Our method does not solve the equations of transport theory by discretizing its governing differential equations, but simulates the transport process itself, so that different considerations have to be taken into account. The differential operators for advection and collision in the transport equation Eq.(4.1) have been replaced with a step along a graph, and an interaction at a vertex, respectively.

Thus, the keyword *here* is *sampling*. Our method solves for the propagation of particles through the free path space of the medium, in which no reference to differential equations is made. The free path space is ideally sampled in such a way that every possible free path, at every point in space, in every possible direction, is resolved. If this is accomplished, every possible trajectory of a particle can be traced, and every possible linear transport process can be solved for. This would be accomplished by distributing an infinite number of points, that would perfectly sample the space-dependent properties of the medium. Moreover, a graph should be constructed from these points, connecting each point to an infinite number of neighbors. These lines would then perfectly sample the unit sphere, and as such all possible trajectories emanating from that one point. The resultant graph, G_∞ , with an infinite number of vertices, each of which has vertex degree ∞ , would then be a perfect representation of the free path space for the interaction between particles and medium. The actual trajectory a particle would make along this graph would, of course, still depend on the type of particle, and interaction, but also the initial position, etc. Whatever the resultant trajectory's form, however, we know it can be represented by a path along our ideal graph G_∞ , which is in effect a graph representation of continuous space.

It is, however, not feasible to construct the graph G_∞ . We lack the required resolution to sample coordinate *and* connectivity space infinitely finely, and even if we could, the resultant transport method would become prohibitively slow. Not only does it scale with the number of points, cf. Eq.(4.13), the line lengths of the graph would be infinitely small too, by which the set of interaction coefficients $\{c_i\}$ would become proportionally small. It would then take a very large number of steps along the graph ($\propto 1/c_i$), before one mean free path is reached, and interaction finally takes place. We therefore need a finite version of G_∞ , that is coarse enough to make the algorithm computationally feasible, but accurate enough to resolve the relevant properties of the free path space.

As this finite representation, we have chosen the adaptive random lattice, a graph that is constructed by computing the Delaunay triangulation of the adaptive point set defined in Eq.(4.6). We have already discussed several issues considering the finite angular and length sampling of the resultant random graphs in Sect. 4.3. In this chapter, however, we shall focus solely on the sampling criteria for the point distribution Eq.(4.6). In Sect. 6.2, we will derive how this distribution should be chosen, so as to accurately resolve all the relevant details of the possibly inhomogeneous medium distribution. Or, alternatively, we shall work out the minimum resolution needed so that the resultant tessellation is still a fair representation of the actual continuous medium. If it so happens that the resolution available is less than what is needed on the basis of the sampling criteria of Sect. 6.2, we will give an estimate of the error made. Moreover, in the case of lacking resolution, there might be a way out. The ergodic property of the Poisson point process opens up other possibilities for increasing the resolution of our method, as will be the topic of Sect. 6.3.

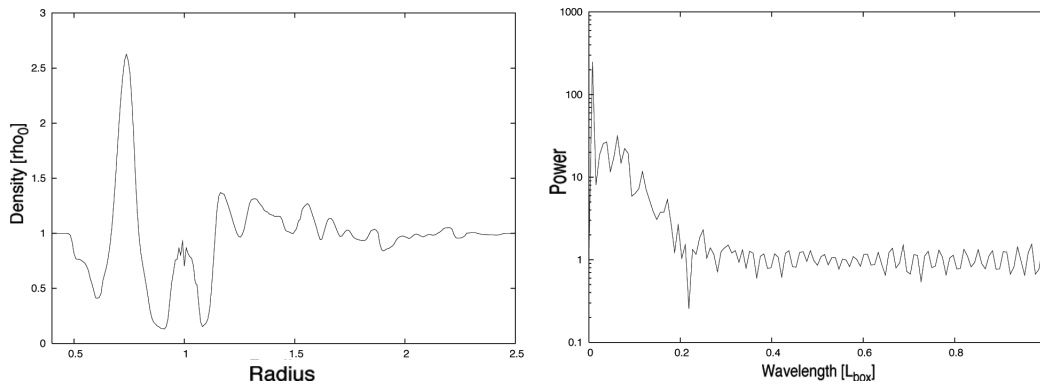


Fig. 6.1: Left: Radial density profile of a proto-planetary disk, as simulated by the *RODED* code (Paardekooper & Mellema 2006). A Jupiter mass planet is embedded at radius unity. Right: The power spectrum of the radial density profile, which shows that most energy is in the high frequency range.

6.2

Sampling Theorems

In order to be sure that our adaptive random lattice is a fair representation of the free path space of the medium, we need to know in advance that the underlying point process samples the medium properties accurately enough. In other words, we need to know that, when a particle moves from one point to the next along a line of the graph, the medium properties do not change too drastically. In all the linear transport examples we gave, we have been sampling the medium density $n(\mathbf{x})$, cf. Eq.(4.6) (for completeness, we point out that $n(\mathbf{x})$ is not necessarily a mass density, but could be a different function relevant to the specific transport phenomenon at hand). Thus, as we already briefly pointed out in Sect. 4.3, we need

$$\frac{1}{n(\mathbf{x})} \frac{\partial n(\mathbf{x})}{\partial x} < \frac{1}{\langle L \rangle}, \quad (6.1)$$

in which $\langle L \rangle$ is the average local line length, cf. Eq.(4.2), or a measure of the local point-to-point distance. Any adaptation of a point process, or a (resultant) mesh, should therefore be based on the gradient of the density, that tracks any spatial-dependent fluctuations of the medium. It merely represents the fact that, when the medium is homogeneous, we in principle need very few points, because the medium does not change when moving along a line between two points.

Eq.(6.1) is the basis for AMR (Berger & Olinger 1984; Berger & Colella 1989), in which a (regular) mesh is refined according to the gradient of a scalar material property, such as the density or the pressure. AMR is most often used in hydrodynamics, the governing equations of which have driving terms in the form of the gradient of the pressure, and not the pressure itself. Our method works by moving particles through a statistical space of free paths that scale inversely with the density. It is, therefore, better to *not* sample with respect to the gradient of the density, but with respect to the density itself. A result of this may be that the sampling of Eq.(4.6) is in conflict with the criterion Eq.(6.1). The point process may be optimal for sampling the free path space, and doing transport, but it can underresolve rapidly fluctuating parts of the medium.

This behavior is evident, when observing Fig. 6.1, left. This depicts a radial cut of a

proto-planetary disk that is the result of a planet-disk interaction simulation performed by the RODEO code (Paardekooper & Mellema 2006). A Jupiter-mass planet is embedded at radius unity, and the disk is perturbed. The resultant density profile shows much fluctuating behavior. If we would use our sampling criterion Eq.(4.6), most of the points available would end up in the relatively homogeneous tail end of the disk, whilst it is apparent that the fluctuations are most prominent around the location of the planet at radius unity. The point intensity should be highest here, via Eq.(6.1), but the density is actually lower than at the tail end, by which the intensity will be lower, via Eq.(4.6). Thus, we need to quantify the error associated with moving from gradient sampling to proportional sampling.

6.2.1 Nyquist Sampling

It is known from classical *information theory* that, if one tries to approximate a continuous function with only a finite set of values, errors are unavoidable. However, the nature of the errors that are introduced by the sampling can be directly controlled. Conventional methods use sampling techniques in the form of a regular array, or mesh, of points that approximate the ideal function. For this uniform sampling, the cell, or pixel, size determines the upper limit to the frequencies that can be displayed (Nyquist 1928; Shannon 1949). This limit, one cycle every two pixels, is called the *Nyquist limit*.

It is a well-known fact from sampling theory that, when our sampling rate is less than what is needed on the basis of the Fourier analysis, or, alternatively, when the highest frequency of the Fourier spectrum is greater than the Nyquist frequency, *aliasing* occurs. The unresolved frequencies above the Nyquist frequency will interfere with the ones below it, and a distorted sampled signal is the result. The inability to reproduce frequencies higher than the Nyquist frequency is inherent in a sampling process, but their appearance as aliases is merely a consequence of the regularity of the sampling mesh. Again, a testimony against the use of regular meshes.

A well known alternative to this regular sampling is the use of Poisson distributed sample points (Dippé & Wold 1985; Cook 1986). The energies in the frequencies exceeding the Nyquist frequency then do not cause aliasing effects, but appear as noise uniformly distributed over frequency space. This artifact is much less objectionable than aliasing. The human eye, for example, has only a limited number of photoreceptors, and it therefore also has a Nyquist limit. Yet, our eyes are not prone to aliasing effects, because the cells outside the fovea, where they are less tightly packed, are not distributed on a regular grid (Williams & Collier 1983). As another example, film grain also appears to have a random distribution (Thomas 1973), and, indeed, film is not known to alias. Thus, by choosing a (generalized) Poisson point process to sample the continuous medium distribution, we have automatically avoided any unwanted aliasing effects that are unavoidable with regular meshes.

Still, any computational implementation of our transport method will use a finite number of points, the method's resolution. Even though these points are distributed randomly, thereby avoiding aliasing (and the symmetry effects discussed in Chapter 3), there is still the issue of only being able to sample frequencies below the Nyquist limit. Given some arbitrary continuous and fluctuating function, defined on our computational domain, we need to find the minimum number of points needed to sample this function accurately.

6.2.2 Analysis

For regular lattices, the task of finding the right sampling frequency, or the ideal resolution, is relatively simple. Given a continuous function defined on a domain \mathcal{D} in \mathbb{R}^d , we can Fourier transform this function, and establish the highest frequency component. Consider, without loss of generality, the 1D case, in which we have a function $n(x)$, the Fourier transform of which is

$$N(f) = F[n(x)] = \int_{\mathcal{D}} n(x) e^{-2\pi i f x} dx, \quad (6.2)$$

in which we defined the frequency f . When $n(x)$ is *band limited*, there exists a highest frequency f_H , such that $N(f) = 0 \forall |f| > f_H$. The Nyquist-Shannon sampling theorem states that regular sampling with sampling intervals (average cell widths) Δx gives a complete representation of the original function, when the associated Nyquist frequency,

$$f_{\text{Nyq}} = \frac{1}{2\Delta x}, \quad (6.3)$$

is higher than the highest frequency component f_H , i.e.

$$f_H < f_{\text{Nyq}}. \quad (6.4)$$

As an example, we have computed the power spectrum of the radial profile of the proto-planetary disk in Fig. 6.1, left. The result is depicted in the same figure, right. It is obvious that the most energy is in the high frequency end, which reinforces our earlier claims of there being much detail around the position of the planet. A more thorough study of this graph would give the minimum resolution needed to resolve all these details.

For random lattices, this matter is a bit more tricky. It would be easy enough if we would use a non-adaptive homogeneous Poisson point process in every case, even if the medium is inhomogeneous. Similar considerations as for the regular lattices would then apply, the sampling interval Δx now being a measure of the average point-to-point distance ($\propto n_p^{-1/d}$). Our sampling criterion inhibits adaptation to the medium density, however, by which the relationship between the point density and the fluctuations in the medium may change from location to location. Thus, it becomes impossible to do a global analysis over the whole domain, as can be done for homogeneous lattices. Still, given the dichotomy between gradient (i.e. $n_p(\mathbf{x}) = f(\partial n(\mathbf{x})/\partial \mathbf{x})$), and direct (i.e. $n_p(\mathbf{x}) = f(n(\mathbf{x}))$) sampling, we need to work out how to ensure that the latter conforms to the resolution imposed by the first.

We therefore move from a global analysis to a local one, in which we try to relate the local sampling density, $n_p(\mathbf{x}) \propto n^d(\mathbf{x})$, to the local fluctuating behavior of the medium density $n(\mathbf{x})$. To accomplish this, we isolate a small patch $\mathcal{S} : |x^i - x_0^i| \leq \delta; i = 1, \dots, d$ around each point \mathbf{x}_0 , where the medium density has a profile $n(\mathbf{x})$. See, for example, Fig. 6.2. We can perform harmonic analysis of the function $n(\mathbf{x})$ on this small patch \mathcal{S} , and determine on the basis of this Fourier transform what the highest frequencies

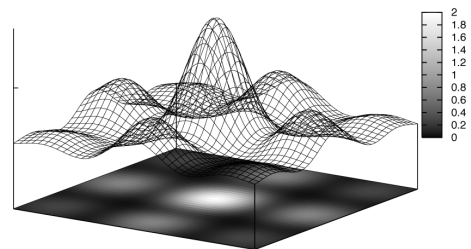


Fig. 6.2: Small patch \mathcal{S} of a 2D density distribution $n(\mathbf{x})$.

components are. Depending on the dimension of space, one can decompose this Fourier transform. In \mathbb{R}^3 , for example, one can decompose it into spherical harmonics, and Bessel

functions. Analysis of the high frequency behavior of both decompositions can give an estimate as to whether or not the local resolution is enough to resolve the local fluctuating behavior. If the resolution is lacking, we can estimate what the highest frequency component resolved. The sum of the amplitudes of the frequencies above this threshold gives an estimate on the sampling error.

The local resolution is determined by the sampling criterion Eq.(4.6), i.e. $n_p(\mathbf{x}) \propto n^d(\mathbf{x})$. Given the computational domain \mathcal{D} , and the total resolution N_p , we can compute the normalization constant, obtaining

$$n_p(\mathbf{x}) = N_p \frac{n^d(\mathbf{x})}{\int_{\mathcal{D}} n^d(\mathbf{x}) d\mathbf{x}}. \quad (6.5)$$

Given this local sampling density, the average point-to-point distance $\langle L \rangle$ is

$$\langle L \rangle \propto n_p^{-1/d}(\mathbf{x}) = n^{-1}(\mathbf{x}) \left(\frac{N_p}{\int_{\mathcal{D}} n^d(\mathbf{x}) d\mathbf{x}} \right)^{-1/d}. \quad (6.6)$$

More specifically, if we choose the Delaunay triangulation procedure to create a graph from these points, the point-to-point distances are defined as the average Delaunay line lengths, Eqs.(3.25) and (4.2),

$$\langle L \rangle = \zeta(1, d) n^{-1}(\mathbf{x}) \left(\frac{N_p}{\int_{\mathcal{D}} n^d(\mathbf{x}) d\mathbf{x}} \right)^{-1/d}. \quad (6.7)$$

Note that the average point-to-point distance scales linearly with $n^{-1}(\mathbf{x})$, proportional to the local mean free path (see Chapter 4). Thus, we can compare the local sampling length Eq.(6.7), and the resultant local Nyquist frequency, $f_{\text{Nyq}} = 1/2 \langle L \rangle$, with the results from the harmonic analysis to draw conclusions about the sampling accuracy. The sampling condition Eq.(6.4) then takes the form

$$f_H < \frac{1}{2 \langle L \rangle}, \quad (6.8)$$

or, alternatively, using Eq.(6.7),

$$f_H < \frac{n(\mathbf{x})}{2\zeta(1, d)} \left(\frac{N_p}{\int_{\mathcal{D}} n^d(\mathbf{x}) d\mathbf{x}} \right)^{1/d}. \quad (6.9)$$

Thus, the higher the energy of the local fluctuations ($\propto f_H$), the more sampling points, or resolution, N_p needed.

As an illustration hereof, consider the 1D example of the function

$$n(x) = \text{sinc}^2(x) = \frac{\sin^2(x)}{x^2}, \quad (6.10)$$

plotted in Fig. 6.3, left. Its Fourier transform can be derived analytically as

$$F[n(x)](k) = \frac{\pi}{2} [(k-2)H(k-2) + (k+2)H(k+2) - 2kH(k)], \quad (6.11)$$

in which $H(x)$ is the *Heaviside function*¹. The graph of Eq.(6.11) is plotted in Fig. 6.3, right. It is clear from this diagram that the Fourier transform of this one-dimensional density

¹The Heaviside step function $H(x)$ is defined as

$$H(x) = \begin{cases} 0 & x < 0 \\ 1/2 & x = 0 \\ 1 & x > 0. \end{cases}$$

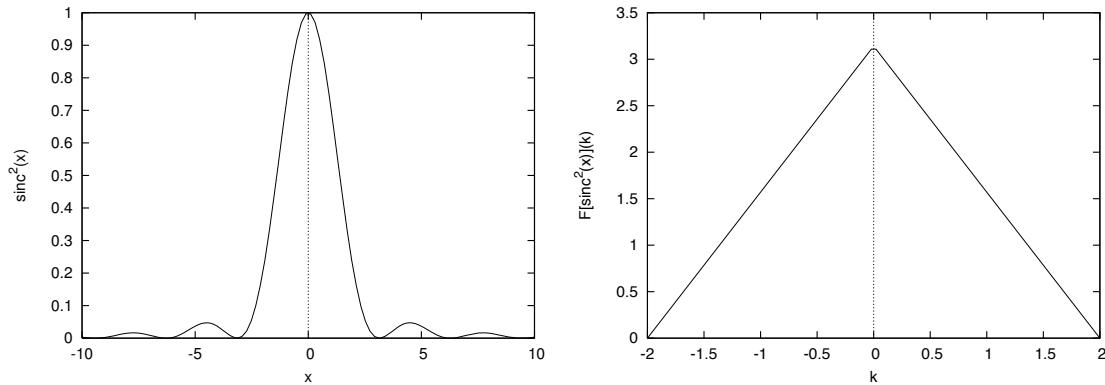


Fig. 6.3: Left: A 1D example of a medium density distribution function $n(x) = \text{sinc}^2(x)$. Right: Its Fourier transform, which is clearly bandlimited. Thus, only a finite sampling resolution is needed to recover the original density function.

function is bandlimited, with a maximum wavelength number $k = 2$. Thus, only a finite number of sampling points is needed to fully reconstruct the continuous density function, and one can immediately work out whether or not the available sampling resolution is satisfactory.

6.2.3 Remarks

It might, however, prove very difficult to do this analysis for every location \mathbf{x} , for different choices of the patch size parameter δ . We can, however, give a somewhat cruder sampling analysis. We can determine the lowest local point intensity n_p^{\min} that is reached at one or more locations inside the computational domain, and compare this to the highest frequency behavior of the harmonic analysis of the continuous function on the whole domain. If we know that this lowest intensity is sufficient, we know the point sampling intensity will be even better than necessary at other points. A somewhat less drastic approach would be to not compare the harmonic analysis to the lowest point intensity, but to the average point intensity $n_p^{\text{aver}} = N_p / \int_{\mathcal{D}} d\mathbf{x}$.

Another complication might be that the actual medium density distribution is not an analytically known function, but rather a d -dimensional output array from some simulation (as will be the case in Part III of this thesis). In that case, one needs to perform a numerical Fourier transform of the whole domain, or small patches \mathcal{S} of it, depending on what we are analyzing. This can easily be performed by using *Fast Fourier Transform* algorithms (Press et al. 1992), open source implementations of which are readily available on the Internet.

Unfortunately, there are also several medium distributions for which the sampling can be shown to be insufficient a priori, no matter what the resolution is. If we were to couple our linear transport method to nonlinear fluid dynamics solvers, it is to be expected that this will be the case very often. Fluid dynamics allows for shocks, in the supersonic flow regime (Landau

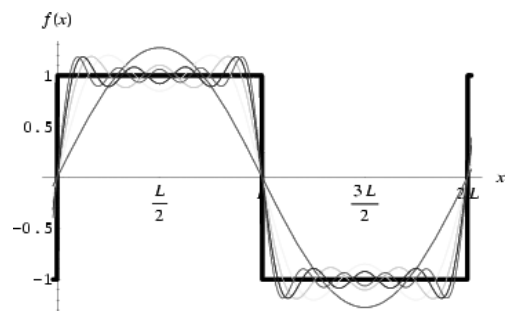


Fig. 6.4: Illustration of the Gibbs phenomenon near sharp discontinuities. Source: Weisstein (2003).

& Lifschitz 1987). This can happen for almost all astrophysical fluid flow. These shocks emerge as sharp discontinuities in the density, which can be described, in the mathematically idealized case, by Heaviside functions, $H(x)$. Take, for example, the square wave, as depicted in Fig. 6.4. The discontinuity at $x = L$ is typical for the density profile in front and behind a strong shock. This square wave has the form

$$f(x) = 2 [H(x/L) - H(x/L - 1)] - 1, \quad (6.12)$$

and can be written as a Fourier series,

$$f(x) = \frac{4}{\pi} \sum_{n=0}^{\infty} \frac{1}{2n+1} \sin \left[\frac{(2n+1)\pi x}{L} \right]. \quad (6.13)$$

Thus, this function is *not* bandlimited, and we would need an infinite number of points to sample this discontinuity correctly. This has been known quite some time as the *Gibbs phenomenon* (Gibbs 1899), which describes the overshoot, or ‘ringing’, of the finitely sampled function at these discontinuities (cf. Fig. 6.4). In Fourier analysis, this overshoot can be removed by adding the *Lanczos factor* (Lanczos 1956), but this does not help when dealing with point processes sampling these type of shocks. All particle based methods face this problem, and the only consolidation is that in reality most shocks do not have a sharp Heaviside-like form, but, for example, a somewhat softer $\tanh(x/\tau)$ form, which in the limit for $\tau \rightarrow 0$ approaches the Heaviside function. This form then only needs a finite number of sample points.

6.3

Ensemble Approach

In the previous section, we showed that we are able to do harmonic analysis, either analytically, or numerically, on a density function, or density array, the result of which is a local condition Eq.(6.9) for the sampling resolution. It might well be that the local density distribution is fluctuating so rapidly, that the sampling resolution is not satisfactory. The total number of points needed in order to fulfill the inequality might be so large that it is beyond our computational limit. It is, in that case, impossible to resolve the medium density profile as accurately as one may want, by which the overall transport process may be faulty.

There is a way out, however. The ergodicity of the Poisson point process, as pointed out in Sect. 3.3.1, is of great importance here. Given only a finite number of points, the algorithms for constructing a (generalized) Poisson point process satisfying Eq.(4.6) will only create one instance of all possible different arrangements. The full set of instances constitutes an ensemble that is homologous to the continuous distribution function describing the Poisson point process. Thus, given some finite resolution N_p that is not satisfactory, one can create several instances sequentially, and combine the results of the transport process on each instance in such a way that the overall result has higher effective resolution. This is schematically depicted in Fig. 6.5, in which an adaptive point process is used to sample a rather artificial density distribution that peaks at the unit circle. Only a small number of points is available, so we construct four different instances of the point distribution. The combination, or sum, of the four has much better resolution.

One can also state this differently. The Poisson point process is defined as being composed of points that have a point-to-point distance determined by an exponential distribution

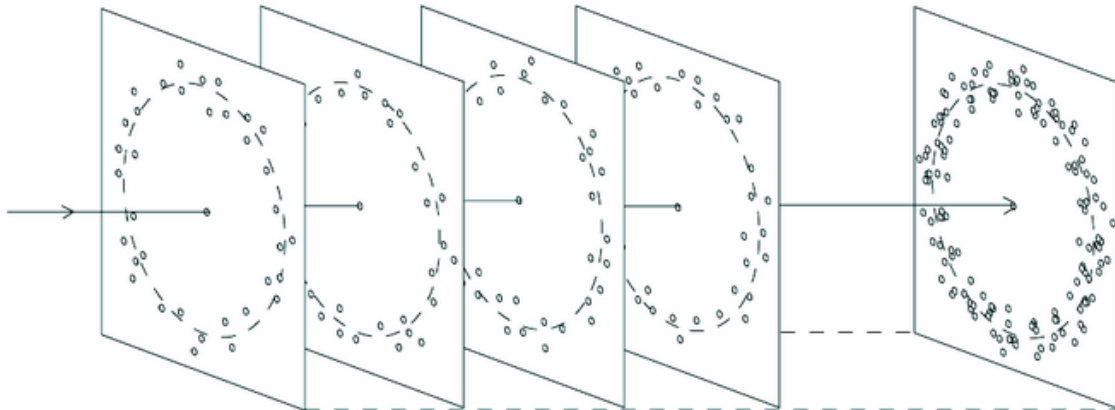


Fig. 6.5: Four instances of a density distribution that is peaked at the unit circle. The sum of the four makes up an overall result that has much better resolution.

function ($\propto e^{-n_p A}$), cf. Sect. 3.3.1. This entails that points are independent from each other, by which a *superposition principle* holds. This principle says that the superposition of two Poisson point processes with intensity n_p^1 and n_p^2 is again a Poisson point process, but now with intensity $n_p^1 + n_p^2$. This is obvious, because the probability that process 1 will occur in a small region dA is $n_p^1 dA$, and similarly $n_p^2 dA$ for process 2. This can also be extended to the inhomogeneous generalized Poisson point process, Eq.(4.6): the superposition of two generalized Poisson point processes, one with point distribution $n_p^1(\mathbf{x})$, and another with point distribution $n_p^2(\mathbf{x})$, is again a generalized Poisson point process, now with a point distribution density $n_p^1(\mathbf{x}) + n_p^2(\mathbf{x})$.

Thus, when the sampling criterion for random spatial point processes, Eq.(6.9), dictates that we need a certain number of points to accurately resolve local fluctuations, and when it turns out that this resolution can not be reached, because our maximum computational resolution is a factor α lower, we can solve this problem by creating not one, but α instances of the same point process $n_p(\mathbf{x})$. The superposition of these instances *does* satisfy the required sampling criterion. Moreover, as discussed in Sect. 4.3.3, we can use this ensemble approach not only to increase the spatial sampling, but also the very important angular sampling.

6.4

Conclusions

In this chapter, we have studied the consequences of using spatial point processes as a tool for sampling continuous density functions. The standard Nyquist sampling techniques, trivial for the uniform sampling of regular meshes, have been adapted to incorporate adaptive point processes, which have the fortunate side effect of avoiding aliasing effects. Globally valid sampling criteria for uniform meshes have been replaced by space-dependent local sampling criteria, that relate the local fluctuating behavior to the maximum number of available points. It might well be that these sampling criteria dictate that that number of points is not satisfactory. The solution is again one of the many advantageous properties of Poisson point processes. The superposition of several instances of the same Poisson point process is again a Poisson point process, but now with the required, higher, resolution.

All this is relevant for our transport method, described in Chapter 4, because we now have a quantitative measure to determine the accuracy of the result, something that was until now only available for uniform mesh-like sampling techniques. Not only that, we also have a means to increase the resolution, and the accuracy, of our method, without the need to buy an extra DIMM for our desktop machine. We conclude this chapter by emphasizing that the analysis in this chapter is as much relevant for other particle based methods, as it is for ours, whether these are digital sampling methods (Cook 1986), or particle based fluid dynamics solvers, such as DPD (Flekkøy et al. 2000) or SPH (Lucy 1977; Monaghan 1992).

Acknowledgements We thank Franco Maschietto for doing most of the inspiring work presented in this chapter, as part of his Master's research project. We are much obliged to Sijme-Jan Paardekooper for making available the density profile of the proto-planetary disk.

Bibliography

- Berger, M. J. & Colella, P. 1989, *J. Comput. Phys.*, 82, 64
 Berger, M. J. & Oliger, J. 1984, *J. Comput. Phys.*, 53, 484
 Cook, R. L. 1986, *ACM Trans. Graph.*, 5, 51
 Dippé, M. A. Z. & Wold, E. H. 1985, in *SIGGRAPH '85: Proceedings of the 12th annual conference on Computer graphics and interactive techniques* (New York, NY, USA: ACM Press), 69–78
 Flekkøy, E. G., Coveney, P. V., & de Fabritiis, G. 2000, *Phys. Rev. E*, 62, 2140
 Gibbs, J. W. 1899, *Nature*, 59, 200 and 606
 Lanczos, C. 1956, *Applied Analysis* (Princeton, NJ: Van Nostrand)
 Landau, L. D. & Lifschitz, E. M. 1987, *Fluid Dynamics* (Oxford: Pergamon)
 Lucy, L. B. 1977, *AJ*, 82, 1013
 Monaghan, J. J. 1992, *ARA&A*, 30, 543
 Nyquist, H. 1928, *Trans. AIEE*, 47, 617
 Paardekooper, S.-J. & Mellema, G. 2006, *A&A*, 450, 1203
 Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, *Numerical Recipes in C: The Art of Scientific Computing* (New York, NY, USA: Cambridge University Press)
 Shannon, C. E. 1949, *Proc. Institute of Radio Engineers*, 37, 10
 Thomas, W., ed. 1973, *SPSE Handbook of Photographic Science and Engineering* (New York: Wiley)
 Weisstein, E. W. 2003, Gibbs Phenomenon, from *MathWorld—A Wolfram Web Resource*
 Williams, D. R. & Collier, R. 1983, *Science*, 221, 385

Part II

Triangulating Radiation

Heavenly radiance fills the aether, its rays parallel and straight and, so long as nothing is there to interrupt them, invisible. The secret of God's creation are all told by those rays, but told in a language we do not understand, or even hear - the direction from which they shine, the spectrum of colors concealed within the light, these are all characters in a cryptogram.

ISAAC NEWTON
The Baroque Cycle
Neal Stephenson

Cosmological Radiative Transfer

In this chapter, we will give a brief introduction into the field of cosmological radiative transfer. We present the equations describing the propagation of ionizing photons in an expanding universe, whereafter we simplify these equations using a series of well-motivated approximations. We finish by giving a concise description of most of the cosmological radiative transfer codes on the market, pointing out their advantages and disadvantages along the way.

7.1

Introduction

In the first part of this thesis, we gave an overview of a new method that solves for transport processes with the use of adaptive random lattices. This was done in a wide perspective, placing emphasis not only on the method's general framework, but also on its versatility, and on its intrinsic connection to other parts of science. In the final two parts of this thesis, we narrow our field of view, focusing on just one specific transport process, namely that of the transport of photons through a possibly evolving background medium. All this will be done in a cosmological setting, propagating photons on the largest scales possible.

7.1.1 Radiative Transfer

Radiative transfer, as already briefly discussed in Sect. 2.4.1, describes the propagation of radiation through a medium that may absorb, scatter, or even reemit photons. As was to be expected, its governing equation is a transport equation, cf. Eq.(2.11), for a photon gas, a Boltzmann-like equation for the specific intensity $I_\nu(\mathbf{r}, \boldsymbol{\Omega}, t)$, to be defined in Sect. 7.2. After Maxwell's description of the dynamical behavior of electromagnetic radiation (Maxwell 1864, 1892), the first half of the 20th century led to a rapid development of the macroscopical treatment of radiation-matter interaction. First mainly dominated by phenomenological laws, the subject of radiative transfer soon grew out to be a mature, theoretically well-founded part of science. Its applications are very diverse, ranging from the description of heat transfer in engineering (Modest 1993), to that of photons trying to penetrate the solar atmosphere (Rutten 2005). The subject is now well-established, and, in an astrophysical context, the definitive treatises are Chandrasekhar (1950), Mihalas & Mihalas (1984), and Rybicki & Lightman (1986).

The preceding decade has shown rapid advances in the field of theoretical and observational cosmology, motivated by the precision cosmology experiments COBE, Boomerang,

and WMAP, and, more recently, by the newly found interest in the *Epoch of Reionization* (EOR). We refer the reader to Part III of this thesis, which is devoted solely to this heavily studied subject. In this context, *cosmological radiative transfer*, dealing with the propagation of ionizing photons from the smallest to the largest scales, has received considerable attention. This field is still under heavy development, and is far from being as mature as other fields in radiative transfer.

7.1.2 A Cosmological Setting

The EOR denotes a phase transition of the Early Universe. When the Universe had an age of about 400,000 years, its expansion had caused it to cool down to the point that the hydrogen recombination rate was higher than the ionization rate, by which almost all protons and electrons combined into neutral hydrogen, making the Universe opaque to ionizing radiation. It took quite some time before the initial density perturbations gave rise to the first generation of stars (dubbed *Population III*) and quasars, forming in high-density, filamentary-like regions. These first sources ended the "Dark Ages" by producing the first new supply of ionizing photons, blowing bubbles of ionized regions that grow until overlap, or percolation (cf. Sect. 5.3), occurs. A much more elaborate description of our current understanding of this period is given in Chapter 10.

The formation of these cosmic structures, such as stars and quasars, is almost certainly dominated by an intricate interplay between (magneto)hydrodynamics, gravity, and radiative transfer, on a cosmological background that sets the initial and boundary conditions. Thus, we must solve for all three parts of physics in order to accurately model this age. Of these, the cosmology is often assumed to be given, while the computation of gravitational potentials is rather well understood (e.g. Greengard 1988). Hydrodynamics must be three-dimensional for this purpose, and 3D hydro is beginning to enter its springtime: adaptive mesh refinement (AMR) and related methods are beginning to produce magnificent results (see e.g. LeVeque et al. 1998, for a review). However, radiative transfer techniques that combine true three-dimensionality with reasonable spectral resolution are, by comparison, the most primitive of the methods needed for realistic simulation of structure formation, and the Epoch of Reionization. Yet it seems essential that the physics of radiation be built in, because the energy budget of nascent structures is heavily influenced, indeed often dominated, by radiative effects. Serious models must therefore be three-dimensional, and spectral coverage must at least be good enough to cover hydrogen ionization and recombination. This makes solving the radiation part of this physical problem seven-dimensional: a daunting computational task.

In this thesis, we will always assume the gravitation and hydrodynamics part of any simulation to be given by some external solver which sets the conditions of the background medium. In this specific part of the thesis, we will focus solely on the radiative transfer aspect of the problem. In Part III, we will present results of simulations of the EOR, in which our radiative transfer approach has actually been coupled to hydro-solvers.

7.1.3 Overview

We will start by stating the relevant transfer equations in a cosmological context in Sect. 7.2. In reality, though, none of the existing cosmological radiative transfer codes solve these exact equations, because of their huge computational complexity. Instead, they solve some simplified version thereof. These versions are derived under several well-motivated assumptions, discussed in Sect. 7.3. Finally, we give a general overview of the classes of

methods existing today in Sect. 7.4. All this will pave the way for the introduction of a new cosmological radiative transfer code, `SimpleX`, an implementation of our new transport method, in the next chapter.

7.2

Cosmological Radiative Transfer Equations

The transfer equation needed to describe what happens during the EOR describes the propagation of ionizing radiation through an inhomogeneous density field. In a cosmological setting, the radiative transfer equations have to be set up to incorporate the expansion of the Universe. Following the derivation in Gnedin & Ostriker (1997), we present the radiative transfer equation in comoving coordinates.

We define $f(x_i, p_k, t)$ to be the distribution function for photons in comoving coordinates x_i , and comoving momentum

$$p_k = a \frac{h\nu}{c} \Omega_k, \quad (7.1)$$

in which we used the scale factor $a \equiv a(t)$, Planck's constant h , the speed of light c , the photon frequency ν , and the unit propagation direction vector Ω_k . The distribution function is the one used in the general transport equation Eq.(2.11), so that

$$N_\gamma = \int f(x_i, p_k, t) d^3x d^3p \quad (7.2)$$

is the number of photons in the Universe. The transport equation for this distribution function is

$$\frac{\partial f}{\partial t} + \dot{x}_i \frac{\partial f}{\partial x_i} + \dot{p}_k \frac{\partial f}{\partial p_k} = \left(\frac{\partial f}{\partial t} \right)_{\text{coll}} + s(x_i, p_k, t). \quad (7.3)$$

The rhs of Eq.(7.3) denotes the collision, or absorption, and the emission terms.

In radiative transfer it is customary, however, to use the specific intensity $I_\nu(x_i, \Omega_k, t)$ of radiation instead of the more general $f(x_i, p_k, t)$. General convention dictates that $I_\nu d\nu d\Omega dA dt$ is the energy of photons in the frequency range $[\nu; \nu + d\nu]$, passing in the time interval dt through an infinitesimal area dA that spans a solid angle $d\Omega$ around Ω_k . Using $a^3 d^3x = cdtdA$ and $d^3p = p^2 dp d\Omega$, we have the following transformation, cf. Eq.(2.15):

$$I_\nu = h\nu c f \frac{d^3x d^3p}{d\nu d\Omega dA dt} = \frac{h^4 \nu^3}{c} f. \quad (7.4)$$

Combining Eqs.(7.3) and (7.4), we obtain the equation of cosmological radiative transfer:

$$\frac{1}{c} \frac{\partial I_\nu}{\partial t} + \frac{\mathbf{\Omega} \cdot \nabla I_\nu}{\bar{a}} - \frac{H(t)}{c} \left(\nu \frac{dI_\nu}{d\nu} - 3I_\nu \right) = j_\nu - \alpha_\nu I_\nu, \quad (7.5)$$

which we now have written in vectorial form. In this equation, $H(t) \equiv \dot{a}/a$ is the time-dependent Hubble constant, and $\bar{a} = \frac{1+z_{\text{em}}}{1+z}$ is the ratio of cosmic scale factors between photon emission at frequency ν , and at the present time t . The terms on the rhs of Eq.(7.5), j_ν and α_ν , have their traditional meanings of source function and absorption coefficient, respectively. Eq.(7.5) resembles the standard equation of radiative transfer Eq.(2.17) closely, except for two modifications: (i) the denominator \bar{a} in the second term, which accounts for the changes in path length along a ray due to cosmic expansion, and (ii) the third term, which accounts for cosmological redshift and dilution, respectively.

7.3

Simplifications

The equation of cosmological radiative transfer in comoving coordinates, Eq.(7.5), is exact, in the sense that one can use this equation to solve for the specific intensity exactly, at every point, once the source function j_ν and the absorption coefficient α_ν are given. This problem is extremely complex, however. The specific intensity is a seven-dimensional entity (three positions, two angles, one frequency and time). Moreover, because of the huge range in density, realistic cosmological simulations require very high spatial and angular resolution. It is therefore mostly impractical to use this approach to do dynamic computations. Fortunately, when numerically solving for the propagation of ionizing photons during the EOR, the complexity of this problem can be reduced by a series of approximations.

7.3.1 Local Approximation

Following the steps taken in Norman et al. (1998) and Abel et al. (1999), we begin by eliminating the cosmological terms. That this can be done is obvious on physical grounds. The ionizing sources produce photons that have to reionize the neutral hydrogen, which is opaque to Lyman continuum photons. The transfer of these photons can therefore be described on a local scale. In other words, if λ is the photon's mean free path, and L the side length of the simulation box, we have by construction that $\lambda \ll L$.

Working this out more formally, the ratio of the third to the second term in Eq.(7.5) is $H(t)L\bar{a}/c \ll 1$, when the scale of interest, L , is much smaller than the horizon, $L_H = c/H(t)$. Thus, the third term can be safely ignored when $|\nu(\partial I_\nu/\partial \nu)| \leq I_\nu$, i.e. for continuum radiation. By the same condition $L \ll L_H$, we obtain $\bar{a} = 1$, and $\nu_{\text{em}} = \nu$. The formal equation Eq.(7.5) thus reduces to the familiar radiative transfer equation, Eq.(2.17),

$$\frac{1}{c} \frac{\partial I_\nu(\mathbf{r}, \boldsymbol{\Omega}, t)}{\partial t} + \boldsymbol{\Omega} \cdot \nabla I_\nu(\mathbf{r}, \boldsymbol{\Omega}, t) = j_\nu(\mathbf{r}, \boldsymbol{\Omega}, t) - \alpha_\nu(\mathbf{r}, t) I_\nu(\mathbf{r}, \boldsymbol{\Omega}, t), \quad (7.6)$$

where ν is now the instantaneous, comoving frequency.

A pitfall of this approach that is often overlooked is that these approximations fail, when the ionization bubbles start to overlap at the end of the EOR, when the mean free path of the photons increases drastically. This is mostly not a problem because the photons just leave the simulation box, but in more recent simulations, where the box sizes are ever increasing, the expansion terms may have to be incorporated.

7.3.2 Quasi-static Approximation

In most realistic reionization simulations, the absorption and emission coefficients change on time scales longer than the light crossing time, L/c . When this is the case, the time derivative in Eq.(7.6) can be dropped, which results in the reduced equation

$$\boldsymbol{\Omega} \cdot \nabla I_\nu(\mathbf{r}, \boldsymbol{\Omega}) = j_\nu(\mathbf{r}, \boldsymbol{\Omega}, t) - \alpha_\nu(\mathbf{r}) I_\nu(\mathbf{r}, \boldsymbol{\Omega}). \quad (7.7)$$

There are various numerical ways of solving this static transfer equation (see e.g. Mihalas & Mihalas 1984), depending on the symmetry of the problem and the properties of the absorption and emission coefficients.

In effect, this approximation amounts to finding the position, velocity and thickness of the ionization front of each source, at each small time step. Because of the assumed time-independence of the medium (which in this approximation means that the matter moves with speeds very much less than the speed of light), the actual propagation of photons during one time step is not solved for. The dynamical evolution of the problem is only incorporated by taking steps in time, resulting in expanding fronts. Ignoring the speed of the individual particles has as a result that this approximation can give rise to ionization fronts expanding faster than the speed of light. Fortunately, this only happens near the sources, where the neutral hydrogen fraction will always be very near zero, but we do not need to resolve the details there anyway.

Time dependent effects are accounted for by updating the local absorption coefficient according to the ionization state. Moreover, the ionized gas can recombine, producing diffuse photons that contribute to the local source function. Details hereof, including how our new cosmological radiative transfer method `SimpleX` implements this, are discussed in the next chapter.

7.3.3 Frequency Reduction

When a certain volume of space is reionized by the first stars or quasars, the emergent spectrum of these sources can by no means be described by a delta function peaking at the hydrogen ionization threshold of $E = 13.6$ eV or $\lambda = 91.2$ nm. In the case of the first stars, it will most probably something resembling a blackbody spectrum peaking above the ionization threshold; in the case of quasars, it will exhibit some power law behavior. The proper way of numerically dealing with this nontrivial frequency dependence of the emitted radiation would be to use very narrow bins to sample the spectrum. This extra dimensional dependence of the computational problem would make any method prohibitively expensive.

Fortunately, several physically well-motivated simplifications can be made (see Mihalas & Mihalas 1984, Chapter 6). A major simplification can be made by using a *mean opacity representation*, in which the spectrum is approximated by only one frequency, for example that of the hydrogen ionization threshold. The overall shape of the spectrum is incorporated by determining a spectrum-averaged absorption coefficient a_ν , which may be lower than the one near the Lyman limit. This is the approach we will use in our new transfer method, and we refer the reader to the next chapter for more details on the implementation.

A more elaborate simplification can be made, when helium reionization is also incorporated in the calculations. In that case, one can use one of the *multigroup methods*, in which the spectrum is divided into a number of frequency groups, each of which spans a definite range $[\nu_g; \nu_{g+1}]$. When helium is included, we could only consider three frequency groups above the ionization thresholds for HI, HeI, and HeII, at energies $E = 13.6$, 24.6, and 54.4 eV, respectively. None of the simulations presented in this thesis will include helium, however, which is why the mean opacity representation will suffice. Another multigroup method, presented in Petrosian et al. (1972) and Icke et al. (1980), in which the spectrum is divided into a Lyman continuum, Lyman alpha, and softer photons group, could prove useful when dust is included and the emergent spectrum has to be calculated.

7.3.4 Front Tracking

Solving the quasi-static cosmological radiative transfer Eq.(7.7) numerically can be quite costly, even when the frequency dependence has been dropped using the mean opacity rep-

resentation. Still, most modern cosmological radiative transfer codes have been designed to solve the modified radiative transfer equation, Eq.(7.7), directly or indirectly, using the type of numerical scheme described in the next section.

In cosmological radiative transfer simulations of the EOR, the subject of interest is the dynamical overlap of ionization fronts propagating outwards from each source. Thus, instead of trying to solve the radiative transfer equation, one may choose to take a more opportunistic approach, based on a drastically idealized treatment of the expansion of ionization fronts, in which the propagation of the front is described by balancing the number of ionizing photons emitted by the source against the number of recombinations within the HII region. This ionization balance equation is a jump condition for the ionization front (Spitzer 1978; Osterbrock 1989)

$$4\pi r^2 \frac{dr}{dt} = S_* + 4\pi\alpha_1(T) \int n^2(r')r'^2 dr - 4\pi\alpha_A(T) \int n^2(r')r'^2 dr', \quad (7.8)$$

in which r is the position of the front with respect to the source, and S_* is the number of ionizing photons (i.e. with frequencies $\nu \geq \nu_0$ above the Lyman limit threshold) emitted by the source per second. The second term on the rhs is the total number of diffuse photons produced by recombinations directly to the ground level $n = 1$, parametrized by the recombination coefficient $\alpha_1(T)$, given the HI density distribution $n(r)$. The last term is the total number of recombinations to every possible level, parametrized by the recombination coefficient $\alpha_A(T)$. The integrations are over the whole HII region. The so-called *Strömgren radius* is reached when the terms on the rhs of Eq.(7.8) cancel, i.e. when the number of recombinations within the HII region are exactly compensated for by the number of ionizing photons emitted by the source. Up to that point, the ionization front has a finite speed dr/dt . The jump condition Eq.(7.8) can be extended to include relativistic propagation speeds (Shapiro et al. 2005). For a given static density field $n(r)$ and a given source luminosity S_* , one can use this bookkeeping formalism to precompute for each location r the arrival time of the ionization front. This *front tracking* mechanism can thus be used to give a simplified picture of the movement of an ionization front through the neutral medium.

Unfortunately, there are several caveats to this approach. First and foremost, front tracking methods need to integrate Eq.(7.8) along radial rays emanating from the source. This becomes problematic when the diffuse recombination photons, the second term on the rhs of Eq.(7.8), are introduced into the problem. These photons have an isotropic source function, and so deviate from the aforementioned radial rays. This problem is often circumvented by using the so-called *on the spot approximation* (Baker & Menzel 1938; Spitzer 1978; Osterbrock 1989), but we show in Chapter 11 that this approximation is far from accurate. Second, Eq.(7.8) is an equation expressing a jump condition. Thus, the boundary of the HII region, i.e. the ionization front, will always be infinitely sharp. Physically realistic ionization fronts, and the asymptotically reached Strömgren sphere have a thickness that scales with the mean free path λ of the Lyman limit photons in neutral hydrogen. We prove in Appendix 7.A that this thickness is in fact of the order of 20λ . Front tracking methods will not resolve this feature, and only real radiative transfer will do this. Finally, there is the inherent difficulty of incorporating more than one source. The ionization balance equation Eq.(7.8) defines the front propagation around one particular source, around which the coordinates r are centered. Including more sources means that the overall ionization balance at each point can be influenced by any one of them. Unfortunately, it is not an option to store the arrival times for each front sequentially, solving Eq.(7.8) in turn for

each source. While in reality two or more sources would influence one patch of medium simultaneously, now the first source will have preference. The overall result may have severe unphysical features. This effect is enhanced, when we do not ignore the diffuse photons.

Although the front tracking mechanism is not very physical, in the sense that it does not resolve several of the distinguishing features of ionization bubbles, and might produce inaccurate models of their overlap, it does prove its use in method validation procedures. As we shall see in Chapters 9 and 11, Eq.(7.8) presents one of the few examples in cosmological radiative transfer for which analytical solutions are available.

7.4

Numerical Methods

The process of reionization is one of the most demanding in the field of astrophysics, considering the computational effort needed to give an accurate picture of the physics involved. Not only do we need accurate hydrodynamics solvers to solve for the nonlinear dynamics of the matter distribution over a very large range of densities, it is also mandatory to incorporate a full radiative transfer treatment. The dimensional complexity of the latter is what puts the largest constraints on what is computationally feasible, mostly because of the fact that the reionization process is highly inhomogeneous. The lack of any spatial symmetry prohibits any dimensional reduction of Eq.(7.5). Over the last decade, much effort has been devoted to the development of methods that solve Eq.(7.5) using one or more of the approximations listed in Sect. 7.3. In what follows, we will give a brief description of several classes of methods, mainly aimed at giving an introduction to our method, described in the next chapter.

As discussed, a good starting point for incorporating radiative transfer into realistic reionization simulations is finding a solver for the static transfer PDE Eq.(7.7). There are numerous schemes that are excellent at solving that transfer equation in one of the limiting opacity regimes. But in passing from one regime to the other most schemes fall short. A solver for the diffusion limit cannot solve the hyperbolic PDE, and vice versa. This is a severe limitation, because the inhomogeneous early Universe covers a wide range of opacities. Fortunately, there are several classes of methods that are, in principle, able to solve Eq.(7.7) for all opacities, given appropriate numerical resolution. All of these methods work by superimposing a mesh on the domain on which the transfer equation has to be solved. These meshes are mostly regular (the drawbacks hereof were discussed in Sect. 3.2), and may include adaptive refinement (Berger & Olinger 1984; Berger & Colella 1989).

7.4.1 Long Characteristics Methods

The most accurate of these methods is without doubt the *long characteristics method* (Mihalas & Mihalas 1984), which has many similarities with ray-casting methods. In this type of method a grid cell is connected to every other relevant grid cell, and the transfer equation is integrated along that 'ray'. We refer to Fig. 7.1, left, in which we give an example of how long characteristics can be used to determine the influence of photons being emitted at point A, near the border of a dense region I, on a point B, near the border of a dense region B. This type of method has the advantage that it incorporates the nonlocality of the transfer equation, by connecting every cell to every other. Thus it is able to solve Eq.(7.7) accurately for arbitrary density configurations. A practical disadvantage is apparent from Fig. 7.1: in

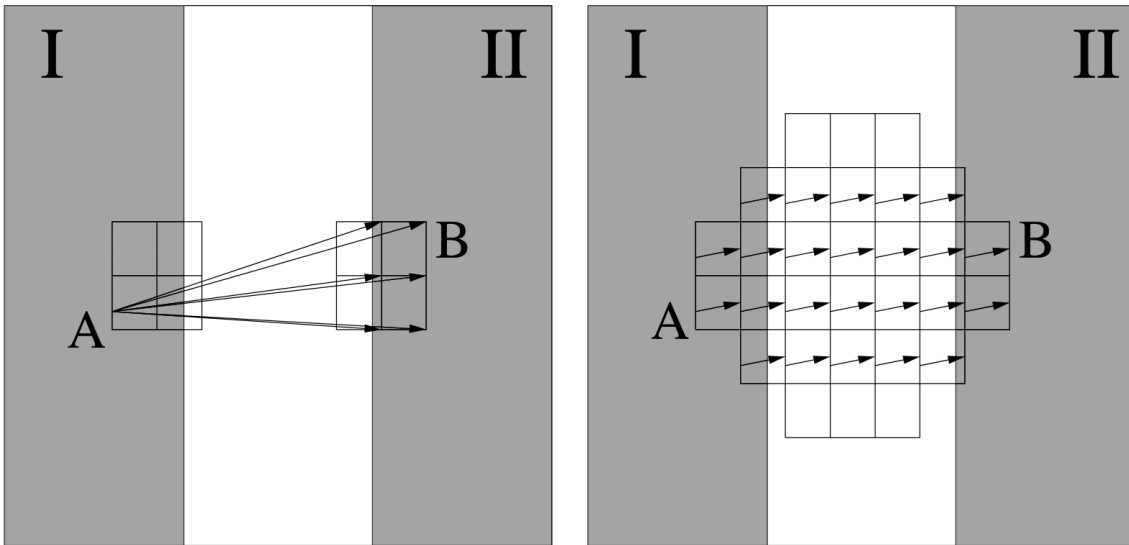


Fig. 7.1: Illustration of two different mechanisms for transporting radiation through a sparse region that separates two dense regions, I and II. Radiation emitted at point A, near the border of region I, eventually reaches point B, near the border of region II. Left is a schematic illustration of the long characteristic method; right is one of a short characteristic method.

the process of connecting all the cells, different characteristics cross the same cell multiple times. This introduces strong redundancy, which makes the method time-consuming.

A close relative of the long characteristics method, which is somewhat less expensive, is the well-known ray-casting approach. In this case, one does not draw a ray from a source to every other cell, but just a finite number of rays from every one source into several predefined directions. Those rays will cross many cells, until the domain boundary is reached. As with the long characteristics method, the transfer equation Eq.(7.7), or alternatively the jump condition Eq.(7.8), is integrated along each ray. Long characteristics methods automatically take care of this, but when using ray-casting methods one has to make sure explicitly that the angular resolution of the rays cast is sufficient to cover all border cells. The high angular resolution needed in realistic computations of the EOR often make these methods prohibitively slow.

Ray-casting methods are the most abundant amongst the current ensemble of cosmological radiative transfer methods. Which differential equation is solved along each ray may differ from method to method. Several of these methods are described in Abel et al. (1999), Razoumov & Scott (1999), Sokasian et al. (2001), and Cen (2002). The latter is somewhat eccentric, in the sense that it does not cast rays starting from the source, but that it does so at the receiving side.

7.4.2 Short Characteristics Methods

In an attempt to solve the redundancy problem of long characteristics methods, Kunasz & Auer (1988) introduced a *short characteristics method*. In this class of methods, the intensity in one grid cell is computed by connecting it to its neighboring cells only. This process is repeated for every angular direction Ω . We refer to Fig. 7.1, right, in which a short characteristics method is used to compute the propagation of photons emitted at A

in a certain fixed direction. As with the long characteristics and ray-casting methods, the transfer equation is solved one-dimensionally along the each line, using interpolation schemes to obtain the required grid variables.

This method has the advantage that it does not have any apparent redundancy. It does, on the other hand, require a very clever scheme to sweep the grid, in order to be sure that the intensities in all the neighboring grid cells are known at the right time. This is necessary because the emissivities may depend on the intensities, for example in the case of scattering. The physical values of the neighboring cells contribute via interpolations along the grid lines, which have to be quadratic or higher order in order to accurately reproduce second-order diffusion terms. The interpolation, intrinsic to the short characteristics method, introduces angular diffusion into the numerical solution, for example causing parallel laser beams to diverge in the downwind direction (see e.g. Steinacker et al. 2002). Kunasz & Auer (1988) showed that a parabolic interpolation reduces this intrinsic numerical diffusion, thereby obtaining a more accurate result, but not only does it make the algorithm more complex, requiring three upwind interpolation points, but it can also cause unphysical under- and overshoots of the interpolated quantities near discontinuities, possibly resulting in negative values of I_ν .

Although short characteristics methods have their numerical disadvantages, they are still among the most popular numerical radiative transfer schemes available, mainly because of a good trade-off between accuracy and speed. Cosmological radiative transfer methods that have successfully implemented this type of method are described in Nakamoto et al. (2001) and Mellema et al. (2006).

7.4.3 Hybrid Methods

Being aware of the computational cost of long characteristics methods on the one hand, and the problem of numerical diffusion problem of short characteristics on the other, much effort has gone into finding an intermediate solution that can combine the best of both worlds.

Jessee et al. (1998), for example, designed a transfer method for patch-based AMR that uses a *discrete ordinate method* (Chandrasekhar 1950). Abel & Wandelt (2002) designed a ray-casting method, in which every ray is split adaptively into sub-rays. Several cosmological transfer codes have implemented this scheme, amongst which Razoumov et al. (2002), and Razoumov & Cardall (2005). Juvela & Padoan (2005) implemented a ray-casting scheme for cell-based AMR, and Rijkhorst et al. (2006) developed a radiative transfer scheme, called *hybrid characteristics*, that uses a combination of both short and long characteristics to trace radiation along quad- and oct-tree type adaptive grids.

With the advent of ever more AMR type implementations of hydrodynamics solvers, and the need for the dynamical and self-consistent coupling between hydrodynamic and radiative transfer, the list of transfer methods using some type of adaptation will continue to grow.

7.4.4 Monte Carlo Methods

As can be inferred from our discussion in Sect. 2.6 on the widespread use of stochastic methods as an alternative to solving transport equations deterministically, it is not surprising that Monte Carlo methods are also a common tool in computational astrophysics. In cosmological radiative transfer, the most well-known implementation hereof has been described in Ciardi et al. (2001) and Maselli et al. (2003).

As with all stochastic transfer methods, one has to be aware of the fact that noise is introduced. This effect dictates the use of a large number of trials, because the stochastic noise only goes down with the square root of the number of trials. Especially when the resolution required increases, as is the case in high-resolution reionization simulations, the number of trials thus needed can increase drastically.

7.4.5 Moment Methods

Instead of trying to solve Eq.(7.6) directly, one can also try to find the same solution by considering moments thereof. The first three moments of the specific intensity $I_\nu(\mathbf{x}, \boldsymbol{\Omega}, t)$ with respect to the direction vector $\boldsymbol{\Omega}$ are (cf. Sect. 2.4.1) the energy density $J_\nu(\mathbf{r}, t)$, the flux vector $\mathbf{q}_\nu(\mathbf{r}, t)$, and the radiation pressure tensor $\mathbf{Q}_\nu(\mathbf{r}, t)$. The set of moment equations for the energy density and the flux vector is closed via the *Eddington tensor* \mathbf{h} , which relates the energy density to the pressure tensor,

$$Q_\nu^{ij} = J_\nu h_\nu^{ij}. \quad (7.9)$$

Gnedin & Abel (2001) have used this approach to design a computationally efficient cosmological radiative transfer code. In order to solve the system of moment equations, one has to specify the Eddington tensor \mathbf{h} that can, in principle, be computed once the sources and medium density distribution are known. In order to make this procedure tractable, Gnedin & Abel (2001) introduce the approximation that the optically thin Eddington tensor can be used, i.e.

$$h_\nu^{ij} = \frac{P_\nu^{ij}}{\text{Tr} P_\nu^{ij}}. \quad (7.10)$$

The resultant method conserves photons and flux, but may be sensitive to the introduction of numerical diffusion. It can also be rigorously shown to be correct only in the case of a single point source, and deviations from the correct solutions are seen in the case of multiple sources. On the other hand, it has the major advantage that in some cases it does not scale with the number of sources.

7.5

Conclusions

The list of methods and implementations presented in this chapter is representative for the current status of numerical cosmological radiative transfer. Even though the overall degree of approximation or the implementation details of each method may differ quite drastically, the results of each method should agree. This will be the subject of Chapter 9, in which most of the methods described in this chapter and our method described in the next will be compared to each other, and to the few configurations that have a closed analytical solution.

Even before this elaborate comparison project is described, we can already make several observations from what we have described. First of all, every one of the methods described in this chapter uses a structured mesh to discretize the medium properties (the absorption coefficient, and the like), and to define their operations on. Although we have already extensively discussed this in Sect. 3.2, we emphasize again that the use of structured meshes introduces unphysical behavior into the numerical solution, either through the introduction of spurious invariants, or the lack of rotational symmetry of the grid's discrete subgroup. Thus

we can claim a priori that structured grids, even if they are adaptive, are to be avoided. A notable exception, not described in this chapter, is the approach of Kessel-Deynet & Burkert (2000) and Susa (2006), that describe a radiative transfer method that is to be combined with SPH methods (Lucy 1977; Monaghan 1992). Another important observation to be made is that the operation count of all the methods in this chapter, with the exception of the methods of Gnedin & Abel (2001) and Cen (2002), scale with the number of sources. This is crucial, because this is precisely the reason why most methods become either very costly, or prohibitively slow, when applied to the EOR. During that epoch, the number of contributing sources may increase to millions, which is why most realistic simulations of the EOR have been running on massively parallel computers.

An essential ingredient for a realistic cosmological radiative transfer treatment of the EOR is the inclusion of chemistry effects. The relevant chemistry in this specific cosmological setting includes collisional and photoionization, hydrogen and helium recombination, but also photo-heating. These effects are described by a set of (coupled) rate equations that determine the overall ionization and temperature balance of the gas, once the radiation field is given. The incorporation of these rate equations is very specific to the code developed, so we have refrained from listing them here, and we refer the reader to the next chapter for any implementation details. Each one of the transfer codes listed in this chapter implements the rate equations by decoupling them from the transfer equation itself. At each time step, the static transfer equation Eq.(7.7) is solved, and the equilibrium position of the ionization front is found, given the properties of the medium (i.e. ionization state, chemical composition, and temperature). Hereafter, at the end of each time step, the rate equations are used to update the properties of the medium. There are transfer methods that solve the transfer and rate equations concurrently (see e.g. Wehrse et al. 2005), but in general, as long as the time steps are small enough, the overall results are shown to be correct, even with this decoupling. Therefore, we will choose this approach for including chemistry into our method.

In the current ensemble of cosmological radiative transfer methods, there is certainly room for improvement. Simulations of the EOR demand very high resolution on the one hand, and the ability to include a large number of sources on the other. All of the methods listed in this chapter scale with at least one of these two. Accordingly, the current state-of-the-art reionization simulations still do not have the resolution range needed to resolve all relevant scales. Thus, there is an obvious niche for cosmological radiative transfer methods that can improve on either scaling law. In the next chapter, we will present our new method which was designed to be optimal on both counts.

Bibliography

- Abel, T., Norman, M. L., & Madau, P. 1999, *ApJ*, 523, 66
 Abel, T. & Wandelt, B. D. 2002, *MNRAS*, 330, L53
 Baker, J. G. & Menzel, D. H. 1938, *ApJ*, 88, 52
 Berger, M. J. & Colella, P. 1989, *J. Comput. Phys.*, 82, 64
 Berger, M. J. & Olinger, J. 1984, *J. Comput. Phys.*, 53, 484
 Cen, R. 2002, *ApJS*, 141, 211
 Chandrasekhar, S. 1950, *Radiative Transfer* (London: Oxford University Press)
 Ciardi, B., Ferrara, A., Marri, S., & Raimondo, G. 2001, *MNRAS*, 324, 381
 Gnedin, N. Y. & Abel, T. 2001, *New Astronomy*, 6, 437

- Gnedin, N. Y. & Ostriker, J. P. 1997, *ApJ*, 486, 581
- Greengard, L. F. 1988, *The Rapid Evolution of Potential Fields in Particle Systems* (Cambridge, MA: MIT Press)
- Icke, V., Gatley, I., & Israel, F. P. 1980, *ApJ*, 236, 808
- Jessee, J. P., Fiveland, W. A., Howell, L. H., Colella, P., & Pember, R. B. 1998, *J. Comput. Phys.*, 139, 380
- Juvela, M. & Padoan, P. 2005, *ApJ*, 618, 744
- Kessel-Deynet, O. & Burkert, A. 2000, *MNRAS*, 315, 713
- Kunasz, P. B. & Auer, L. 1988, *JQSRT*, 39, 67
- LeVeque, R., Mihalas, D., Dorfi, E., & Müller, E. 1998, *Computational Methods for Astrophysical Fluid Flow* (Berlin: Springer-Verlag)
- Lucy, L. B. 1977, *AJ*, 82, 1013
- Maselli, A., Ferrara, A., & Ciardi, B. 2003, *MNRAS*, 345, 379
- Maxwell, J. C. 1864, *Scientific Papers*, 526
- . 1892, *Treatise on Electricity and Magnetism* (London: Oxford Univ. Press)
- Mellema, G., Iliiev, I. T., Alvarez, M. A., & Shapiro, P. R. 2006, *New Astronomy*, 11, 374
- Mihalas, D. & Mihalas, B. W. 1984, *Foundations of radiation hydrodynamics* (New York: Oxford University Press, 1984)
- Modest, M. F. 1993, *Radiative Heat Transfer* (New York: McGraw-Hill)
- Monaghan, J. J. 1992, *ARA&A*, 30, 543
- Nakamoto, T., Umemura, M., & Susa, H. 2001, *MNRAS*, 321, 593
- Norman, M. L., Paschos, P., & Abel, T. 1998, *Memorie della Societa Astronomica Italiana*, 69, 455
- Osterbrock, D. E. 1989, *Astrophysics of gaseous nebulae and active galactic nuclei* (Mill Valley, CA: University Science Books)
- Petrosian, V., Silk, J., & Field, G. B. 1972, *ApJ*, 177, L69+
- Razoumov, A. O. & Cardall, C. Y. 2005, *MNRAS*, 362, 1413
- Razoumov, A. O., Norman, M. L., Abel, T., & Scott, D. 2002, *ApJ*, 572, 695
- Razoumov, A. O. & Scott, D. 1999, *MNRAS*, 309, 287
- Rijkhorst, E.-J., Plewa, T., Dubey, A., & Mellema, G. 2006, *A&A*, 452, 907
- Rutten, R. J. 2005, *Radiative Transfer in Stellar Atmospheres* (University of Utrecht: Rob J. Rutten), www.astro.uu.nl/~rutten/
- Rybicki, G. B. & Lightman, A. P. 1986, *Radiative Processes in Astrophysics* (New York: Wiley-Interscience)
- Shapiro, P. R., Iliiev, I. T., Alvarez, M. A., & Scannapieco, E. 2005, *Relativistic Ionization Fronts*
- Sokasian, A., Abel, T., & Hernquist, L. E. 2001, *New Astronomy*, 6, 359
- Spitzer, L. 1978, *Physical processes in the interstellar medium* (New York: Wiley-Interscience)
- Steinacker, J., Hackert, R., Steinacker, A., & Bacmann, A. 2002, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 73, 557
- Susa, H. 2006, *PASJ*, 58, 445
- Wehrse, R., Wickramasinghe, D. T., & Davé, R. 2005, *ArXiv Astrophysics e-prints*, astro-ph/0507359

7.A

Ionization Front Thickness

The standard Strömgren sphere derivation uses the jump condition Eq.(7.8) to derive the radius of the HII region for which the front velocity dr/dt equals zero. The radius at which this occurs is called the Strömgren radius, denoted by R_S , and it delimits that region in which the number of ionizing photons emitted per second exactly compensates for the number of recombinations per second. Assuming, for now, that the hydrogen density distribution is

homogeneous, i.e. $n(r) = n$, we obtain

$$S_* = \frac{4}{3}\pi\alpha_B n^2 R_S^3, \quad (7.11)$$

in which we have defined the recombination coefficient $\alpha_B = \alpha_A - \alpha_1$, which is a measure for the number of recombinations to every level, except for the ground level. From this, we obtain an expression for the Strömngren radius

$$R_S = \left(\frac{3S_*}{4\pi\alpha_B n^2} \right)^{1/3}. \quad (7.12)$$

This derivation assumes that within the Strömngren sphere the ionization fraction $x = 1$, and that everywhere outside the fraction is $x = 0$. It therefore assumes that the Strömngren radius itself represents an infinitely thin transition layer from fully ionized to fully neutral. This is, of course, not the case, and we need a slightly more sophisticated derivation to determine the actual thickness of the front. Introducing J as the flux of ionizing photons from the star at distance r , we have

$$J = \frac{S_*}{4\pi r^2}. \quad (7.13)$$

In a small shell of thickness dr at radius r , the flux of ionizing photons decreases due to the local optical depth, parametrized by the local mean free path $\lambda^{-1} = a_0 n_{\text{HI}} = a_0 n(1-x)$, in which a_0 is the hydrogen ionization cross section for Lyman limit photons (Spitzer 1978; Osterbrock 1989). Thus,

$$J + dJ = J - a_0 n(1-x)Jdr \longrightarrow \frac{dJ}{dr} = -a_0 n(1-x)J. \quad (7.14)$$

The number of ionizing photons thus retained compensates for the number of local recombinations:

$$J = \frac{\alpha_B n}{a_0} \frac{x^2}{1-x} \longrightarrow \frac{dJ}{dx} = \frac{\alpha_B n x(2-x)}{a_0 (1-x)^2}. \quad (7.15)$$

Using the chain rule in combination with Eqs.(7.14) and (7.15), we can eliminate J , and we obtain a differential equation for the radial dependent ionization fraction:

$$\frac{dx}{dr} = -a_0 n \frac{x(1-x)^2}{2-x}. \quad (7.16)$$

Introducing the dimensionless unit of length $\Lambda = a_0 nr = r/\lambda$, which is a measure for the number of mean free paths, we obtain

$$\frac{dx}{d\Lambda} = -\frac{x(1-x)^2}{2-x}. \quad (7.17)$$

Choosing the boundary condition $\Lambda = 0$ at $x = 1/2$, we find as a solution to Eq.(7.17),

$$\Lambda = 2 - 2 \ln \left(\frac{x}{1-x} \right) - \frac{1}{1-x}. \quad (7.18)$$

Using Eq.(7.18), we can give a direct estimate of the thickness of the front. Assuming that in the region of the front the ionization fraction drops from $x = 0.9$ to $x = 0.1$, we

find that the corresponding dimensionless distance parameters are $\Lambda = -12.4$ and $\Lambda = 5.3$, respectively.

Thus, we find that the thickness of an ionization front has a value of approximately 20 mean free paths. It depends on the medium properties whether or not this thickness can be considered small with respect to the Strömgren radius. As we shall see in one of the test problems of Chapter 9, it might well be that this thickness of the front, 20λ , is of the same order as the Strömgren radius itself.

SimpleX

Because of its advantageous scaling properties and its adaptive nature, the transport method described in Chapter 4 is an ideal candidate for doing complex cosmological radiative transfer simulations for the Epoch of Reionization. Therefore, we implemented it into a C++ package specifically designed for this purpose. In this chapter, we will go into the details of implementing our new transport method in general terms, focusing on the overall functionality, whereafter we will specify how the relevant physics, specific to this particular problem, was added.

Triangulating Radiation: Radiative Transfer on Unstructured Grids
J. Ritzerveld, V. Icke, and E.-J. Rijkhorst

8.1

Introduction

The transport method that was introduced in Part I of this thesis was already described in a broad context in Chapter 4. One of the distinguishing characteristics of the transport method is that it does not use stiff, regular grids, but makes use of an adaptive random lattice that represents the properties of the background medium. The lattice is constructed by performing a tessellation of a generalized Poisson point process, cf. Eqs.(3.4) and (4.6), which is defined in such a way that it correlates directly with some power of the medium density. This type of mesh construction technique has the intrinsic property that the grid itself follows the inhomogeneous properties of the medium. Consequently, our new method is optimally suited to deal with physical problems in which the medium distribution is highly inhomogeneous, simply because it places the available points where they are needed most.

This recipe does not only use the available resolution in an optimal way, but it also allows the propagation of particles by performing a (constrained) random walk on the resultant mesh. In other words, by choosing a grid that is a direct representation of the interaction properties of the medium, the solution of the transport problem is encoded into the grid. To find the solution for a certain initial condition of sources, an emitted particle merely has to scan the grid in an effort to find its predefined trajectory. Effectively, the adaptive mesh is a graph representation of what particles would do once emitted somewhere in the medium. The resultant random walk procedure can be shown to result in a method that is independent of the number of source locations, as was expressed in the scaling law Eq.(4.6).

In the light of what we discussed on the requirements for numerical simulations of the Epoch of Reionization (EOR) in the previous chapter, it is immediately apparent that a

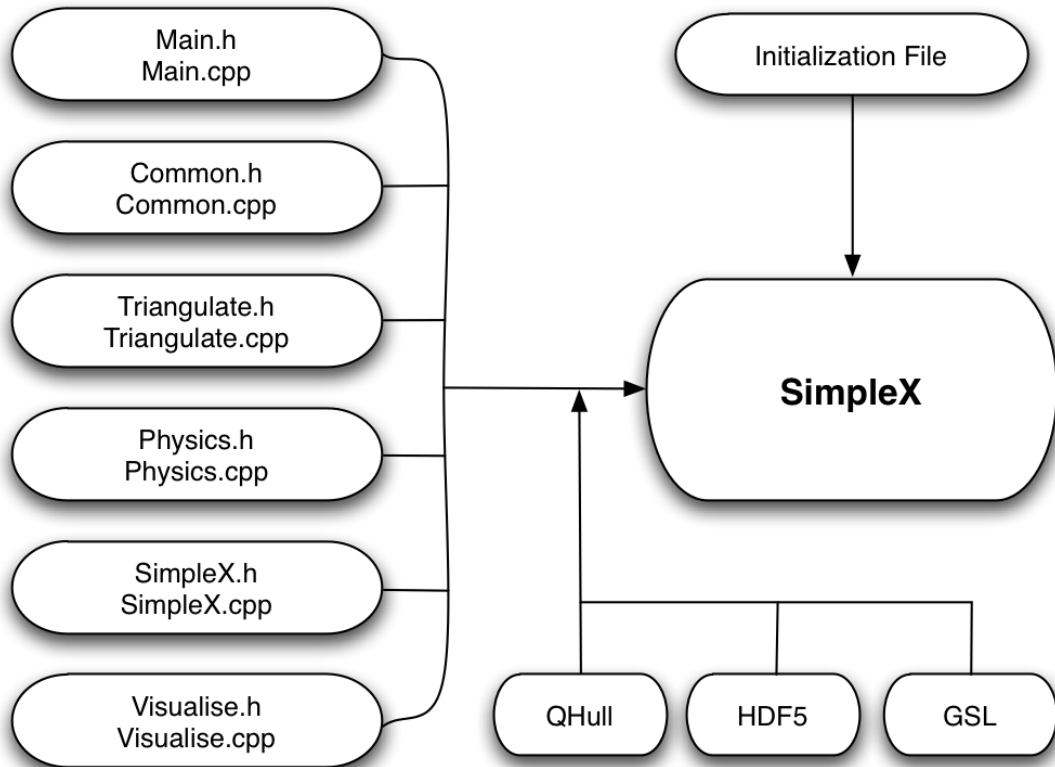


Fig. 8.1: Schematic illustration of the dependency structure of the different source files of the SimpleX package. Several external libraries are linked to our own files, the result of which is an executable that is run with the input of an initialization file.

transport method that can handle both inhomogeneity and a large number of sources is of immediate use, and might even prove essential in relieving the radiative transfer bottleneck of current reionization simulations. Accordingly, we have developed an implementation of our new transfer method that was specifically designed to solve for the cosmological transfer of photons in the early universe. This chapter will serve as an overview of the specifics of the resultant code, mainly aimed at illustrating how to design an implementation of our general transport method, and how to implement the relevant radiation-matter interactions and rate equations. We will start by giving an impression of how our package is structured in Sect. 8.2. In Sect. 8.3, we delve deeper into the details of implementing several steps of the transport algorithm as described in Chapter 4. Applications of our transport code will be given in the next chapter, and results of coupling it to hydrodynamics simulations will be given in Part III of this thesis.

8.2

Code Overview

When we were comparing certain aspects of our transport method to several different well-established branches of science in Chapter 5, we emphasized that our adaptive lattice can be

directly related to a mesh of automata, each point of which is a cellular automaton that has some simple local rules for the redistribution of particles and the interaction with the medium. This particular point of view dictates the use of *object oriented programming* (Budd 2001) as a basis for our implementation. Since it is not intended as a commercial package and because we expect to couple it to large scale multi-processor hydro simulations, the most logical choice of programming language was C++.

The resultant package was given the name SimpleX, after its most elementary constituent. It was designed to function in a modular fashion, with functions written in such a way that they can easily be plugged into the main routines. The overall dependency structure of SimpleX is schematically depicted in Fig. 8.1. Functions with similar purpose were grouped into separate files, and several external libraries are linked to our own libraries. The result is an executable that works according to the parameters specified in an initialization file that conforms to means of a predefined syntax. Comments were added to the source code in Doxygen¹ format, through which an online catalogue of functions and variables can be easily generated.

The different groups of files, together with their functionality, are:

- `Main.h`: Root file, containing startup functions.
- `Common.h`: Class definitions of graph objects, several commonly used functions.
- `Triangulate.h`: Contains all functions dedicated to the mesh construction, i.e. point distribution creation, periodic boundary condition implementation, tessellation procedures, and other geometric point preprocessing steps.
- `Physics.h`: Rate equations, optical depth implementation, and other interaction-related functions.
- `SimpleX.h`: Encompasses all functions that do the actual transfer.
- `Visualise.h`: The user has the choice of several output options, the routines for which are contained in this file.

The SimpleX code uses three packages of external libraries:

- `QHull`: An open-source C implementation of the Quickhull algorithm for performing convex hull constructions, Voronoi tessellations, and Delaunay triangulations. We refer to Sect. 3.4.1 for more details.
- `HDF5`: *Hierarchical Data Format*² is a library and multi-object file format for the transfer of graphical and numerical data between computers. Many modern astrophysical hydrodynamics solvers store their output in the HDF5 standard. In order to enable coupling of SimpleX to hydro output, this library bundle is included.
- `GSL`: The *GNU Scientific Library*³ is an open-source C and C++ library that implements almost all relevant mathematical routines, ranging from random number generation, to least square fitting. We include this library, because we did not aim to reinvent this particular wheel.

Platform independence is ensured by avoiding the use of any platform specific libraries, or

¹<http://www.doxygen.org>

²<http://hdf.ncsa.uiuc.edu/HDF5/>

³<http://www.gnu.org/software/gsl/>

graphical user interface elements. Once compiled and linked to the separately compiled external libraries, SimpleX is ready to run.

8.3

Implementation

The SimpleX package was designed in an effort to incorporate radiative transfer more realistically in modern reionization simulations. To accomplish this, it has to be coupled to either dark matter only simulations, or full-fledged hydrodynamics simulations, that solve for the nonlinear dynamics of the dark matter and gas in the early universe, and predict the position and source functions of the first sources of ionizing radiation. The most realistic results would be obtained if the radiative transfer were incorporated self-consistently into the hydrodynamics solver itself, but this is very costly, even with our new fast method. Fortunately, when one compares the relevant time-scales for the dynamic behavior of the hydrodynamics and the radiative transfer, one can conclude that it is justified to decouple them. The radiation field adapts almost instantaneously to changes on the hydro time-scale, and one can take time slices of the hydro simulation, and for each slice solve the cosmological radiative transfer equation on that quasi-static density field. All reionization simulations done up to now have taken this approach, with the exception of the method presented in Gnedin & Abel (2001).

Although there are current developments in coupling SimpleX self-consistently to hydrodynamics solvers, for example to model the chaotic atmospheres around AGB stars (see e.g. Woitke & Niccolini 2005), we here stick to the common approach just described. Thus, SimpleX solves the quasi-static cosmological radiative transfer equation Eq.(7.7) for every time slice of a separate hydro solver. In this section, we will describe the somewhat more technical aspects of using the general method of Chapter 4 to perform this task. A schematic overview of the steps involved in solving the transfer equations for each hydro slice is depicted in Fig. 8.2. As can be seen from this figure, the procedural approach of SimpleX can be divided into two main sequential groups of algorithms: several preprocessing steps, involving procedures for creating the adaptive random lattice based on the hydro slice input; and the actual transfer steps, in which the physical transfer is solved.

8.3.1 Preprocessing Steps

The first category of algorithms in Fig. 8.2 involves the transformation of the density field, as established from the output of some hydro solver, to the adaptive random lattice that is used for our transport method. The input can have a variety of forms, ranging from analytically known density distributions, to 3D data arrays, or even a distribution of point particles used in SPH methods (Lucy 1977; Monaghan 1992). The first two were used for doing the code comparison test cases, described in the next chapter, and the latter is the input format used in the reionization simulation in Part III of this thesis. In what follows, we will always assume that our computational domain is a 3D cube, the size of which may vary. In contrast with other transfer methods that operate on structured grids, our method does not have a preferred domain topology. However, in order to conform to most input formats, we adhere to the conventional cubical domain.

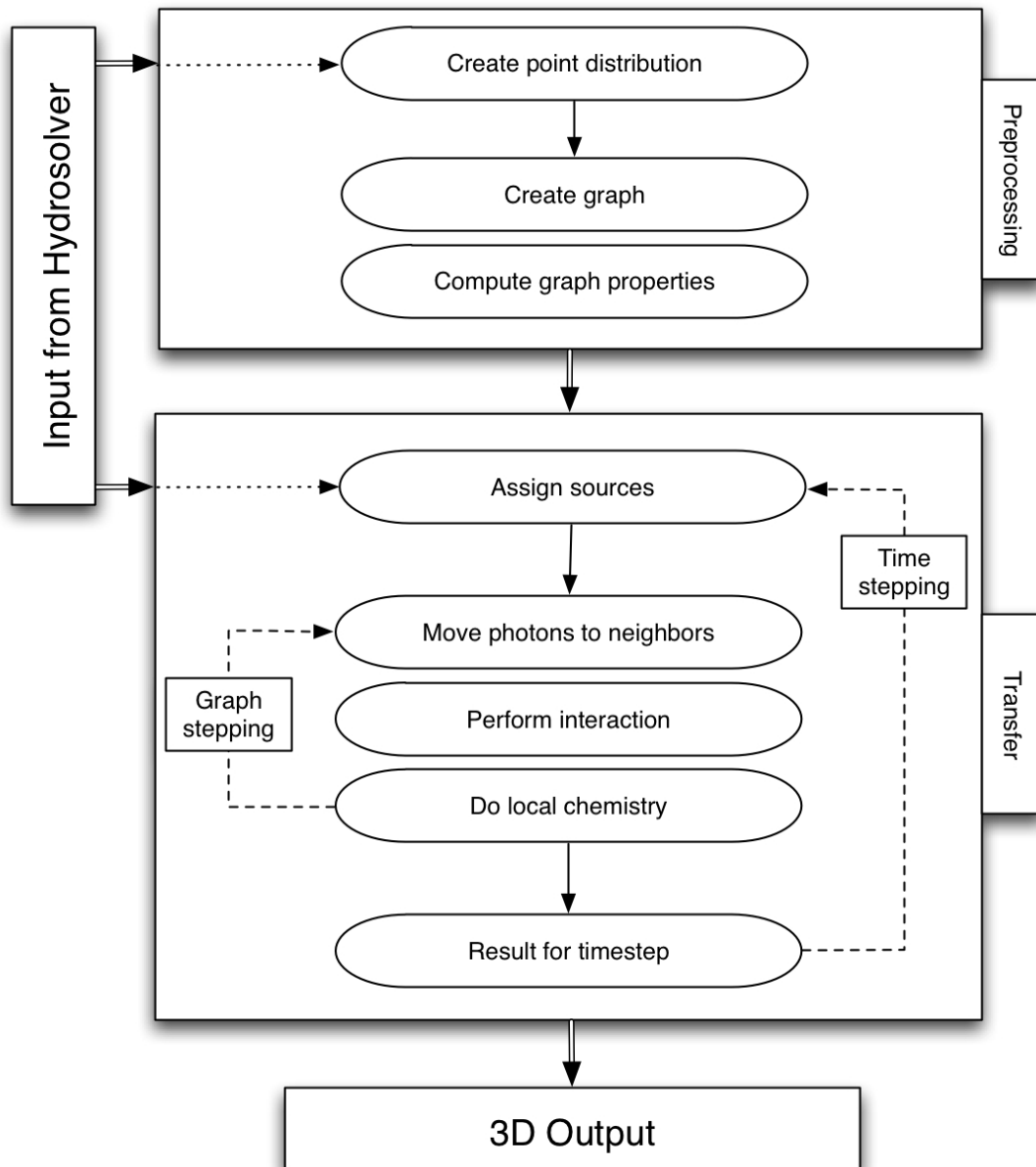


Fig. 8.2: Schematic overview of the steps taken by SimpleX in solving the transfer equation for each input time slice. The overall approach can be split into two main sequential groups: one involving preprocessing steps for the creation of the adaptive random lattice, and one that governs the actual algorithms for doing the transfer.

Generating Point Distributions

The format of input may vary from case to case. There is an appropriate way of generating point distributions from it for each of the ones described.

For a density distribution that can be described by a closed analytical expression there are several ways of obtaining a representative point distribution, conforming to the desired

properties expressed in Eq.(4.6). When the function happens to be monotonically increasing or decreasing on the domain, it might be possible to directly invert the function. The well-known *direct inversion method* (Press et al. 1992) can then be used. In every other case, except when the function has severe discontinuities, *rejection methods* (Press et al. 1992) can be used to generate a point distribution that mimics the density structure. The local properties of the resulting point distribution will then conform to the properties of a Poisson point process.

A bit less trivial is the procedure followed when dealing with regular 3D density arrays. This type of input is encountered most frequently, because it conforms to the type of mesh used in most hydro solvers. We deal with this type of input by normalizing the density array to the number of SimpleX points available for our simulation. A point is placed in every cell with a value above unity, and it obtains the density value of that particular cell. Note that the point is placed at a random position within each cell to ensure that the resultant point distribution has a local Poissonian character. In this approach, one has to be aware that one tends to overestimate the density, because placing the points is biased towards high density regions. We amend this by interpolating the density values of any unused cell, i.e. one without a point, to the closest SimpleX points. The resultant distribution of points, each of which has been assigned a density from the 3D input array, can be checked to give a total number of atoms that is within error margin of the number obtained from the original 3D array.

The last type of input format is one generated by SPH methods for doing cosmological hydrodynamics. The output of this type of method mostly consists of a collection of particles that have certain properties, such as mass, density and range of influence. As long as the number of particles is large enough, their local distribution is random enough to conform to the requirements for our adaptive random lattice, and we can directly use this particle distribution, without further need for manipulation. If the available SimpleX resolution is less than the SPH resolution, we can just take a random subset.

We conclude by noting that, in this description for point distribution generation, we have not strictly used the rule Eq.(4.6). In the case of the SPH input, for example, the particle density will mostly scale with the density, and not the required d -th power of the density. This is not a problem, because the only thing that will change is that the local interaction coefficients will now vary from location to location. The lattice, however, will still be adaptive.

Performing Tessellation

Once a point distribution has been generated on the computational domain, the lattice has to be constructed. This is achieved by calling the external QHull routines. These create the convex hull, Voronoi tessellation, and Delaunay triangulation of the set of input points. Moreover, they provide us with the volume of every Delaunay simplex, which is stored for later use. SimpleX stores the graph internally by creating an array of point objects, which has as member variables a number of pointers, equal to the connectivity number of the vertex, each of which is pointing to one of the neighbors.

It is desirable for cosmological simulations to impose periodic boundary conditions. For structured lattices, this topological Clifford torus is trivially implemented. For unstructured lattices, like the one we are considering now, the implementation is a bit more involved. In order to establish periodicity, we copy the points in the domain to each side of the boundary of the domain (cf. Fig. 8.3 for a two-dimensional example). In order to save memory and computational expense, we only consider a limited region (*white*) that is small enough to

minimize the extra computations needed, but large enough to ensure that the tessellation is still correct near the boundaries. This can be verified by using a scheme using buffer points near the boundary regions. We refer the reader to Neyrinck et al. (2005) for more details.

The result is a tessellation that conforms to the periodic boundary conditions. `SimpleX` points near the boundary may have lines to neighbors that cross the boundary due to these conditions. Internally, these points then have a pointer that points to the relevant point at the other side of the domain, by which radiation moving from these points to their respective neighbors will move along that line, or pointer, appearing to disappear from one end of the domain, entering on the other side.

The user can specify in the initialization file whether or not `SimpleX` should operate with or without the periodic boundary conditions. Although it is often dictated by external hydro solvers, it is by no means necessary to stick to cubical domains. A similar procedure for constructing a periodic adaptive random lattice can be used when the domain is not cubical, but, for example, spherical.

The memory consumption even of quite efficient tessellation packages such as `QHull` can become prohibitively high when the number of points is of the order of millions, a typical number when dealing with realistic reionization simulations. In order to circumvent this becoming a bottleneck in our transfer code, we have arranged for the computational domain to be split into N parts in each direction. Each part then contains a number of points that is more tractable. Buffer regions around each sub-box are included to ensure that the N^3 resultant sub-graphs can be seamlessly tied up into a whole.

Other Preprocessing Steps

Once the adaptive random lattice has been constructed based on the generated point distribution, the preprocessing phase of `SimpleX` is concluded by performing several housekeeping steps. There is a category of geometric procedures, in which the lengths of the lines of the graph, but also the d most straightforward paths (cf. Sect. 4.4) for each line are computed. Finally, there is a series of procedures that determine from the density and volume at each point the local number of atoms, and other relevant physical properties.

8.3.2 Transfer Steps

As soon as the first preprocessing phase of `SimpleX` has been completed, the actual transfer can commence. As with the construction of the point distribution and the resultant adaptive random lattices, the transfer is decomposed into a sequence of steps.

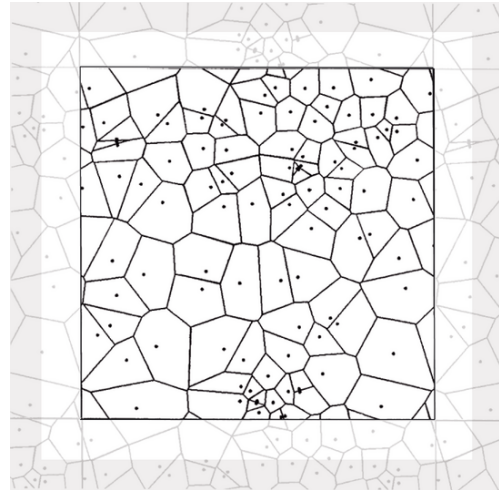


Fig. 8.3: A periodic tessellation is constructed by copying points along the boundaries. For optimization purposes, only points within the white region are considered.

Assigning Sources

The input from the hydrodynamics simulations does not only provide us with a neutral hydrogen density field, it also contains a catalogue of sources, and their source function, or spectra. Consequently, several points of our adaptive random lattice are marked as a source, with a certain luminosity of ionizing photons. This luminosity is obtained by integrating the source spectrum S_ν divided by the energy of each photon,

$$L_{\text{ion}} = \int_{\nu_0}^{\infty} \frac{S_\nu}{h\nu} d\nu, \quad (8.1)$$

with a lower limit ν_0 denoting the Lyman limit. By integrating, we lose the information contained in the spectrum, neglecting the effect that hard photons have larger mean free paths, and might therefore penetrate deeper into the neutral hydrogen. This is compensated for by using the mean opacity approach, cf. Sect. 7.3.3,

$$\bar{a} = \int_{\nu_0}^{\infty} S_\nu^{\text{norm}} a(\nu) d\nu, \quad (8.2)$$

in which we used the frequency-dependent hydrogen ionization cross section⁴

$$a(\nu) = A_0 \left(\frac{\nu_0}{\nu} \right)^3 \quad \forall \nu \geq \nu_0, \quad (8.3)$$

with reference value (Spitzer 1978; Osterbrock 1989)

$$A_0 = a(\nu_0) = 6.3 \cdot 10^{-18} \text{cm}^2. \quad (8.4)$$

The weighing function in Eq.(8.2) is the source spectrum normalized above the Lyman limit, i.e.

$$S_\nu^{\text{norm}} = \frac{S_\nu}{\int_{\nu_0}^{\infty} S_\nu d\nu}. \quad (8.5)$$

Thus, this approach entails that we effectively use only one frequency, storing the number of ionizing photons emitted per second only, but it does account for the fact that the spectrum may be very hard by using a lower effective absorption coefficient, Eq.(8.2), from which the local interaction coefficients are computed, as explained in Sect. 8.3.2.

Moving Photons

The hydrodynamics simulations that were run separately from our transfer method are performed over an interval in redshift, or time. At a certain point in time, the first sources form, and it is at this point that radiative transfer is needed to resolve for the feedback effect of the HII regions being blown. The density fields computed by the hydro solvers are sliced in time. Thus, each slice presented to SimpleX spans a certain number of, say, megayears. When computing the evolution of the HII regions being blown in each separate slice, a time step

⁴A more exact expression for the hydrogen ionization cross section is

$$a(\nu) = A_0 \left(\frac{\nu_0}{\nu} \right)^4 \frac{\exp(4 - 4 \arctan \epsilon/\epsilon)}{1 - \exp -2\pi/\epsilon},$$

in which $\epsilon = \sqrt{\nu/\nu_0 - 1}$. Fortunately, this function closely resembles the much more simple expression Eq.(8.3), which is therefore used more frequently.

Δt is used that is typically much smaller than the time spanned by one slice. This ensures that the front movement can be accurately followed, cf. Sect. 7.3.2.

At the beginning of each time step t_i , the points that have been marked as a source isotropically emit a number of ionizing photons $N_{\text{ion}} = L_{\text{ion}}\Delta t$ (i.e. an equal number of photons is emitted to each of the neighboring points). Looping over all points in the graph, photons are redistributed to neighboring points according to recipes such as depicted in Fig. 4.3, at which interaction takes place. Continuing the loop, the photons perform a random walk, moving from one interaction center to the next.

Interaction

In this specific case, we only consider as interaction the ionization of the local hydrogen atoms by the influx of photons. Consider a typical point of our graph that receives a number of photons N_{ion} from one of its neighboring points. Its hydrogen density n , total number of local atoms N , and the physical length ΔL to its neighbor have all been determined in the preprocessing steps, so that they are constant in this step. The time dependence of the local absorptive properties is incorporated via an extra factor $x(t) = n_{\text{HII}}/n$, denoting the local ionization fraction.

When moving from the neighboring point to the current one, along a line with distance ΔL , the photons N_{ion} encountered an optical depth

$$\Delta\tau = n_{\text{HII}}\bar{\alpha}\Delta L = (1 - x)n\bar{\alpha}\Delta L, \quad (8.6)$$

in which we used the mean absorption coefficient Eq.(8.2), associated with the spectrum of these photons. Eq.(8.6) is a measure of the number of mean free paths ($\lambda = 1/n_{\text{HII}}\bar{\alpha}$) in the line length ΔL , and is equivalent to the interaction coefficient c_{ion} , as discussed in Sect. 4.4. Thus, the number of ionizing photons propagating onwards from this point is

$$N_{\text{ion}}^{\text{out}} = N_{\text{ion}}e^{-\Delta\tau}, \quad (8.7)$$

and, similarly, the number of photons used to ionize the local neutral hydrogen atoms is

$$N_{\text{ion}}^{\text{used}} = N_{\text{ion}}(1 - e^{-\Delta\tau}). \quad (8.8)$$

The number of photons withheld are used to ionize the same number of neutral hydrogen atoms locally present, changing the relative number of neutral and ionized hydrogen atoms. This instantly alters the local ionization fraction x .

Note that this recipe ensures that photons are explicitly conserved, that is $N_{\text{ion}}^{\text{out}} + N_{\text{ion}}^{\text{used}} = N_{\text{ion}}$, which makes certain that the ionization fronts will have the correct speed. Moreover, the local opacity is immediately updated, not just at the end of a time step, as is the case in other methods.

Chemistry

At the end of every time step t_i , we have looped over all points, propagating the photons from neighbor to neighbor, ionizing neutral material along the way. At this point in the simulation, we account for recombination by integrating, at each point of the graph, the recombination rate over the time step Δt , i.e. (Spitzer 1978; Osterbrock 1989)

$$N_{\text{rec}} = \alpha_{\text{A}}n_{\text{HII}}^2\Delta t = \alpha_{\text{A}}x^2n^2\Delta t. \quad (8.9)$$

Here, we have used the total recombination coefficient $\alpha_A(T)$, cf. Sect. 7.3.4, but in the on the spot approximation (see Chapter 11) one might wish to use $\alpha_B(T) = \alpha_A(T) - \alpha_1(T)$. The number of recombined atoms are locally added to the number of neutral atoms, by which the local ionization fraction is updated.

If we do not follow the on the spot approximation, a fraction $\alpha_1(T)/\alpha_A(T)$ of the number of recombination actually produces ionizing photons that have energies that are highly peaked just above the Lyman limit. These diffuse photons can then compete with the other ionizing photons, and this particular point is then added as a source. Once the next time step begins, and the loop over all points commences, these photons contribute to the total field of ionizing radiation.

8.3.3 Output

At the end of the time spanned by one of the hydro slices, SimpleX has looped over many small time steps with length Δt , looping over the whole adaptive random lattice during each. The result is an ionization fraction at each point of the lattice, that might have changed due to the influence of the ionizing photons emitted by the sources present.

There are several things that can be done with this overall result. Data, such as the ionizing photons still unused, new sources that were added, and the local ionization fractions, can be stored, so that they can be used as input together with the hydro input from a possible next slice. A large scale simulation, spanning a large range in redshift, can be run in this fashion.

Of course, there are also multiple options for producing output for visualization purposes. Geometrically, the change in ionization fractions at each point can be interpreted as the expansion of HII regions that overlap. In order to be able to get a 3D impression of this process, we have implemented the option to interpolate the data on our adaptive random lattice onto a regular datacube with a certain resolution. This interpolation routine introduces some smearing, or noise, but the regular datacube is almost always the standard input format, especially when our method is to be compared to other transfer methods, as in the next chapter, or when it needs to be dynamically coupled to hydro solvers that use this type of mesh. It is also much easier to produce power spectra and the like from this type of mesh. Alternatively, we have written a routine that computes the intersection of a plane with the adaptive lattice. The cells that are sliced are projected onto the plane, and their data can be visualized two-dimensionally in this fashion, the result being dumped in a PostScript file.

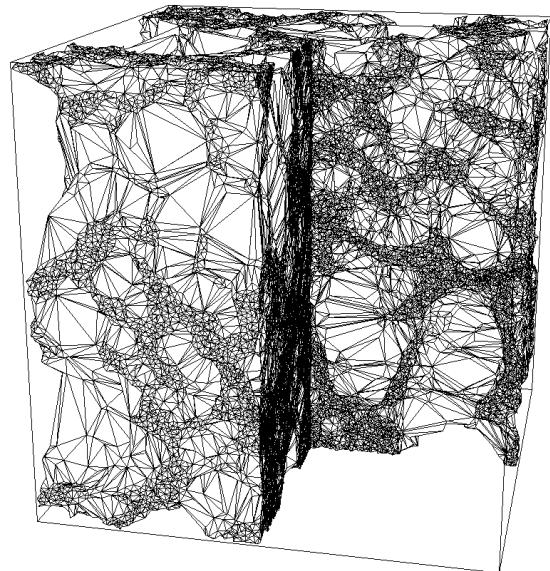


Fig. 8.4: A visualization of the adaptive random lattice in a typical reionization simulation.

The most elegant output format SimpleX can produce is in the form of VTK-files⁵. This

⁵<http://public.kitware.com/VTK/>

Visual Toolkit file format does not impose restrictions on the type of mesh used, by which our adaptive random lattice, consisting of many tetrahedron-shaped cells, can be directly written to this type of file, without the need for any interpolation. See, for example, Fig. 8.4, in which is depicted a typical adaptive random lattice, with a cutout to show more clearly the internal inhomogeneous properties of the graph. This lattice was generated from a typical density distribution of a reionization simulation. It is obvious that we would need a very high resolution datacube to capture all of the features of this adaptive lattice. The open source *VisIt*⁶ package is a 3D visualization tool, which was designed to be multi-platform. It reads in these VTK-files without any problem, and it has many useful features for creating isosurfaces, volume renderings, and even movies based on the data defined on the mesh. All 3D visualizations in this thesis, including the one in Fig. 8.4 were created using *VisIt*.

8.4

Conclusions

This concludes our discussion on how the general method for the transport of particles on an adaptive random lattice, as described in Chapter 4, was implemented into a package that covers all of the essentials for doing cosmological radiative transfer.

It is still not complete, however. There are several secondary physical effects that remain to be included. Most important is the lack of the self-consistent solution of the energy equation for the gas. Effects such as photo-heating and cooling, which influence the temperature and pressure of the gas, are ignored. This is important, because recombination and collisional ionization rates depend on the local gas temperature. For now, we restrain ourselves to making the justified approximation that neutral gas is assigned a fixed temperature of 100K, whereas ionized gas is assigned a typical temperature of 10^4 K (Osterbrock 1989).

Still, *SimpleX* has many advantages that make it ideal for doing realistic reionization simulations. Its adaptive properties ensure that effectively high resolution simulations can be done on simple desktop computers within several minutes wall clock time. The added advantage that the overall method does not scale with the number of sources entails that we do not have to refrain from including many inhomogeneously distributed sources, or from including diffuse radiation. The method is explicitly photon conserving, ensuring the correct front speed. Moreover, because the ionization fractions are updated dynamically throughout the loop over the graph, hopping randomly from one point to the next, any under- or overshoots of the ionization front, as discussed at the end of Sect. 7.3.4, are avoided.

The results presented in the next chapter and the last part of this thesis have all been obtained by using the version of *SimpleX* as described in this chapter, and we refer the reader to those parts of the text for more on our new cosmological radiative transfer method.

Acknowledgements We are much obliged to Hank Childs at LLNL for providing us with efficient source code for creating VTK-files of our adaptive random lattices.

⁶<http://www.llnl.gov/visit/>

Bibliography

- Budd, T. 2001, *An Introduction to Object-Oriented Programming* (Reading, MA: Addison Wesley)
- Gnedin, N. Y. & Abel, T. 2001, *New Astronomy*, 6, 437
- Lucy, L. B. 1977, *AJ*, 82, 1013
- Monaghan, J. J. 1992, *ARA&A*, 30, 543
- Neyrinck, M. C., Gnedin, N. Y., & Hamilton, A. J. S. 2005, *MNRAS*, 356, 1222
- Osterbrock, D. E. 1989, *Astrophysics of gaseous nebulae and active galactic nuclei* (Mill Valley, CA: University Science Books)
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, *Numerical Recipes in C: The Art of Scientific Computing* (New York, NY, USA: Cambridge University Press)
- Spitzer, L. 1978, *Physical processes in the interstellar medium* (New York: Wiley-Interscience)
- Woitke, P. & Niccolini, G. 2005, *A&A*, 433, 1101

Code Comparison Project

We present a detailed comparison of our new cosmological radiative transfer method `SimpleX` with a large set of other cosmological radiative transfer methods, on several common tests. The participating codes represent a full variety of existing approaches, ranging from ray-casting methods to statistical ones, which solve the transfer equation on regular or adaptive meshes. The tests examine the implementation of our basic physics, the behavior of expanding ionization fronts, and the overall morphology of a cosmological density field that is reionized.

Cosmological Radiative Transfer Codes Comparison Project - I. The Static Density Fields

I.T. Iliev, B. Ciardi, M.A. Alvarez, A. Maselli, A. Ferrara, N.Y. Gnedin,
G. Mellema, T. Nakamoto, M.L. Norman, A.O. Razoumov, E.-J. Rijkhorst,
J. Ritzerveld, P.R. Shapiro, H. Susa, M. Umemura, and D.J. Whalen
MNRAS (2006), **371**, 1057-1086

9.1

Introduction

As discussed in Chapter 7, the radiative transfer equation in 3D space is seven dimensional. Although in specific cases certain kinds of symmetry or approximations can be exploited, leading to a partial simplification, most problems of astrophysical and cosmological interest remain very complex. For this reason, although the basic physics involved is well understood, the detailed solution of the complete radiative transfer equation is presently beyond available computational capabilities. In addition, the technical implementation of the radiative transfer equation is a very young and immature subject in astrophysics. As the constraint on the available computational power has become less demanding, the number and type of desirable applications have expanded extremely rapidly, particularly spreading into the field of galaxy formation and cosmology.

Following this wave of excitement, several groups then attacked the problem from a variety of perspectives by using completely different, independent and dedicated numerical algorithms, most of which were described in Sect. 7.4. It was immediately clear that the validation and assessment of the various codes were crucial in order not to waste limited computational and human resources. At that point, the community faced the problem that, in contrast with e.g. gasdynamical studies, very few simple radiative transfer problems admit exact analytical solutions that could be used as benchmarks. Just a few years ago, only a few radiative transfer codes were available and these were mostly still in the testing and

optimization phase, and therefore lacking the necessary degree of stability required to isolate truly scientific results from uncertainties due to internal programming bugs. In the last few years, the subject has rapidly changed. Not only has the reliability of existing codes matured, but also a new crop of codes based on novel techniques, such as our `SimpleX` method, has been developed. Thus, a comparison of them was timely. This chapter presents the detailed outcome of comparing our method to most of the others.

The project is in a similar spirit to the well-known *Santa Barbara Cluster Comparison project* (Frenk et al. 1999), but we have utilized a somewhat different approach from that project. Instead of considering a single, complex problem like the Santa Barbara Cluster Comparison did, we considered a set of relatively simple problems. The aim of the comparison was to determine the type of problems the codes are able to solve, to understand the origin of the differences inevitably found in the results, to stimulate improvements and further developments of the existing codes and, finally, to serve as a benchmark for testing future ones.

The project was a collaboration of a large number of cosmological radiative transfer researchers, and includes a wide range of different methods. The comparison was made among 11 independent codes: `C2-ray` (Mellema et al. 2006), `OTVET` (Gnedin & Abel 2001), `CRASH` (Maselli et al. 2003), `RSPH` (Susa 2006), `ART` (Nakamoto et al. 2001), `FTTE` (Razoumov & Cardall 2005), `ZEUS-MP` (Whalen & Norman 2006), `FLASH-HC` (Rijkhorst et al. 2006), `IFT` (Alvarez et al. 2006), `CORAL` (Mellema et al. 1998), and, of course, our own novel technique, `SimpleX`. The list of codes is representative for the available numerical radiative transfer techniques, as discussed in Sect. 7.4, and includes ray-casting codes, moment methods, Monte Carlo solvers, and short characteristics techniques. We refer the reader to the relevant papers for more details on each code.

In this chapter, we describe the tests we have performed, along with their detailed parameters, geometry and set-up, and the results we obtained. When constructing these tests, we aimed for the simplest and cleanest, but nonetheless cosmologically interesting problems. We designed them in a way which allows us to test and compare all the important aspects of any radiative transfer code. We start with describing the results of comparing the implementation of the basic physics of each method in Sect. 9.2, after which we move on to the classical test of an expanding HII region in Sect. 9.3. We finish with a more realistic test of multiple sources emitting ionizing photons in an actual cosmological density field in Sect. 9.4.

All test problems are solved in three dimensions, with grid dimensions 128^3 cells. All of the codes listed above, except our `SimpleX` and the `RSPH` code, operate on structured grids, and comparing those codes to ours is not entirely fair in the sense that we need to interpolate our adaptive random lattice to their 128^3 fixed one, in order to enable comparison. Issues associated with this interpolation will be pointed out in the relevant sections. In all the tests, we use as many `SimpleX` points as there are grid cells. Unless otherwise noted, the sources of ionizing radiation are assumed to have a blackbody spectrum with effective temperature $T_{\text{eff}} = 10^5\text{K}$, except for the classical HII region test, in which the assumed spectrum is monochromatic with all photons having energy $h\nu_0 = 13.6\text{eV}$, the ionization threshold of hydrogen. In all the tests, recombination is incorporated via the on the spot approximation. Diffuse photons, that are generated by the capture of electrons directly to the ground level, are ignored, and an effective recombination coefficient $\alpha_{\text{B}}(T)$ is used. The validity of this approach is discussed in Chapter 11.

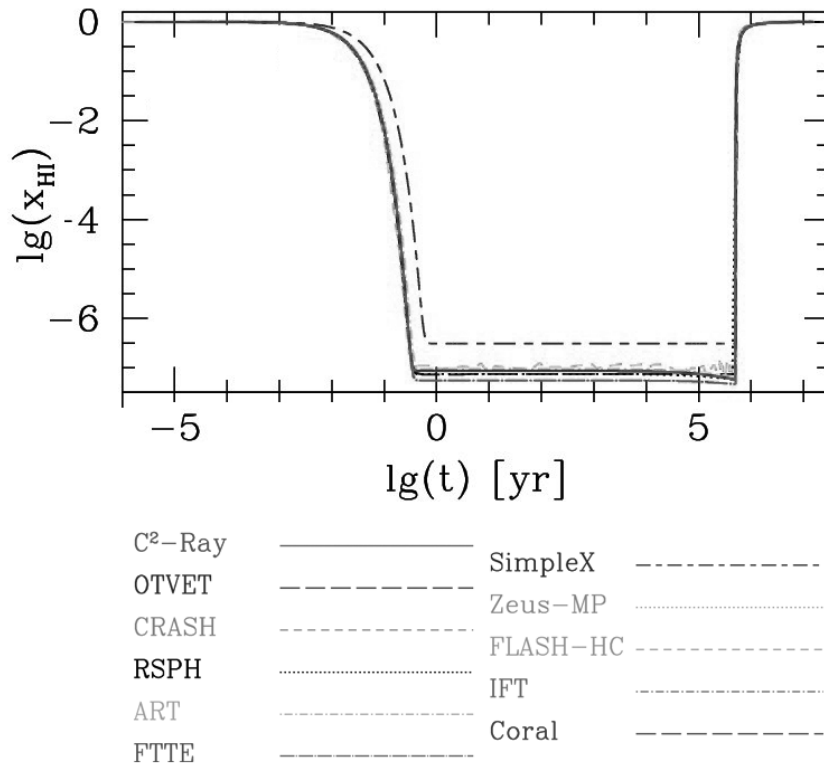


Fig. 9.1: Results for the basic physics test, showing the evolution of the neutral fraction, $x_{\text{HI}} = 1 - x$, versus time. A single zone of gas is ionized by a certain influx of ionizing photons. After a certain time, the source of photons turns off, and the zone recombines. Each code will have its own line type, as indicated in the bottom half.

9.2

Basic Physics

The solution of the radiative transfer equation is closely related to the ionization and thermal states of the gas. These depend on the atomic physics reaction rates, photoionization cross sections, as well as the cooling and heating rates used. As said, SimpleX does not include the latter two, and the specifics of the first two were already discussed in Sect. 8.3.2.

One of the first things we tested in this comparison project was the robustness of the methods we use for solving these non-equilibrium rate equations. These equations are stiff and thus generally require implicit solution methods. Such methods are generally expensive, however, so often certain approximations are used to speed up the calculations. In order to test them, we performed the following simple test with a single, optically thin zone. We start with a completely neutral zone at time $t = 0$. We then apply a photoionizing flux of $F = 10^{12}$ photons $\text{s}^{-1}\text{cm}^{-2}$, with a 10^5K blackbody spectrum for 0.5 Myr, which results in the gas parcel becoming heated and highly ionized. Thereafter, the ionizing flux is switched off, and the zone cools down and recombines for a further 5 Myr. The zone contains only hydrogen gas, with a number density of $n = 1 \text{ cm}^{-3}$, and an initial temperature of $T_i = 100\text{K}$.

The results for all the codes are depicted in Fig. 9.1, top, in which the logarithm of the neutral hydrogen fraction is plotted versus the logarithm of the time. Fig. 9.1, bottom,

provides us with a legend of all the codes used. The same line types for each code will be used throughout this chapter. Most of the times, though, the lines will overlap, and we will then specifically address the ones that deviate. One can see that all codes agree well in terms of the evolution of the neutral fraction. The only one that sticks out is our `SimpleX` code. The reason for this discrepancy is that our code currently does not solve the energy equation to find the gas temperature, because we anticipate using `SimpleX` in hydrodynamical situations where the temperature is, in fact, found from the equation of motion for the energy density. The temperature of the gas will in reality be higher than the one assumed by our code, and the resultant recombination coefficient will be lower. Still, the overall difference is minor.

It is worthwhile to compare these results to what can be analytically derived. The relevant rate equation to be solved here is the one for the hydrogen ionization fraction $x = n_{\text{HII}}/n$. The change in ionization fraction is due to photoionizations, and due to ionized atoms recombining, i.e.

$$\frac{dx}{dt} = (1-x)A_p - x^2\alpha_B(T)n. \quad (9.1)$$

In this equation, we have ignored collisional ionization, as it has not yet been incorporated into `SimpleX`. Fortunately, this is mostly a small secondary effect in the cosmological context. The recombination coefficient $\alpha_B(T)$ depends on the gas temperature in the following approximate way (Spitzer 1978):

$$\alpha_B(T) = 2.59 \cdot 10^{-13} \left(\frac{T}{10^4} \right)^{-0.7} \text{ cm}^3\text{s}^{-1}. \quad (9.2)$$

As said, our transfer code does not solve self-consistently for the temperature of the gas, and we will assume a temperature of 100K for neutral gas, and a temperature of 10^4K for ionized gas. The photoionization rate for a given flux of ionizing photons F , with a blackbody spectrum profile B_ν ($T_{\text{eff}} = 10^5\text{K}$), has the form

$$A_p = F \int_{\nu_0}^{\infty} B_\nu^{\text{norm}}(T_{\text{eff}}) a(\nu) d\nu = F\bar{a}, \quad (9.3)$$

in which we used the mean opacity \bar{a} defined in Eq.(8.2). Given the normalized blackbody spectrum

$$B_\nu^{\text{norm}}(T_{\text{eff}}) = 15 \frac{h^4}{\pi^4 k^4 T_{\text{eff}}^4} \frac{\nu^3}{e^{h\nu/kT_{\text{eff}}} - 1}, \quad (9.4)$$

with Planck's constant h and Boltzmann's constant k , we can use the expression Eq.(8.3) for the hydrogen ionization cross section $a(\nu)$ to integrate Eq.(9.3) exactly. Hence, we obtain as the photoionization rate for a certain effective temperature of the source T_{eff} ,

$$A_p(T_{\text{eff}}) = F \cdot \frac{15A_0}{\pi^4} \left(\frac{h\nu_0}{kT_{\text{eff}}} \right)^3 \ln \left(\frac{1}{1 - e^{-h\nu_0/kT_{\text{eff}}}} \right). \quad (9.5)$$

Given a constant photoionization rate A_p , recombination coefficient α_B (10^4K), and density n , one can solve Eq.(9.1) exactly, and we refer the reader to Appendix 9.A for more details. Using Eq.(9.5) and Eq.(9.2), we find that in this test case, $A_p = F \times A_0 \times 0.15 \text{ s}^{-1} = 10^{12} \times 6.3 \cdot 10^{-18} \times 0.15 \text{ s}^{-1}$ (i.e. $\bar{a} = 0.15A_0$), and $\alpha_B = 2.59 \cdot 10^{-13} \text{ s}^{-1}$, respectively. At the end of the first phase of this test, we have as an equilibrium solution to the rate equation, cf. Eq.(9.19),

$${}^{10}\log(1 - x_+) = -6.56, \quad (9.6)$$

which exactly matches our equilibrium solution in Fig. 9.1. Once the second phase of the test commences, and the photoionizing flux turns off, $A_p = 0$, and Eq.(9.1) has the exact solution

$$x(t) = \frac{1}{\alpha_B n t + x_+}, \quad (9.7)$$

which also exactly matches the right part of the curve in Fig. 9.1.

Thus, we can conclude that the implementation of the basic physics into our `SimpleX` code, as explained in Sect. 8.3.2, is very accurate, because it produces results that exactly conform to the analytical solutions. All these solutions assume fixed temperatures, and including temperature evolution into our code will only modify the equilibrium solution Eq.(9.6) to the ones obtained by the other codes, as depicted in Fig. 9.1.

9.3

Simple Ionization Front

After we had ascertained that the implementation of the basic physics is correct, we could do a more sophisticated test. For this, we have chosen the classical problem of an HII region expanding in uniform gas around a single ionizing source (Strömgren 1939; Spitzer 1978). A steady, monochromatic $h\nu_0 = 13.6\text{eV}$ source, emitting S_* ionizing photons per unit time, is turning on in an initially neutral, uniform density, static environment with hydrogen number density n . For this test, we assumed a fixed gas temperature of $T = 10^4\text{K}$. Under these conditions, and if we assume that the front is sharp (i.e. that it is infinitely thin, with the gas inside fully ionized and the gas outside fully neutral), there is a well-known analytical solution for the evolution of the ionization front radius, r_1 , and its velocity, v_1 , given by

$$r_1 = R_S \left(1 - e^{-t/t_{\text{rec}}}\right)^{1/3} \quad (9.8)$$

$$v_1 = \frac{R_S}{3t_{\text{rec}}} \frac{e^{-t/t_{\text{rec}}}}{\left(1 - e^{-t/t_{\text{rec}}}\right)^{2/3}}, \quad (9.9)$$

where

$$t_{\text{rec}} = \frac{1}{\alpha_B(T)n} \quad (9.10)$$

is the recombination time, and

$$R_S = \left(\frac{3S_*}{4\pi\alpha_B(T)n^2}\right)^{1/3} \quad (9.11)$$

is the Strömgren radius, cf. Eq.(7.12), which delimits the final, maximum size of the ionized region at which point recombinations inside balance the incoming photons. The HII region initially expands quickly, and then slows down considerably as the evolution time approaches the recombination time, at which point the recombinations start balancing the ionizations. This is when the HII region starts approaching its Strömgren radius. After a few recombination times the ionization front stops at radius $r_1 = r_S$, and in the absence of gas motion remains static thereafter.

The particular numerical parameters we used for this test are as follows. The dimension of the computational box is $L = 6.6$ kpc, the gas number density is $n = 10^{-3}\text{cm}^{-3}$, the initial ionization fraction (given by collisional equilibrium) is $x = 1.2 \times 10^{-3}$, and the ionizing photon

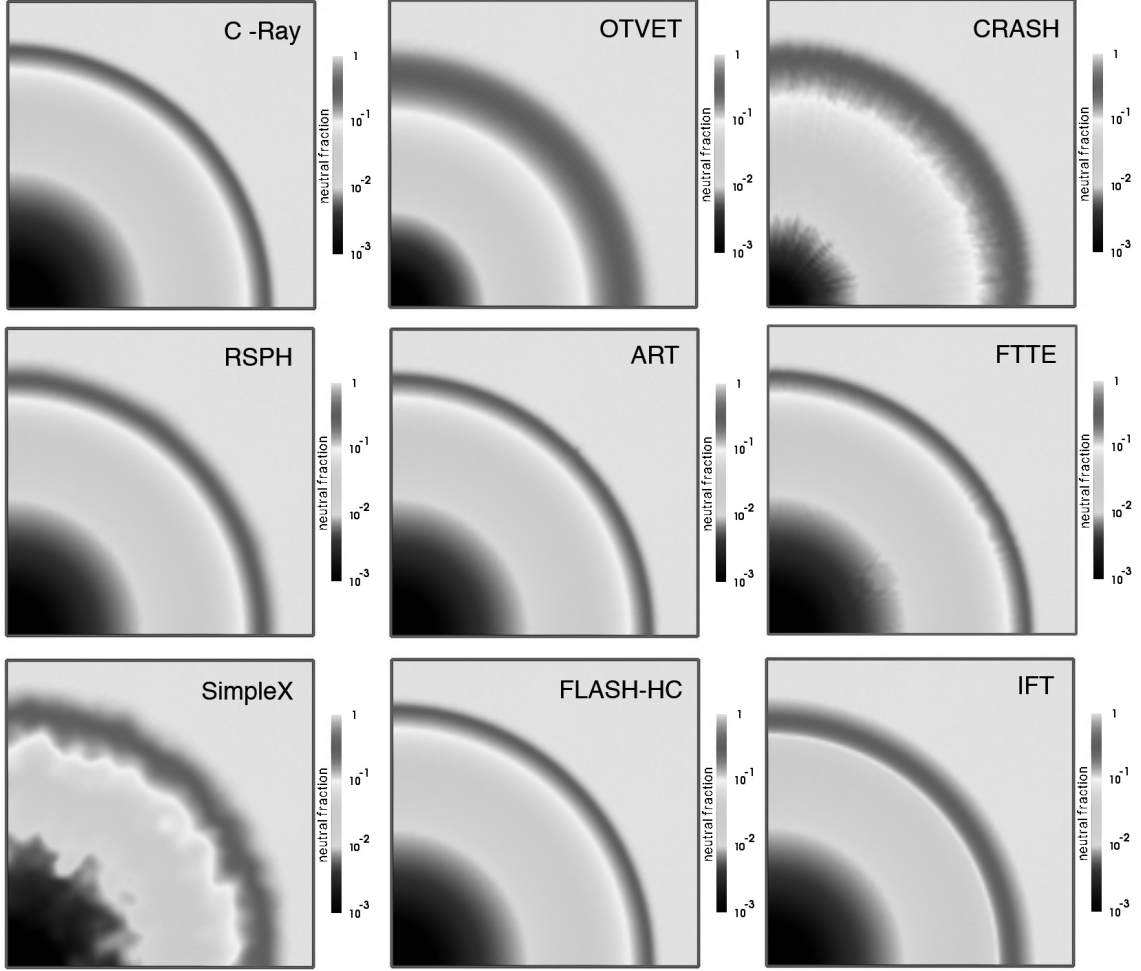


Fig. 9.2: Results for the HII region expansion in a uniform gas at fixed temperature. Shown are images of the HI fraction, cut through the simulation volume at $z = 0$ at time $t = 500$ Myr, when the final Strömgen sphere has been reached. Each panel indicates which of the codes produced that specific result.

rate is $S_* = 5 \times 10^{48}$ photons s^{-1} . The source is defined to be located at the corner of the box. For these parameters, the recombination time is $t_{\text{rec}} = 3.86 \times 10^{15} s = 122.4$ Myr. Assuming a recombination rate $\alpha_B(T) = 2.59 \times 10^{-13} \text{cm}^3 s^{-1}$ at $T = 10^4 \text{K}$, the Strömgen radius will be at $R_S = 5.4$ kpc. The simulation time is $t_{\text{sim}} = 500$ Myr $\approx 4t_{\text{rec}}$. The required outputs are the neutral fraction of hydrogen on the whole regular grid at times $t = 10, 30, 100, 200$ and 500 Myr, and the ionization front position (defined by the 50% neutral fraction threshold) and velocity versus time along the x-axis.

In Fig. 9.2, we show for each code images of the neutral fraction in the $z = 0$ plane at time $t = 500$ Myr, at which point the equilibrium Strömgen sphere has been reached. As can be seen from the figure, the size of the final-ionized region is in very good agreement between the codes. Most codes produce nice spherical HII regions, with the exception of CRASH and our SimpleX. In the first case, this is due to inherent Monte Carlo sampling nature of the code, and in the case of our method, this is partly due to the statistical nature of our algorithm, but mainly due to the use of an adaptive lattice, instead of the regular grid. Interpolation from our grid to the fixed one was needed, and errors associated with this are

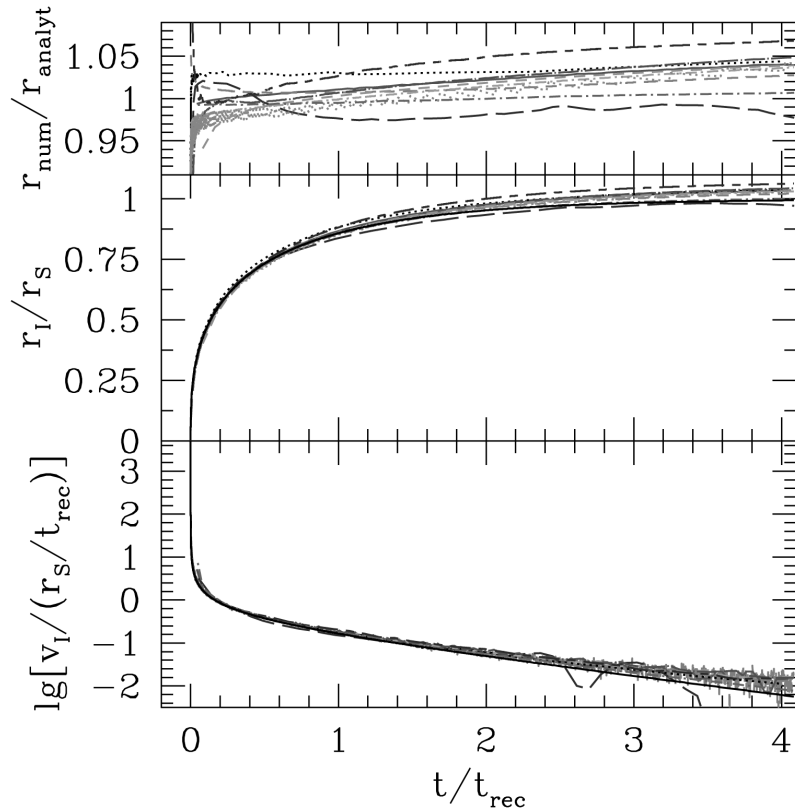


Fig. 9.3: Results for the evolution of the position and velocity of the ionization front of an HII region expanding in a uniform gas at fixed temperature. The analytic solutions Eqs.(9.8) and (9.9) are indicated by the black solid lines in the bottom two graphs.

exaggerated in a logarithmic plot of the fractions. As can be seen from the analysis in Sect. 4.B, the noise in our method goes down with the square root of the number of steps taken. Thus, we expect that the front will become more spherical, when the distance to the source increases. Indeed, one can see in Fig. 9.2 that the front produced by SimpleX is more spherical further out. As stated above, we have chosen to use a number of Delaunay points equal to the number 128^3 of cells used by the other methods. Had we chosen to compare results for the same wall-clock computing time, we could have a few orders of magnitude more points.

There are also certain differences in the thickness of the ionized-neutral transition at the Strömberg sphere. The inherent thickness of this transition, defined as the radial distance between 0.1 and 0.9 ionized fraction points, for a monochromatic spectrum is $\approx 18\lambda$, cf. Sect. 7.A. For the parameters in this test, this thickness is approximately equal to 0.74 kpc, or to about 14 regular simulation cells, which amounts to 11% of the simulation box size. Most codes yield widths very close to this value; only OTVET, CRASH, and our SimpleX find somewhat thicker transitions, mostly due to the inherent diffusivity of the first two methods, and the adaptive-to-regular mesh interpolation procedure of ours. The first two methods also find smaller high ionized proximity regions due to this diffusivity, while our method resolves this very well compared to the others.

In Fig. 9.3, we show the evolution of the ionization front position and velocity. The

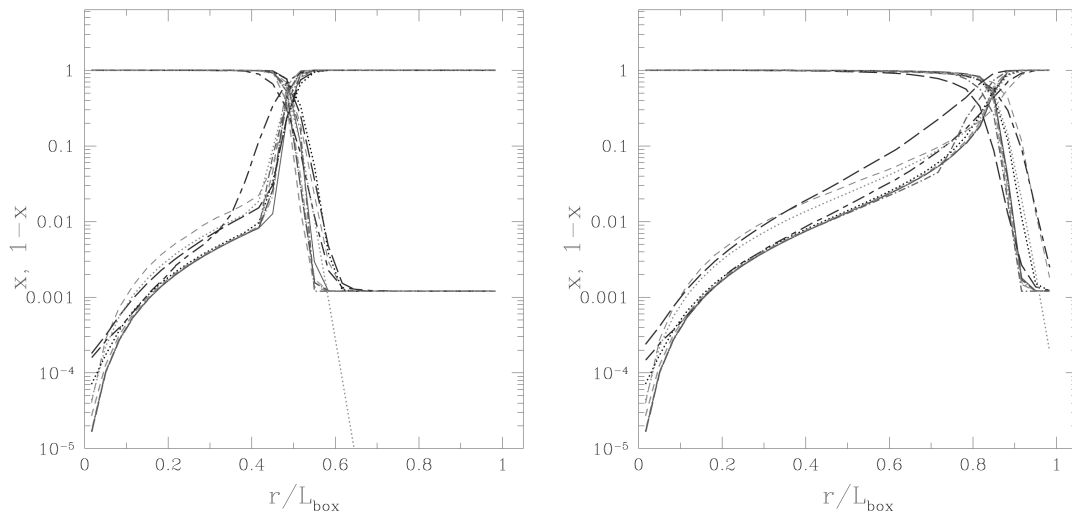


Fig. 9.4: Spherically averaged profiles for the ionized fraction x and the neutral fraction $x_{\text{HI}} = 1 - x$ at times $t = 30$ Myr (left) and 500 Myr (right) versus dimensionless radius (in units of the box size).

analytical results in Eqs.(9.8) and (9.9) are shown as well (*black, solid lines*). All codes track the ionization front correctly, with the position never varying by more than 5% from the analytical solution. These small differences are partly due to differences in the recombination rates used, and partly a consequence of our, somewhat arbitrary, definition of the ionization front position as the point of 50% ionization. The chosen parameters are such that the ionization front internal structure is well resolved, and the ionization front intrinsic thickness is larger than the discrepancies between the different codes. The ionization front velocities also show excellent agreement with the analytical result, at least until late times (i.e. after a few recombination times), at which point the ionization front essentially stops, and its remaining slow motion forward is impossible to resolve with the relatively coarse resolution adopted for the test. The ionization front is moving so slowly at this point that most of the remaining motion takes place with a single grid cell for extended periods of time, and thus fall below the resolution there.

In Fig. 9.4, we plot the spherically averaged radial profiles of the ionized and the neutral fraction. The left-hand panel shows these profiles at $t = 30$ Myr during the early, fast expansion of the ionization front. Most of the characteristics and ray-based methods (*C²-Ray*, *ART*, *FLASH-HC*, and *IFT*) agree very well at all radii. Again, *OTVET*, *CRASH*, and *SimpleX* find a somewhat thicker ionization front transition and lower ionized fraction inside the HII region. The *ZEUS* code also yields lower ionized fractions inside the HII region, mainly due to its slightly higher recombination coefficient. The *RSPH* code is intermediate between the two groups of codes, finding essentially the same neutral gas profile inside the HII region as the ray-based methods, but a slightly thicker ionization front, i.e. the ionized fraction drops more slowly ahead of the ionization front.

The same differences persist in the ionized structure of the final Strömgren sphere at $t = 500$ Myr (Fig. 9.4, right-hand panel). The majority of the characteristics codes again agree fairly well. The *IFT* code is based on the exact analytical solution of this particular problem, and thus to a significant extent could be considered a substitute for the analytical

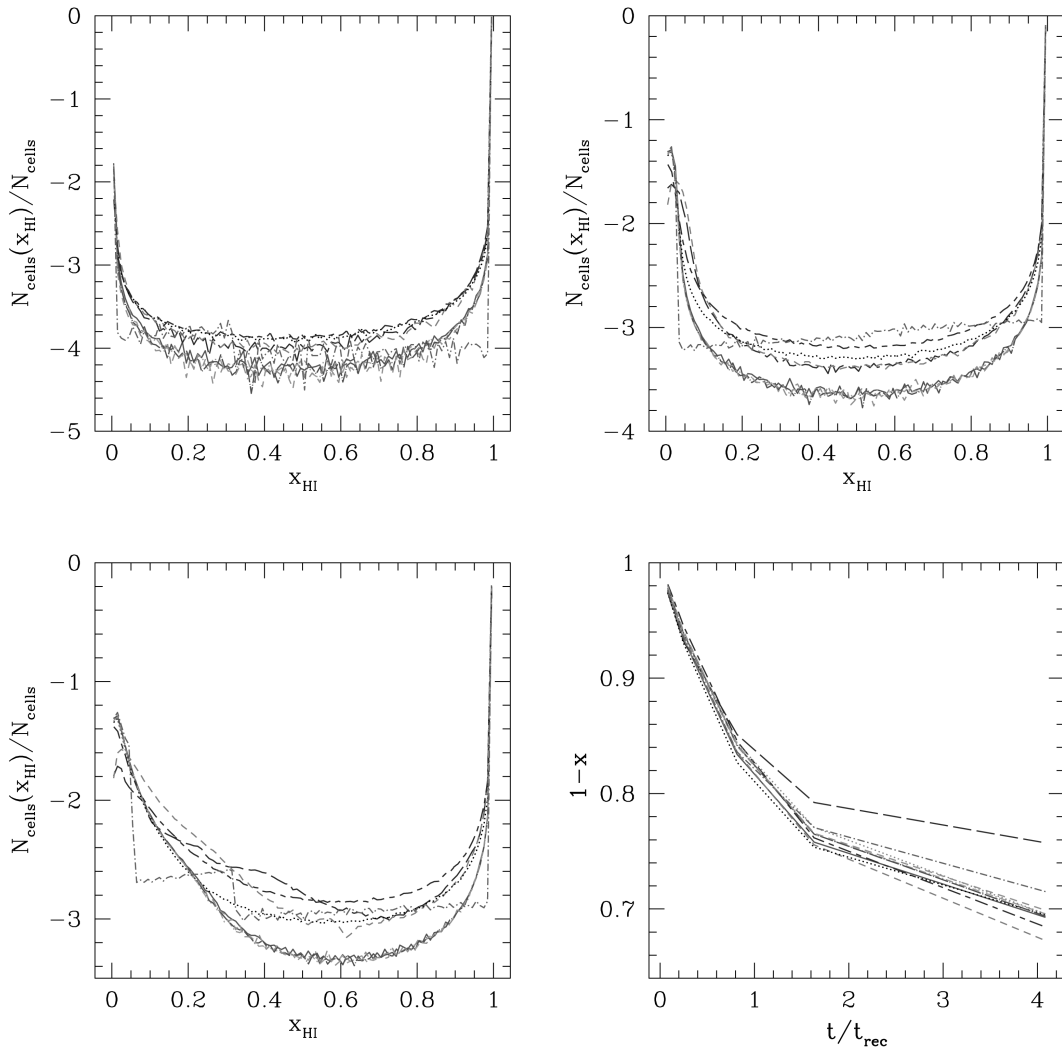


Fig. 9.5: Results for the classical ionization front test: the fraction of cells with a given neutral fraction, $x_{\text{HI}} = 1 - x$, at times $t = 10$ Myr (top-left), $t = 100$ Myr (top-right), and $t = 500$ Myr (bottom-left); also depicted is the global evolution of the total neutral fraction versus time (bottom-right).

HII region structure. Its differences from the exact solution are only close to the ionization front, where the nonequilibrium effects dominate. Away from the ionization front, however, the ionized state of the gas is in equilibrium and there all characteristics codes agree very well. The CRASH, OTVET, and SimpleX codes find thicker sphere boundaries and lower ionized fractions inside, but the first two codes find a slightly smaller Strömgren sphere, while our code finds a slightly larger one.

In Fig. 9.5, we show histograms of the fraction of cells with a given neutral fraction during the early, fast expansion phase (at time $t = 10$ Myr; top-left panel), when it starts slowing down ($t = 100$ Myr, close to the recombination time; top-right panel), and when the final Strömgren sphere has been reached ($t = 500$ Myr; bottom-left panel). These histograms reflect the differences in the ionization front transition thickness and internal

structure. All codes predict a transitional region of similar size, which contains a few percent of the total volume. In detail, however, once again the results fall into two main groups. One group includes most of the characteristics based methods, which agree perfectly at all times, and predict thin ionization fronts close to the analytic solution. The other group includes the more diffusive schemes, CRASH, OTVET, RSPH, and SimpleX, which all find somewhat thicker ionization fronts. During the expansion phase of the HII region, these four codes agree well among themselves, but they disagree somewhat on the structure of the final equilibrium Strömgren sphere, particularly in the proximity region of the source. The IFT code histograms differ significantly from the rest, due to its assumed equilibrium rates, which is not quite correct at the ionization front. Finally, the evolution of the globally averaged neutral fraction is shown in Fig. 9.5, bottom right. The same trends are evident, with the characteristics based methods agreeing well amongst themselves, while OTVET finds about 10% more neutral material at the final time, due to the different ionization structure obtained by this method. SimpleX agrees very well with the characteristics based methods, in contrast with the other statistical methods.

The classical problem of an ionization front expanding in a homogeneous neutral medium is one of the rare examples for which an analytic solution can be found. Comparing SimpleX to the analytic solutions, and the other cosmological radiative transfer codes, we find that our method solves for this problem very well. Although anisotropies are introduced by the stochastic noise of our method, mainly because of our use of a very different lattice, the overall results, expressed in the histograms and the spherically averaged plots, show that the SimpleX implementation, as given in the previous chapter, solves this basic problem very well. Of course, this is not all too surprising. As long as energy, or photons, are conserved, the speed of the ionization front will be correct. Indeed, the local interaction rules of our method were specifically designed to conform to this conservation rule.

9.4

A Cosmological Density Field

The final test we performed involves the propagation of ionization fronts from multiple sources in a static cosmological density field. This test case most closely resembled an actual reionization simulation. The initial condition is provided by a time slice, at redshift $z = 9$, from a cosmological PM+TVD code (Ryu et al. 1993). The simulation box size is $0.5 h^{-1}\text{Mpc}$ comoving, in which $h = 0.7$, the resolution is 128^3 cells, 2×64^3 particles. The halos in the simulation box were found using a *friends-of-friends halo finder* (see e.g. Neyrinck et al. 2005) with linking length 0.25. For simplicity, the initial temperature is fixed at $T = 100\text{K}$ everywhere. The ionizing sources are chosen so as to correspond to the 16 most massive halos in the box. We assume that these have a blackbody spectrum with effective temperature $T_{\text{eff}} = 10^5\text{K}$. The ionizing photon production rate for each source is constant, and is assigned assuming that each source lives $t_s = 3\text{ Myr}$ and emits $f_\gamma = 250$ ionizing photons per atom during its lifetime. Hence,

$$S_* = f_\gamma \frac{M\Omega_b}{\Omega_0 m_p t_s}, \quad (9.12)$$

in which M is the total halo mass, m_p is the proton mass, and we use $\Omega_0 = 0.27$ and $\Omega_b = 0.043$. For simplicity, all sources are assumed to switch on at the same time. The boundary conditions are transmissive, i.e. photons leaving the computational box are lost,

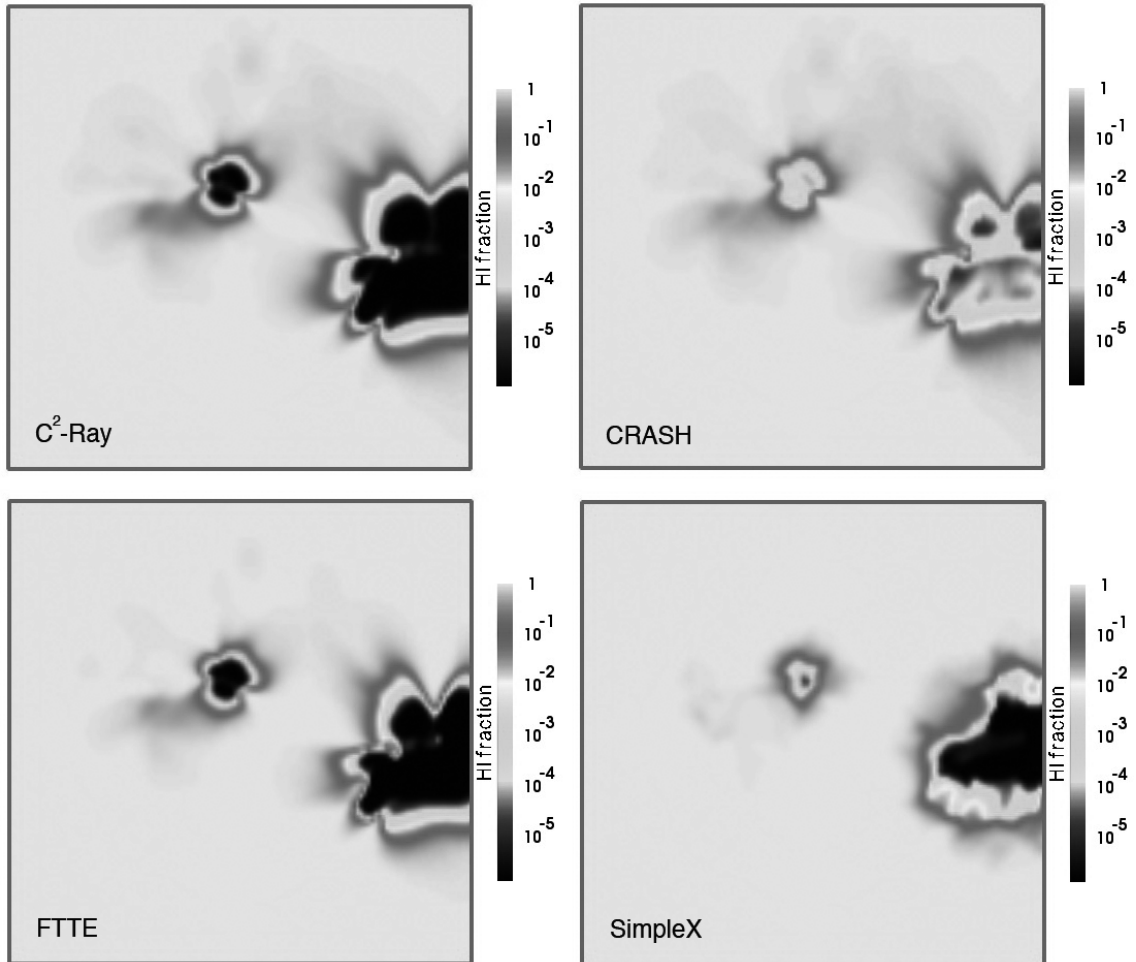


Fig. 9.6: Results for a test case of the reionization of a cosmological density field. Shown are images of the HI fraction, $x_{\text{HI}} = 1 - x$, cut through the simulation volume at coordinate $z = z_{\text{box}}/2$ and time $t = 0.05$ Myr. The blackbody spectrum of each source has an effective temperature of $T_{\text{eff}} = 10^5$ K.

rather than coming back in as in periodic boundary conditions. SimpleX can manage these boundary conditions by adding a boundary layer of absorptive points around the original box. The evolution time is $t = 0.4$ Myr.

In Fig. 9.6, we show slices of the HI fraction cut through the simulation box at coordinate $z = z_{\text{box}}/2$ and time $t = 0.05$ Myr. Fig. 9.7 shows the same, but at a time $t = 0.2$ Myr. Some discrepancies are already evident from a visual inspection, which shows somewhat different morphologies, but still general agreement.

More instructive is the temporal evolution of the volume-weighted (x_v ; thin lines) and mass-weighted (x_m ; thick lines) ionized fractions, depicted in Fig. 9.8, top-left. While CRASH and FTTE find comparable ionized fraction, C²-Ray and SimpleX produce slightly higher and lower values, respectively. The lower value in our code is obtained as a consequence of the temperature being fixed at 10^4 K, which is a little lower than that obtained by other codes, resulting in a higher recombination coefficient. In Fig. 9.8, we also show histograms of the ionized fraction at times $t = 0.05$ (top-right panel), 0.2 (bottom-left panel), and 0.4 Myr (bottom-right panel). As can be seen from these figures, the codes compare very well, and

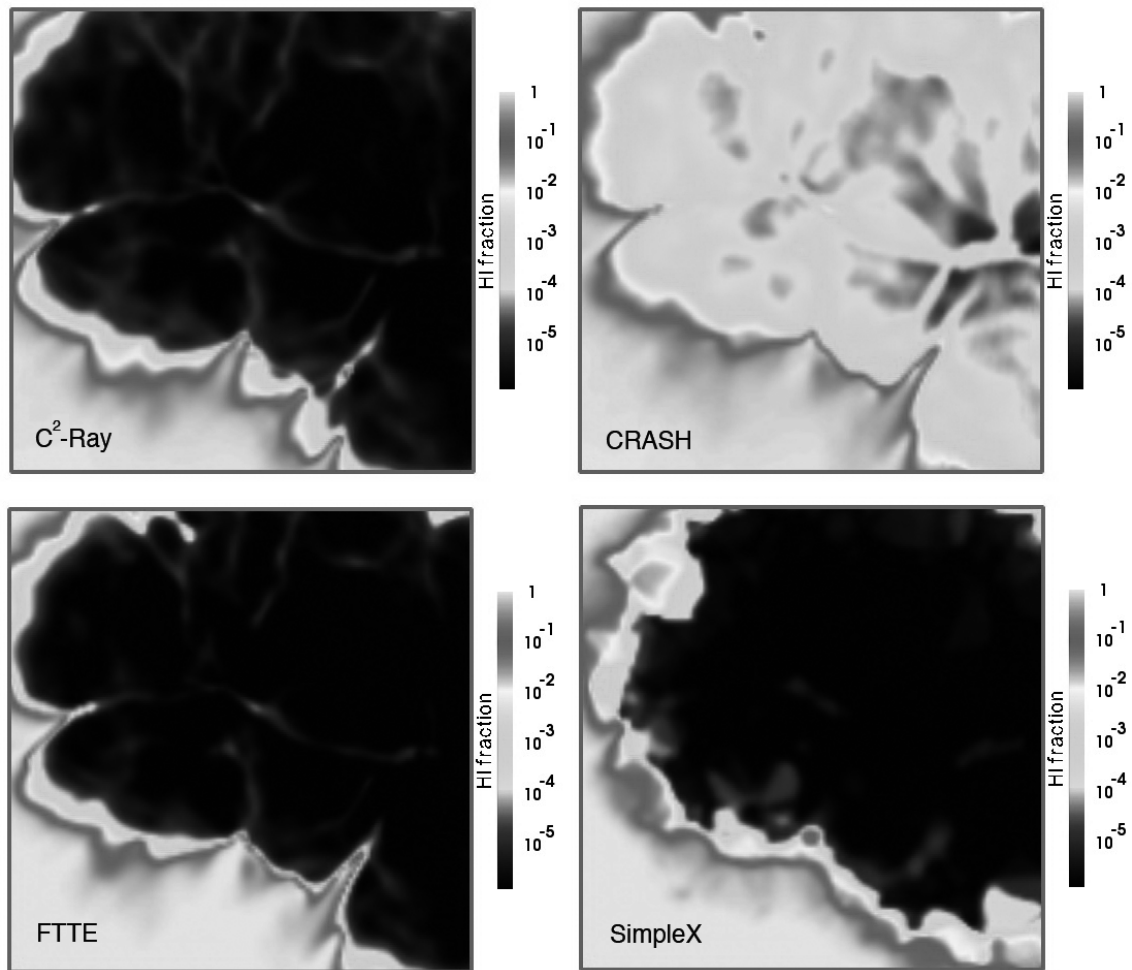


Fig. 9.7: Same as in Fig. 9.6, but now at time $t = 0.2$ Myr.

the origin of the differences is understood.

Although it did not matter much in the classical ionization front test in Sect. 9.3, where the medium was chosen to be homogeneous, our adaptive random lattice deviates drastically from the uniform 128^3 grid of the other three codes for this particular density field. The field, and thus also our lattice, is highly inhomogeneous, so that our method has a much more effective resolution in high density regions. Thus, the overall morphological differences between the results from the different codes in Figs.(9.6) and (9.7) may be caused by the low grid resolution of the uniform grids. In spite of this, the global ionized fractions and the histograms match very well, since these are a measure for the energy distribution. The morphological differences will become important when considering the overall topology of reionization, as will be discussed in Part III of this thesis.

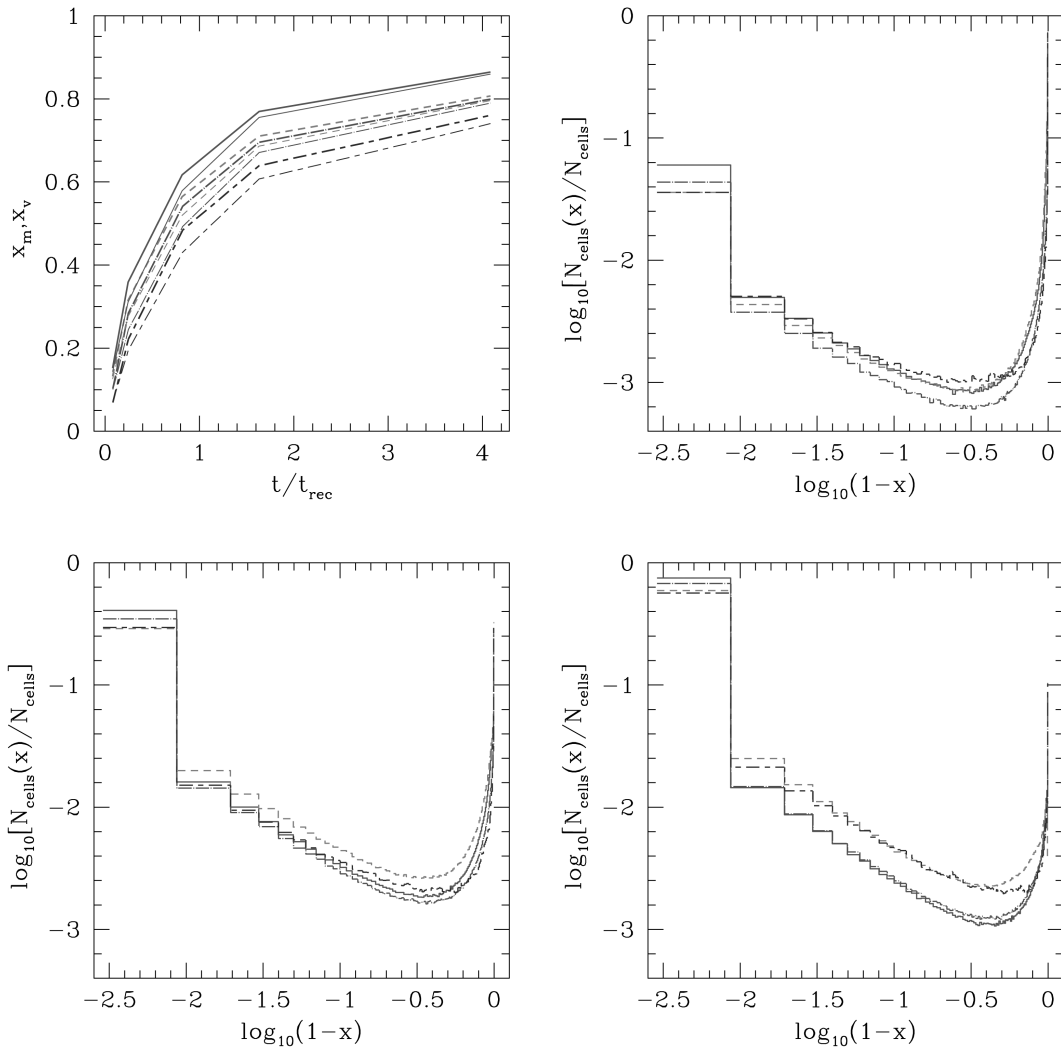


Fig. 9.8: Results for the test case of the reionization of a static cosmological density field. Shown in the top-left panel is the evolution of the volume- and mass-weighted ionized fractions, x_v (thin lines) and x_m (thick lines), for a blackbody source spectrum with $T_{\text{eff}} = 10^5 \text{K}$. Also shown are histograms of the neutral fraction, $x_{\text{HI}} = 1 - x$, at times $t = 0.05 \text{ Myr}$ (top-right), $t = 0.2 \text{ Myr}$ (bottom-left), and $t = 0.4 \text{ Myr}$ (bottom-right).

9.5

Conclusions

We presented a detailed comparison of our new cosmological radiative transfer method SimpleX with a large set of other cosmological radiative transfer methods, on several common tests. The participating codes represent a full variety of existing approaches, ranging from ray-casting methods to statistical ones, which solve the transfer equation on regular or adaptive meshes. This comparison is a collaborative project involving many cosmological radiative transfer researchers, and the result of this comparison is publicly available¹ for testing

¹<http://www.cita.utoronto.ca/~iliev>

of future codes during their development.

We began by comparing our basic physics and evaluated the effect these have on the ionization of a parcel of gas. It turned out that the results of the different codes match very well, and even errors associated with our approximation of assuming a fixed temperature turned out to be relatively minor. Also for the classical test of an ionization front propagating into a homogeneous neutral medium did our code produce results matching the other codes and the analytical results very well. Some anisotropy was introduced, but this was mainly due to the fact that we use a different lattice, which needed to be projected onto a regular grid. Finally, we studied the ionization of a cosmological density field by multiple (but a small number of) ionizing sources. The overall morphology of the HII regions shows good agreement at all times, and the evolution of the global ionized fractions agrees within $\sim 10\%$.

Several extra tests were done to compare methods that do a full multifrequency treatment of source spectra to methods like ours, that use a mean opacity approach. It turned out that the difference is negligible when doing large scale simulations of reionization, in which the internal structure of the propagating ionization fronts is not resolved. When pre-heating of the IGM by hard photons and X-rays is considered, multifrequency approaches must be used.

Many crucial things were not considered in these tests, though. First of all, of all methods only OTVET and SimpleX have incorporated periodic boundary conditions, and only SimpleX can do both periodic and non-periodic simulations. In order to do realistic large scale cosmological simulations, periodicity is mandatory. Secondly, no performance comparison was made at all. Of all methods, only OTVET and SimpleX do not scale with the number of sources. The reionization test only had a small number of sources, making the simulation feasible for all codes. Realistic reionization simulations would include millions of sources, drastically slowing down the other codes. Moreover, in all tests the on the spot approximation was assumed, in which diffuse photons are ignored. We will show in Chapter 11 that this approximation is not valid. In that case, the number of sources will increase beyond the reach of all other methods.

Even though our method uses a lattice that is very different from the one specified in these particular test cases, our SimpleX method compared very well to the results from other regular grid methods. SimpleX has several improvements on other methods, though. Our method includes periodicity, and does not scale with the number of sources, wherefore it is possible to do realistic large scale simulations including the often ignored diffuse photons. Knowing that SimpleX has these advantageous characteristics, we have used our adaptive transfer method to study several aspects of the Epoch of Reionization, as described in the final part of this thesis.

Acknowledgements We are very grateful to the Canadian Institute for Theoretical Astrophysics (CITA) and the Lorentz Center at Leiden University for their hospitality at the two workshops in 2005, which made this comparison project possible.

Bibliography

- Alvarez, M. A., Bromm, V., & Shapiro, P. R. 2006, ApJ, 639, 621
 Frenk, C. S., White, S. D. M., Bode, P., et al. 1999, ApJ, 525, 554
 Gnedin, N. Y. & Abel, T. 2001, New Astronomy, 6, 437
 Maselli, A., Ferrara, A., & Ciardi, B. 2003, MNRAS, 345, 379

- Mellema, G., Iliev, I. T., Alvarez, M. A., & Shapiro, P. R. 2006, *New Astronomy*, 11, 374
 Mellema, G., Raga, A. C., Canto, J., et al. 1998, *A&A*, 331, 335
 Nakamoto, T., Umemura, M., & Susa, H. 2001, *MNRAS*, 321, 593
 Neyrinck, M. C., Gnedin, N. Y., & Hamilton, A. J. S. 2005, *MNRAS*, 356, 1222
 Razoumov, A. O. & Cardall, C. Y. 2005, *MNRAS*, 362, 1413
 Rijkhorst, E.-J., Plewa, T., Dubey, A., & Mellema, G. 2006, *A&A*, 452, 907
 Ryu, D., Ostriker, J. P., Kang, H., & Cen, R. 1993, *ApJ*, 414, 1
 Spitzer, L. 1978, *Physical processes in the interstellar medium* (New York: Wiley-Interscience)
 Strömberg, B. 1939, *ApJ*, 89, 526
 Susa, H. 2006, *PASJ*, 58, 445
 Whalen, D. & Norman, M. L. 2006, *ApJS*, 162, 281

9.A

Solving the Rate Equation

When a parcel of hydrogen gas is irradiated by a constant flux of ionizing photons, as is the case in the first phase of the test in Sect. 9.2, we can solve the rate equation governing its ionization state analytically. To accomplish this, we have to assume that the ionized gas has a fixed temperature (e.g. 10^4K), so that the recombination coefficient can be assumed to be a constant. The rate equation for the ionization fraction $x = n_{\text{HII}}/n$ then has the form, cf. Eq.(9.1),

$$\begin{aligned}\frac{dx}{dt} &= a(1-x) - bx^2 \\ &= -b(x-x_+)(x-x_-),\end{aligned}\quad (9.13)$$

in which $a = A_p$ and $b = \alpha_B n$ are constants, and

$$x_{\pm} = \frac{a}{2b} \left(-1 \pm \sqrt{1 + 4b/a} \right). \quad (9.14)$$

Given the initial condition $x(0) = x_0$, Eq.(9.13) has the exact solution

$$x(t) = \frac{x_-(x_+ - x_0) - x_+(x_- - x_0)e^{b(x_+ - x_-)t}}{(x_+ - x_0) - (x_- - x_0)e^{b(x_+ - x_-)t}}, \quad (9.15)$$

which reduces to a simpler form when $x(0) = 0$,

$$x(t) = \frac{x_-x_+ - x_+x_-e^{b(x_+ - x_-)t}}{x_+ - x_-e^{b(x_+ - x_-)t}}. \quad (9.16)$$

One can immediately see from Eq.(9.16) that it has the asymptotic solution, obtained by using l'Hôpital's rule,

$$\lim_{t \rightarrow \infty} x = x_+. \quad (9.17)$$

When $a \gg b$, as is the case in Sect. 9.2, we can truncate the expansion of Eq.(9.14) with respect to b/a , i.e.

$$x_{\pm} = \frac{a}{2b} \left[-1 \pm \left(1 + \frac{2b}{a} - \frac{2b^2}{a^2} + \dots \right) \right], \quad (9.18)$$

so that

$$x_+ = 1 - \frac{b}{a} + \dots \rightarrow^{10} \log(1 - x_+) =^{10} \log \left(\frac{b}{a} \right). \quad (9.19)$$

This equation can be used to compute the equilibrium solution of the gas in the initial phase of the test in Sect. 9.2.

Part III

The First Light

[...looking up at the heavens...]

I wish God were alive to see this.

HOMER SIMPSON

The Epoch of Reionization

In this chapter, we give a concise introduction into the Epoch of Reionization, putting emphasis on the numerical side of trying to understand this intricate cosmological era. We discuss several physical and numerical aspects that have previously been neglected, and show that these may have a profound impact on the overall process of reionization. In the final part of this thesis, we will systematically study the effect of these aspects.

10.1

Introduction

At a redshift $z \sim 1000$, equivalent to about 400,000 years after the Big Bang, the expansion of the Universe had caused the primordial photon-baryon plasma to cool down to such a low temperature that locally the Compton scattering time for a photon exceeded the Hubble time. Thus, photons decoupled from the baryons, traveling onwards to form the *Cosmic Microwave Background* (CMB) for ages to come, while the electrons and protons were left to recombine. Left in an almost uniform and neutral state, the baryonic material in our Universe was only perturbed by the tiny density fluctuations still observable in the current high precision WMAP maps of the CMB (Hinshaw et al. 2006). Under the action of gravity, these initial perturbations collapsed into bound objects, giving rise to the formation of the first stars.

Several physical processes counteracted this formation process: the Jeans mass scale is finite ($M \approx 10^4 M_\odot$) and prevents the gas from following the collapse of the underlying dark matter component; the available molecules, such as H_2 , are fragile and have low cooling efficiency, and the standard cooling agents, such as heavier elements and dust, were absent; moreover, feedback effects, both radiative and mechanical, counteract the otherwise rapid formation of stars. Therefore, the Universe was in a relatively dark and quiet state for a fair amount of cosmic time. This epoch is denoted as *the Dark Ages*, after a suggestion by Sir Martin Rees. At a certain point in time, the end of the Dark Ages is heralded by the formation of the first generation of stars (dubbed Pop III), which are so massive that they produce a copious amount of energetic UV photons, enabling the ionization of the neutral hydrogen pervading the post-recombination Universe. This process of *reionization* can be thought of as the second major phase change in the ionization state of hydrogen (and also helium), and is believed to have ended at redshift $z \sim 6$ (Fan et al. 2004).

A theoretical description of this whole evolutionary process is rather involved, requiring not only an accurate description of the nonlinear collapse of the initial density perturbations into

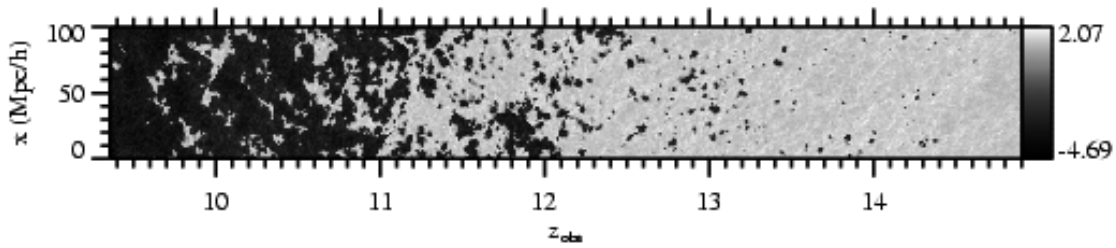


Fig. 10.1: Cosmic evolution during the EOR, depicted in the form of a 21-cm slice against redshift based on a numerical simulation from Mellema et al. (2006). Shown is the brightness temperature relative to the CMB. At an observed redshift of $z \sim 10$, the HII regions almost fill the computational domain: the reionization process has come to an end.

bound objects, but also an accurate hydrodynamical and chemodynamical description of the formation of the first stars and galaxies. Although this may still be somewhat tractable, the subject becomes even more drastically convoluted by the necessity to include an accurate description of radiative feedback. Excellent reviews on our current understanding of the physical processes involved in the formation of first stars and the process of reionization are given in Barkana & Loeb (2001), Loeb & Barkana (2001) and Ciardi & Ferrara (2005), and we will refrain from going into more detail here.

Cosmic reionization is, without doubt, of immense importance in modern-day cosmology, providing a direct link between cutting-edge theoretical efforts and pioneering observational experiments. On the one hand, for example, the process of reionization is an essential component in the study of structure formation. The emergent photons are a product of the formation of the first structures and luminous sources, but also affected subsequent structure formation that lead to the present day Universe. On the other hand, we are at the frontier of observing the relics of the reionization epoch. The study of the 21-cm hyperfine triplet-singlet level transition of the ground state of neutral hydrogen allows a superb tracing of the neutral hydrogen distribution during these early times (Madau et al. 1997). The prospect of studying reionization in this way has motivated several teams to construct arrays of low-frequency radio telescopes, for example the Low Frequency Array¹, and the Primeval Structure Telescope². Over the next few years, these arrays will search for 21-cm emission or absorption from redshifts $z \geq 6.5$, which will be observed redshifted to low frequencies corresponding to wavelengths of, say, 1.5 to 4 meters. An example of a 21-cm slice against redshift of the process of bubble overlap, based on a numerical simulation from Mellema et al. (2006), is given in Fig. 10.1.

The link between these opposite sides of trying to get a better understanding of the Epoch of Reionization (EOR) is provided by numerical studies. Structure formation is highly nonlinear, and gasdynamical and radiative transfer effects are too difficult to track analytically. Given the input, or initial conditions, as provided by the fluctuations in the CMB, we can use sophisticated hydrodynamics solvers to follow structure formation up until the formation of the first stars and galaxies. Once the first photons are produced, an accurate and efficient radiative transfer solver can provide information on the effects of radiative feedback, but it can also in principle produce the 21-cm maps, to be observed in the near future, by

¹<http://www.lofar.org>

²<http://web.phys.cmu.edu/~past/>

predicting the sizes and morphology of the overlapping HII regions being blown, as a function of redshift.

In the final part of this thesis, we will use our transport method, as explained in general terms in Part I of this thesis, and more specifically in the context of cosmological radiative transfer in Part II, to study both of these aspects, examining the effects of radiative feedback, but also analyzing the emergent morphology of the bubble overlap phase. In this chapter, we will give an overview of the ingredients essential to successfully tackle these problems numerically. We start with giving a brief review of these requirements in Sect. 10.2, after which we more specifically list the aims of our efforts in Sect. 10.3.

10.2

Requirements

An accurate and self-consistent numerical simulation of structure formation starting with the initial conditions as seen in the CMB until the formation of the first stars and galaxies, which reionize the Universe, is far from trivial. Indeed, it turns out to be a tremendously involved and formidable task. Accurate hydrodynamics solvers that implicitly solve for gravitational interaction are needed to follow the motion of the gas and dark matter as they evolve from the linear into the nonlinear regime. Chemical evolution routines have to be incorporated to resolve atomic and molecular cooling necessary for the formation of the first stars. And, finally, radiative transfer solvers that capture the interaction of photons, produced by those first stars, as they move through the initially neutral hydrogen, are mandatory. These photons are the essential ingredient in the process of reionization: they provide the feedback on structure formation, and they produce the HII regions surrounding the first sources.

10.2.1 Hydrodynamics

Hydrodynamical solvers that are able to follow the growth of the initial overdensities into collapsed haloes and, coupled to chemistry solvers, into the first sources are entering their springtime. Adaptive codes, such as the Lagrangian SPH method (Lucy 1977; Monaghan 1992), have been adapted for cosmological purposes, culminating in open source packages such as GADGET2 (Springel 2005), and are able to track the dynamical evolution of the large scale structure of baryonic and dark matter. Although they are known to provide inaccuracies, especially when dealing with strong shocks (see Chapter 6), they have as yet no rival in this cosmological context, and the overall results have been shown to agree fairly well with the more robust grid based methods.

10.2.2 Radiative Transfer

As discussed in Part II of this thesis, it is the radiative transfer that provides the current bottleneck in modern reionization simulations. The first sources are expected to form in high density regions, which are distributed throughout space along very inhomogeneous, filamentary structures. Moreover, they are certain to be large in number, and, to top it off, the medium with which the photons will interact is distributed very inhomogeneously as well. Thus, any transfer solver that needs a simple geometry for its medium distribution or one that scales with the number of sources is ruled out a priori.

10.2.3 Box Size and Resolution

If this not might seem worrying enough, there is the issue of box size and resolution. Simple analytical models that give a global description of the EOR (see e.g. Furlanetto et al. 2004; Wyithe & Loeb 2004) predict that the characteristic sizes of HII regions at the end of reionization is of the order of 10 Mpc and above. Thus, the size of the computational box should be at least of that order to give an accurate description of the process of bubble overlap. Moreover, Barkana & Loeb (2004) showed that large fluctuations in the statistics of galaxy formation demand box sizes of the order of 100 Mpc and above to resolve the long-wavelength modes of density fluctuations so as to avoid cosmic scatter.

On the other hand, high resolution is needed to resolve the formation and evolution of the first sources. Even with adaptive grid codes, full simulations of the formation of the first stars (e.g. Abel et al. 2002) are limited to tiny regions. Even when drastic sacrifices are made in the treatment of the relevant physics, by excluding chemistry and hydrodynamics, the large cosmological scales that need to be resolved cannot be reached (Naoz et al. 2006). A $10^5 M_{\odot}$ halo needs to be represented by at least 500 particles in order to even crudely track its evolution (Springel & Hernquist 2003). Given the relevant cosmological scales and the mean cosmic density, we obtain that a total of 10^{21} particles are needed to have large enough resolution to resolve the large scale density modes on the one hand, and to resolve the formation of stars in enough detail on the other. Given that the maximum resolution achieved thus far is of the order of 10^{10} particles (Springel et al. 2005), we are still far away from doing a full realistic simulation of the EOR.

10.3

Aims

A full simulation of the EOR that combines realistic hydro- and chemodynamics with a fast and accurate radiative transfer solver, encompassing not only the range but also the resolution needed, is far beyond our current computational capabilities. That does not mean that it is futile to numerically study the EOR at all; on the contrary: the problem can be reduced to smaller, more tractable, subproblems, studying the effect of certain isolated physical processes, or the approximate geometry of the HII patterns in smaller subboxes. The only thing to beware of is that one does not forget that one is studying an approximation, even though the results might seem convincingly realistic.

There have recently been several efforts that try to simulate the EOR on such large scales that one can accurately follow the process of HII region overlap. Iliev et al. (2006) and Zahn et al. (2006), for example, use an N-body solver to solve for the nonlinear evolution of the dark matter in a computational box that encompasses a comoving volume of the order of $100/h$ Mpc. An approximate description for the formation of the first sources predicts the flux of photons, and a radiative transfer solver is used to post-process the transport of this radiation. These simulations have contributed greatly to our understanding of the EOR, analyzing the growth and overlap of the HII regions in several different ways.

Unfortunately, there are still several shortcomings to their approach, inherent to the computational technique. Hydrodynamical feedback effects can not be resolved, because of the N-body only approach to the evolution of baryonic matter, eliminating a priori the chance to inspect effects like that of reheating. Moreover, the radiative transfer solvers used are the bottleneck for these simulations: they scale with the number of sources, making the

simulation as a whole so costly that one is limited to running only very few simulations at all. This makes it impossible to do convergence tests on the effect of box size and/or resolution, which are needed to check the reliability of the overall result. Moreover, the radiative transfer solvers use stiff regular grids of finite, relatively coarse resolution, undersampling regions of high density and rich detail. The effect of things like diffuse photons also has to be neglected, in order to prevent the number of sources to increase exponentially. Iliev et al. (2006) and Zahn et al. (2006) try to circumvent several of these problems by using approximate models for the unresolved physical processes, but it is, of course, better to solve for them directly.

It is the effect of these previously underemphasized aspects that we wish to study in the final part of this thesis.

10.3.1 The Different Aspects

The `SimpleX` transport method, described in the previous two parts of this thesis, does *not* scale with the number of sources. Using this method to do the radiative transfer in reionization simulation thus resolves the most important bottleneck. This enables us to more systematically check the impact and importance of these effects on the simulation as a whole, by repeating the same experiment over and over again, tweaking several of the physical effects and parameters. Moreover, the random lattice used by the `SimpleX` method is highly adaptive, so that much higher effective resolutions can be reached, resolving parts of high density regions that have been out of reach for more standard radiative transfer methods. As a bonus, `SimpleX`, in contrast with the transfer methods used in previous EOR simulations, can be used without and with periodic boundary conditions, the latter of which is of course most correct in a cosmological context.

We conclude this chapter by listing the three specific effects we are aiming at, each of which may have a profound impact on the overall result of the simulation.

Resolution and Box Size Effects

When using a cosmological SPH method, such as `GADGET2`, to solve for the structure formation, the resolution is determined by the number of SPH particles of the simulation. Once the size of the computational box is specified, we can compute the mass of each particle, assuming the masses of the particles have been chosen to be identical. This mass resolution directly determines the level of detail at which we can resolve high density regions, and it so happens that in these regions the first sources form. Moreover, the ionization balance during the EOR is given by a delicate interplay between the production of UV photons and the number of recombinations, cf. Eq.(7.8). The local recombination rate scales with n^2 , so that these same high density regions have the largest contribution to the total number of recombinations.

The maximum resolution that can be used is mostly a fixed number, dictated by the limits of the available hardware. Increasing the size of the computational box therefore immediately results in a rapid decrease of the mass resolution, so that high density regions will not be resolved properly. This has two consequences. While the number of sources formed in these high density regions may be large, the number of SPH particles present may not be sufficient to resolve each and every one of them. Furthermore, shot noise severely limits the validity of conclusions even for structures that consist of hundreds of particles. A simulation that has only few, very luminous sources will be the result, and this is of course not satisfactory, not in the least because the morphology of the HII regions will be different from what it

should be. The second problem is that most photons are consumed in high density regions, where the recombination rate is highest. When not properly resolving these regions, the recombination rate computed may be too low, and the structure of the emerging HII regions may be erroneous.

In Chapter 12, we systematically study the effect of changing the resolution and the box size, examining their impact on the star formation history, the number of recombinations per baryon, and the overall reionization history.

The Effect of Reheating

Numerical simulations such as the ones described in Iliev et al. (2006) and Zahn et al. (2006) use an N-body method to follow the evolution of gravitational collapse of dark matter. They hereby ignore any hydrodynamical effects that may occur during this process, ranging from shock heating to the effects of turbulence. Most of these effects are merely secondary, though. One effect that can not be ignored, however, is that of reheating.

When the process of reionization has almost finished, the HII regions are beginning to overlap, so that the mean free path of ionizing photons rapidly increases. Thus, the intensity of the cosmic ionizing background increases sharply, which causes the intergalactic medium to heat to a temperature of $\geq 10^4$ K. Two things happen: first, the increase in temperature causes the Jeans mass to increase dramatically, suppressing the total star formation rate, thereby quenching the total production of ionizing photons; second, the increase in temperature causes the gas to respond by counteracting the otherwise unhampered collapse. This will decrease the clumping of gas in high density regions, thereby lowering the total number of recombinations.

Both the effect on the formation of stars and photons, and the effect on the clumping and recombination rate, will have a profound impact on the global behavior of the intergalactic medium during the EOR. As will be described in Chapter 12, we will couple our transfer code `SimpleX` to a variant of the cosmological SPH code `GADGET2`, so that we will be able to study the effect of reheating.

Diffuse Photons

The final effect we have studied is that of the diffuse radiation field. Ionized gas has a finite recombination time ($t_{\text{rec}} \propto n^{-1}$). This means that the baryons within an HII region recombine and thereby produce photons that are energetic enough to contribute to the overall ionization field. An accurate treatment of this diffuse radiation field would therefore need a radiative transfer code that solves for the propagation of photons from each location within the HII regions. Unfortunately, most radiative transfer solvers scale with the number of sources (cf. Chapter 7), so that the computations would become prohibitively slow.

In order to circumvent this problem, most transfer codes implement an often used simplification. This so-called ‘On The Spot Approximation’ (Spitzer 1978; Osterbrock 1989) assumes that all diffuse UV photons are immediately absorbed very near their origin, so that the recombinations directly into the ground state can effectively be ignored. In Chapter 11, we show that this simplification is far from accurate, so that it is mandatory to incorporate the diffuse radiation field into simulations of the EOR. Moreover, we show that the inclusion of these diffuse photons will alter the overall morphology of anisotropic HII regions.

The `SimpleX` transfer code does not scale with the number of sources, so that it no effort at all to incorporate the diffuse ionization field. Thus, this enables us to study the

effect of diffuse photons on the morphology of the structure of overlapping HII regions, and we refer the reader to Chapter 12, in which the results are presented.

Bibliography

- Abel, T., Bryan, G. L., & Norman, M. L. 2002, *Science*, 295, 93
Barkana, R. & Loeb, A. 2001, *Phys. Rep.*, 349, 125
—. 2004, *ApJ*, 609, 474
Ciardi, B. & Ferrara, A. 2005, *Space Science Reviews*, 116, 625
Fan, X., Hennawi, J. F., Richards, G. T., et al. 2004, *AJ*, 128, 515
Furlanetto, S. R., Zaldarriaga, M., & Hernquist, L. 2004, *ApJ*, 613, 1
Hinshaw, G., Nolta, M. R., Bennett, C. L., et al. 2006, *ArXiv Astrophysics e-prints*, astro-ph/0603451
Iliev, I. T., Mellema, G., Pen, U.-L., et al. 2006, *MNRAS*, 369, 1625
Loeb, A. & Barkana, R. 2001, *ARA&A*, 39, 19
Lucy, L. B. 1977, *AJ*, 82, 1013
Madau, P., Meiksin, A., & Rees, M. J. 1997, *ApJ*, 475, 429
Mellema, G., Iliev, I. T., Pen, U.-L., & Shapiro, P. R. 2006, *MNRAS*, 372, 679
Monaghan, J. J. 1992, *ARA&A*, 30, 543
Naoz, S., Noter, S., & Barkana, R. 2006, *MNRAS*, 373, L98
Osterbrock, D. E. 1989, *Astrophysics of gaseous nebulae and active galactic nuclei* (Mill Valley, CA: University Science Books)
Spitzer, L. 1978, *Physical processes in the interstellar medium* (New York: Wiley-Interscience)
Springel, V. 2005, *MNRAS*, 364, 1105
Springel, V. & Hernquist, L. 2003, *MNRAS*, 339, 312
Springel, V., White, S. D. M., Jenkins, A., et al. 2005, *Nature*, 435, 629
Wyithe, J. S. B. & Loeb, A. 2004, *Nature*, 432, 194
Zahn, O., Lidz, A., McQuinn, M., et al. 2006, *ArXiv Astrophysics e-prints*, astro-ph/0604177

The Diffuse Nature of Strömgren Spheres

In this chapter, we argue that the standard analytical derivations of properties of HII regions, such as the speed, shape and asymptotic position of ionization fronts require a more precise treatment. These derivations use the *on the spot approximation*, which ignores the diffuse component of the radiation field. We show that, in fact, HII regions are diffusion dominated. This has as a result that the morphology of inhomogeneous HII regions will be drastically different, because shadowing effects have a less profound impact on the apparent shape. Moreover, it will have influence on the propagation speed of ionization fronts. We quantify our claims by analytically deriving the internal radiation structure of HII regions, taking diffusion fully into account, for several different cosmologically relevant density distributions, and we check these results by doing a full radiative transfer treatment. Several effects that may diminish the consequences of the diffuse radiation field are discussed.

The Diffuse Nature of Strömgren Spheres

J. Ritzerveld

Astronomy & Astrophysics (2005), **439**, L23-L26

11.1

Introduction

Ionized hydrogen is produced when hot stars emit UV radiation that ionizes the surrounding interstellar or intergalactic medium. Recently, the interest in the resultant HII regions has been rekindled by current developments in the field of cosmological astrophysics. When the first hot and massive Pop III stars form, the energetic photons expelled blow ionization bubbles that will overlap, filling the Universe at the end of the Epoch of Reionization (EOR). It is therefore not surprising that much effort has been put into the physical, analytical and numerical understanding of the mechanisms involved.

One of the more relevant properties of HII regions is the structure of the diffuse radiation field, as this will influence not only the dynamics of the expansion of the ionization front, but also the morphology of shadows formed within the region. It is well known that the diffuse component of the overall ionizing radiation field dominates in certain parts of the HII region, but for practical reasons this is often ignored under the assumption that it is a small effect. In this chapter, we quantitatively show that it is a non-negligible effect, especially in the presence of density gradients. This conclusion will have a profound effect on the dynamics

of expanding ionization fronts, but also on the morphology of inhomogeneous HII regions, the 21cm-signature of which is to be observed in the near future.

Much of the analytical work done on HII regions heavily depends on the use of the so-called *on the spot approximation*, part of which has its origin in Baker & Menzel (1938) and Zanstra (1951). It has been discussed in the standard texts Spitzer (1978) and Osterbrock (1989), and, consequently, it has been considered a standard and valid approximation ever since. In this argument, one considers the gas behind the ionization front, or within the Strömgren sphere when the static equilibrium solution has already been reached, and does some bookkeeping to equate the number of recombinations within that volume to the number of ionizing photons. Electron captures directly into level $n = 1$, parametrized by the coefficient $\alpha_1(T)$, produce photons energetic enough to ionize another atom (the spectrum of these diffuse photons can be approximated by a delta function just above the Lyman limit). The ionization balance can be drawn up as follows:

$$4\pi r^2 n(r) \frac{dr}{dt} = S_* + 4\pi \alpha_1(T) \int n^2(r') r'^2 dr' - 4\pi \alpha_A(T) \int n^2(r') r'^2 dr', \quad (11.1)$$

in which r is the position of the front and S_* is the number of ionizing photons (i.e. with frequencies $\nu \geq \nu_0$ above the Lyman limit threshold) emitted by the source per second. The second term on the rhs is the total number of diffuse photons, given the HI density distribution $n(r)$. The last term on the rhs accounts for the total number of recombinations to every level, parametrized by the recombination coefficient $\alpha_A(T)$. The integrations are over the whole HII region. If the Strömgren radius has already been reached, the three contributions on the rhs of Eq.(11.1) cancel, but up until that point the ionization front has a finite speed dr/dt .

Unlike the source photons, which are directed radially outwards, the diffuse photons may cross the nebula in any direction from their point of creation. Thus, one has to do a full treatment of radiative transfer, which complicates matters drastically. This is where the on the spot approximation enters the picture. For typical HII regions, the physical parameters are such that the optical depth for photons near the Lyman limit is ~ 30 (cf. Osterbrock 1989). Thus, one argues that the diffuse photons produced by recombinations to the ground level will be re-absorbed very close to where they were produced ('on the spot'). This has as a result that effectively the recombinations directly to the ground level do not count, because they are exactly balanced by the photons they produce. Henceforth, one can ignore these recombinations, and the diffuse photons created by them, and use the effective *Case B* (using terminology introduced by Baker & Menzel 1938) recombination coefficient

$$\alpha_B(T) = \alpha_A(T) - \alpha_1(T), \quad (11.2)$$

with Eq.(11.1) reducing to

$$4\pi r^2 n(r) \frac{dr}{dt} = S_* + 4\pi \alpha_B(T) \int n^2(r') r'^2 dr'. \quad (11.3)$$

The Strömgren sphere radius can be easily obtained from Eq.(11.3) by considering the equilibrium solution with zero velocity of the ionization front $dr/dt = 0$.

In the past, the effect of the diffuse radiation field has been examined by relaxing the on the spot approximation somewhat. See, for example, the classical papers of Hummer & Seaton (1963), and Rubin (1968), but also that of Hong & Sung (1989). More recent analyses on the effects of the diffuse radiation field have been given in e.g. Canto et al.

(1998), and López-Martín et al. (2001). Every one of these incorporate the effect of the diffuse field by determining its local magnitude from the value of the local stellar ionizing field as obtained by using the on the spot approximation, i.e. (López-Martín et al. 2001)

$$\frac{F_d}{F_*} = \frac{\alpha_1}{4\alpha_B\kappa}, \quad (11.4)$$

in which $\kappa = \bar{a}_d/\bar{a}_*$ is the ratio of the mean hydrogen absorption coefficients (cf. Sect. 7.3.3), averaged over the diffuse photon spectrum and the stellar one, respectively. It is obvious that taking this particular approach to validate the use of the on the spot approximation is a first step in a circular reasoning process.

In this chapter, we will argue that it is not permitted to use the on the spot approximation for anything more than just calculating the radius of the Strömgren sphere. The first and foremost reason: the on the spot approximation is based on the argument that the mean free path for Lyman limit photons is very small, and that thus all diffuse photons are locally re-absorbed. But this argument is not correct, because the mean free path for the source photons (assuming these are also near the Lyman limit) is just as small. Thus, the diffuse and the source photons are on equal footing energy-wise, and no distinction between them can be made, with the exception of their directionality. Thus, at each location r within the HII region, not only are the diffuse photons used to compensate for the α_1 recombinations, but also a weighted fraction of the source photons, by which a fraction of the diffuse photons can actually escape. In effect, a fraction of the mono-directional source photons is converted into diffuse radiation. It is important to stress that we define the diffuse radiation field to consist of photons that were emitted at every location except that of the source itself. Once created by a recombination directly to the ground level somewhere in the HII region, a diffuse photon may or may not interact with the gas until reaching the front. Following this line of argument, one can see that the number of directional source photons decreases, when moving outwards from the source, until at a certain point the diffuse radiation starts to dominate.

In Sect. 11.2, we will analytically derive the diffuse versus mono-directional structure of the radiation within a Strömgren sphere for several astrophysically important density distributions. We will double-check these results by using our `SimpleX` code to solve for the real radiative transfer in these problems in Sect. 11.3. We end this chapter by discussing the impact of several effects that may lessen the role of the diffuse part of the radiation field in Sect. 11.4.

11.2

Analytic Derivation of the Diffuse Radiation Field

In this section, we will derive analytically the distribution and relative intensity of the diffuse and stellar components of the radiation field within an HII region. In order to make analytic derivations possible, we assume a hydrogen-only matter distribution that is isotropic with respect to the source. We only consider an HII region that is already in equilibrium, i.e. the Strömgren radius has already been reached. We also follow the classic approach in Strömgren (1939), assuming an infinitely sharp ionization front, separating the fully ionized HII region from its fully neutral surroundings. A more realistic approach, incorporating the thickness of the front, will be given in the next section. In Appendix 11.A, we give the exact analytic derivation for every \mathbb{R}^d , and every possible isotropic density distribution. Here, we

concentrate on the most relevant ones, in 3D space: a homogeneous, an r^{-1} , and an r^{-2} density distribution.

11.2.1 Homogeneous HI Matter Distribution

In what follows, we assume a homogeneous hydrogen medium density $n_{\text{HI}}(r) = n_{\text{HI}}$, a temperature $T = 10^4\text{K}$ for the ionized plasma, and, following the considerations that led up to the on the spot approximation, we assume that the mean free path of the photons is very small, until locally the gas has been fully ionized. Here, we perform the calculations for a Strömgren sphere in \mathbb{R}^d , because we can use the general result in the following subsection.

Because the Strömgren radius has already been reached, we know that there is an equilibrium between the number of recombinations per timestep and the total number of ionizing photons emitted by the source and the gas, so we may drop the time dependence. The number of recombinations to *every* level and the number of diffuse photons produced within a spherical shell at distance r is

$$\begin{cases} N_{\text{rec}}(r) &= \xi(d)\alpha_A n^2 r^{d-1} dr \\ N_{\text{diff}}(r) &= \xi(d)\alpha_1 n^2 r^{d-1} dr, \end{cases} \quad (11.5)$$

in which $\xi = 2\pi^{1/2}/\Gamma(d/2)$ ($= 4\pi$ in 3D) is just a geometrical factor and $\Gamma(x)$ is the Gamma function. The number of photons emitted by the source plus the number of diffuse photons within the whole region must compensate all the recombinations, from which we obtain

$$\begin{aligned} S_* &= \xi(d)\alpha_B n^2 \int_0^{R_S} r'^{d-1} dr' \\ &= \frac{\xi(d)}{d} \alpha_B n^2 R_S^d, \end{aligned} \quad (11.6)$$

in which R_S is the Strömgren radius.

If we define $I_s(r)$ and $I_d(r)$ as the number of source and diffuse photons, respectively, left at radius $r \leq R_S$, we can set up the following system of coupled differential equations on the domain $0 \leq r \leq R_S$

$$\begin{cases} \frac{dI_s(r)}{dr} &= -\xi(d)\alpha_A n^2 r^{d-1} \frac{I_s(r)}{I_s(r)+I_d(r)} \\ \frac{dI_d(r)}{dr} &= -\xi(d)\alpha_A n^2 r^{d-1} \frac{I_d(r)}{I_s(r)+I_d(r)} + \xi(d)\alpha_1 n^2 r^{d-1} \end{cases} \quad (11.7)$$

The first term on the rhs of both equations is the total number of recombinations being compensated for via a weighing term, that takes care of the fact that, if there are more source than diffuse photons present, the recombinations will mainly be compensated by the source photons, and vice versa. This system of equations is closed via the initial conditions $I_d(0) = 0$ and $I_s(0) = S_*$. We should note that we have used a symmetry condition here, in the sense that the diffuse radiation sent into the opposite direction (away from the front) is exactly balanced by the diffuse radiation on the other side of the source.

If we add the two equations, the weighing terms add up to unity, and using the initial conditions, we easily obtain an equation for the total radiation field

$$(I_s + I_d)(r) = \frac{\xi(d)}{d} \alpha_B n^2 R_S^d \left[1 - \left(\frac{r}{R_S} \right)^d \right]. \quad (11.8)$$

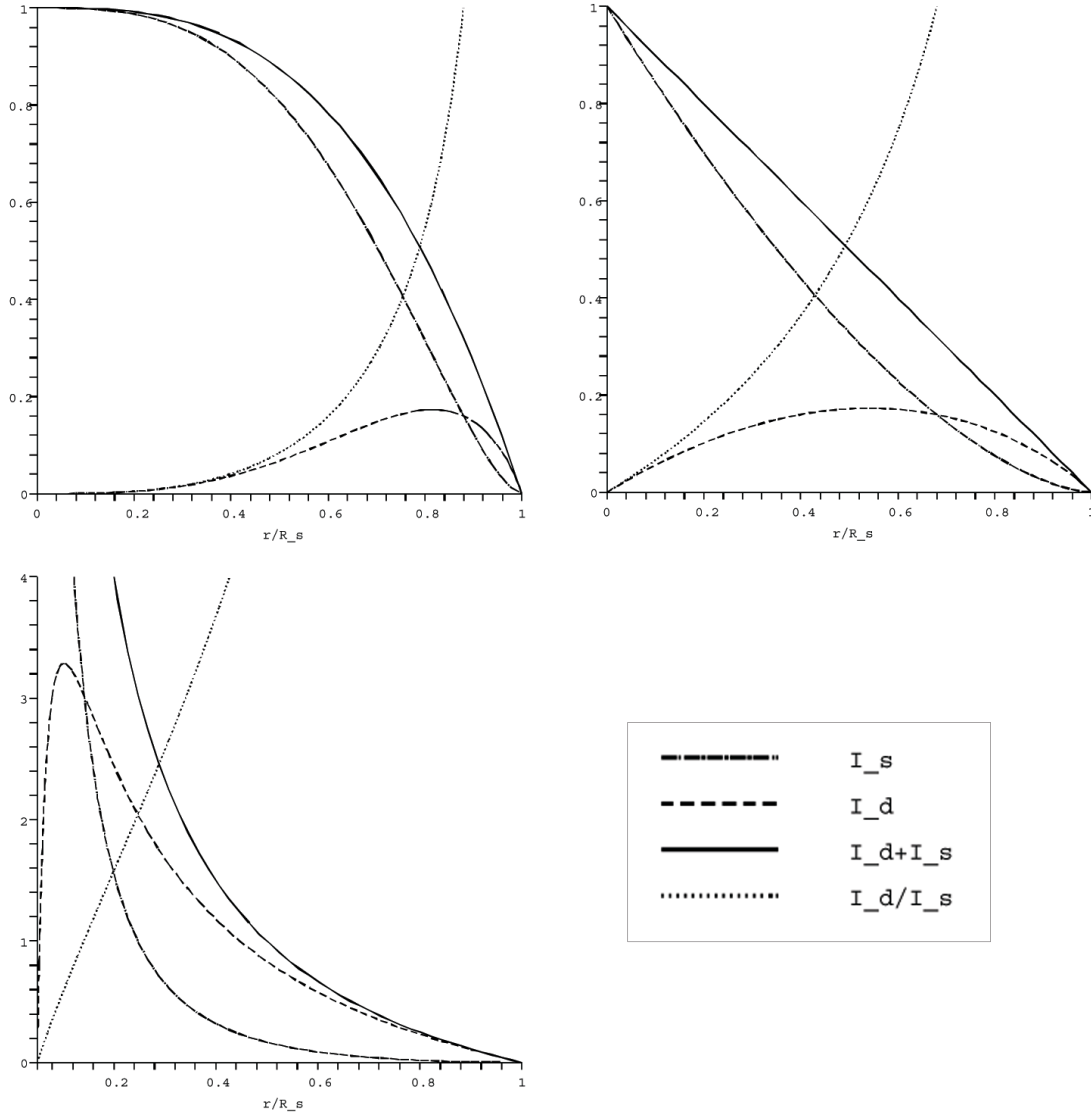


Fig. 11.1: Graphs of the functions $(I_s + I_d)(r)$ (solid), I_s (dashed-dotted), I_d (dashed), and I_d/I_s (dotted) for a homogeneous (top-left), an r^{-1} (top-right), and an r^{-2} (bottom-left) medium distribution in \mathbb{R}^3 .

We can plug this into the weighing terms in Eqs.(11.7), and upon integrating we obtain

$$I_s(r) = (I_s + I_d)(r) - I_d(r) = \frac{\xi^{(d)}}{d} \alpha_B n^2 R_S^d \left[1 - \left(\frac{r}{R_S} \right)^d \right]^{\alpha_A/\alpha_B}. \quad (11.9)$$

After choosing units such that $\frac{\xi^{(d)}}{d} \alpha_B n^2 R_S^d = 1$, and using the ratio $\alpha_A/\alpha_B = 4.18/2.60$ at 10^4K (e.g. Osterbrock 1989), we can draw a figure that plots $(I_s + I_d)(r)$, $I_s(r)$, $I_d(r)$ and the ratio $I_d(r)/I_s(r)$. This has been done for \mathbb{R}^3 in Fig. 11.1, top-left. One can see that at a certain radius r_{eq} the diffuse radiation component takes over. We can derive r_{eq} exactly by solving $I_s(r) = I_d(r)$, and obtain

$$r_{\text{eq}} = R_S \left(1 - 2^{-\alpha_B/\alpha_1} \right)^{1/d}. \quad (11.10)$$

Filling in the values results for \mathbb{R}^3 in $r_{\text{eq}}/R_S = 87.6\%$. Thus, we can immediately conclude that for a homogeneous gas distribution the outer 12% of the HII region is dominated by diffuse radiation, and that very near the front only diffuse photons are present. This means that, in terms of volume, 33% of the HII region is diffusion dominated. This effect is even greater for higher temperatures, for which the ratio α_B/α_1 is smaller.

11.2.2 An r^{-1} HI Matter Distribution

The effect described in the previous section is even more drastic for an HI density distribution in the form $n(r) = n_c(r/r_c)^{-1}$. This density profile has been used to describe the central regions of a dark matter halo (Navarro et al. 1997).

Following the same arguments as in the previous subsection, we obtain for \mathbb{R}^3 ,

$$\begin{cases} N_{\text{rec}}(r) &= 4\pi\alpha_A n_c^2 r_c^2 dr \\ N_{\text{diff}}(r) &= 4\pi\alpha_1 n_c^2 r_c^2 dr \\ S_* &= 4\pi\alpha_B n_c^2 r_c^2 R_S, \end{cases} \quad (11.11)$$

which are similar to Eqs.(11.5) and (11.6) for $d = 1$, but with a different constant factor. Thus, we can use Eqs.(11.8), (11.9) and (11.10), and obtain

$$\begin{cases} (I_s + I_d)(r) &= 4\pi\alpha_B n_c^2 r_c^2 R_S \left(1 - \frac{r}{R_S}\right) \\ I_s(r) &= 4\pi\alpha_B n_c^2 r_c^2 R_S \left(1 - \frac{r}{R_S}\right)^{\frac{\alpha_A}{\alpha_B}} \\ r_{\text{eq}} &= R_S \left(1 - 2^{-\alpha_B/\alpha_1}\right). \end{cases} \quad (11.12)$$

The graphs of these functions are plotted in the top-right panel of Fig. 11.1. Filling in the value for α_B/α_1 at $T = 10^4\text{K}$, gives $r_{\text{eq}}/R_S = 67.2\%$, from which we can conclude that for an isotropic r^{-1} distributed HII region the outer 33% is diffusion dominated, which is equivalent to roughly 70% of the total volume.

11.2.3 An r^{-2} HI Matter Distribution

The effect is strongest for an HI density distribution in the form $n(r) = n_c(r/r_c)^{-2}$, which has, for example, been used to describe the density profile of halos just interior to the virial radius and of collapsing isothermal spheres. In the derivation for this density distribution, we have to be more careful, because the equivalents of Eqs.(11.5) and (11.6) will contain singularities at $r = 0$, which can be resolved by choosing an inner edge $0 \leq r_c \leq R_S$. Thus, we obtain

$$\begin{cases} N_{\text{rec}}(r) &= 4\pi\alpha_A n_c^2 r_c^4 r^{-2} dr \\ N_{\text{diff}}(r) &= 4\pi\alpha_1 n_c^2 r_c^4 r^{-2} dr \\ S_* &= 4\pi\alpha_B n_c^2 r_c^4 (r_c^{-1} - R_S^{-1}). \end{cases} \quad (11.13)$$

Using similar steps as in the previous subsections, we obtain

$$\begin{cases} (I_s + I_d)(r) &= 4\pi\alpha_B n_c^2 r_c^4 (r^{-1} - R_S^{-1}) \\ I_s(r) &= 4\pi\alpha_B n_c^2 r_c^4 (r_c^{-1} - R_S^{-1})^{1 - \frac{\alpha_A}{\alpha_B}} (r^{-1} - R_S^{-1})^{\frac{\alpha_A}{\alpha_B}} \\ r_{\text{eq}} &= R_S [1 + 2^{-\alpha_B/\alpha_1} (R_S/r_c - 1)]^{-1}. \end{cases} \quad (11.14)$$

The graphs of these functions are plotted in the bottom-left panel of Fig. 11.1 for $r_c \leq r \leq R_S$. Taking a typical value of $r_c/R_S = 0.05$, we find that $r_{\text{eq}}/R_S = 13.8\%$, which means that, in terms of volume, 99.7% of the HII region is diffusion dominated.

11.3

Numerical Analysis

The original argument presented in Strömgren (1939) made several simplifying assumptions. Effectively, Strömgren ignored all radiative transfer effects, and used a smart bookkeeping trick to solve a jump equation in the form of Eq.(7.8) or Eq.(11.1). In the previous section, we extended this line of thought by specifically keeping track of all distinct components of the ionizing photon number, solving for both the stellar and the diffuse radiation field simultaneously. Unfortunately, our extension has the same limitation as the approach in Strömgren (1939): it can only be used to find the solution for an HII region in equilibrium.

In an attempt to relieve this restriction, and to double-check the results obtained analytically, a full radiative transfer treatment of the problem is needed. As discussed in Chapter 7, this used to be a major struggle for most radiative transfer methods: they scale with the number of sources, and, solving for the diffuse photons, every cell within the HII region becomes a source, increasing the operation count of the method by a factor 10^6 for a 128^3 cells domain. Fortunately, our `SimpleX` package, described in Chapter 8, has the useful property that it does *not* scale with the number of sources. We therefore took the approach of repeating the classic ionization front test of the comparison project described in Sect. 9.3, but this time *without* using the on the spot approximation, treating the diffuse and stellar photons on equal footing. To accomplish this, we use the same parameters as the test described in Sect. 9.3, but, using `SimpleX`, we do not only transport stellar, but also diffuse photons, produced by α_1 recombinations. Local ionizations are not solely induced by the stellar field, but also by the diffuse field, and this is incorporated by using the two weighing terms used in Eq.(11.7), making a local trade-off between the two types of photons available.

The result of this experiment is depicted in Fig. 11.2. The top-left panel shows the intensity (i.e. the number of photons present in a infinitesimally thin shell at a particular radius) of both diffuse and stellar, and the sum thereof, as a function of the distance to the source. Comparing this plot to what was analytically derived in Fig. 11.1, top-left, we see that the similarity is striking. There is some difference at the inner boundary, due to the numerical noise near the source, but most distinct is the difference at the location of the Strömgren sphere radius (indicated by the dashed line). In contrast with the analytical solution, the intensities do not drop to zero at R_S , but are smeared out. This is obviously because of the inherent thickness of the ionization front in this test problem, that can not be resolved in our analytical approximation. Zooming in on the region around the Strömgren sphere, cf. the top-right panel Fig. 11.2, we more clearly see the effect of the HII region boundary. But, we also see that the radius at which the diffuse part of the radiation field takes over is indeed very close to what we analytically derived, cf. Eq.(11.10), $r_{eq}/R_S = 0.876$. This turnover point is also visible in the bottom-left panel of Fig. 11.2, in which the diffuse and stellar fraction of the total radiation field are plotted, together with the diffuse-stellar photon ratio. Note that this final result also agrees very well with what was analytically derived.

The results given so far have been for the radiation field structure in the equilibrium state, when the Strömgren radius has already been reached. In this equilibrium case, our analytical solution obviously agrees with the numerical solution. Our analytical approach, however, can not be extended to a time-dependent scenario, in its present form. It is therefore worthwhile to use `SimpleX` to study how the diffuse-stellar photon ratio evolves in time, when the ionization front expands, moving towards its equilibrium solution. The result hereof is

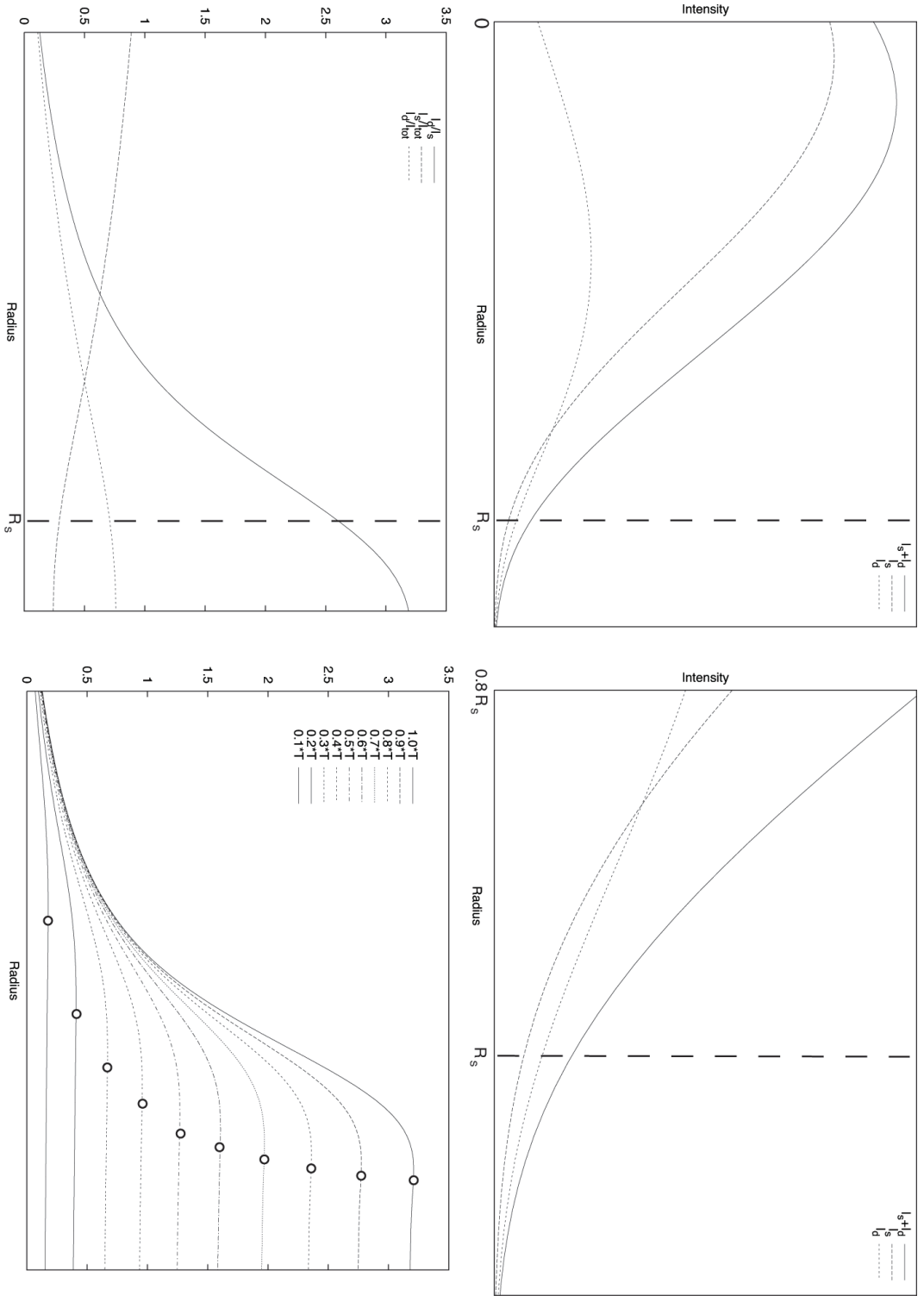


Fig. 11.2: Results for the homogeneous density profile. Top-left: Plot depicting the diffuse (I_d) and stellar part (I_s) of the total radiation field ($I_d + I_s$) versus the distance to the source. Top-right: Zoom in of top-left plot. Bottom-left: Plots of the fractional diffuse (I_d/I_{tot}) and stellar photons (I_s/I_{tot}), and of the diffuse-stellar photon ratio (I_d/I_s). Bottom-right: Plots of the diffuse-stellar ratio for various points in time. The small circles indicate the position of the ionization front.

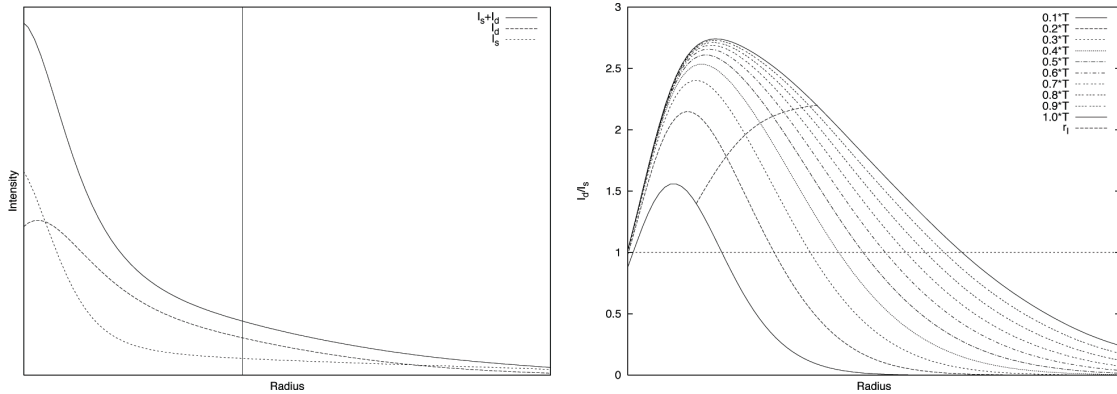


Fig. 11.3: Results for the r^{-2} density profile. Left: Plots of the fractional diffuse (I_d/I_{tot}) and stellar photons (I_s/I_{tot}), and of the diffuse-stellar photon ratio (I_d/I_s). Right: Plots of the diffuse-stellar ratio for various points in time. The dashed line indicates the position of the ionization front.

depicted in the bottom-right panel of Fig. 11.2, in which the diffuse versus stellar photon ratio is given versus the distance to the source, for various points t in time, measured with respect to the final simulation time T . The result for $t = T$ matches the graph in Fig. 11.2, bottom-left. For each different line, a small circle is given, indicating the position of the front. It is apparent from this figure that already at $t = 0.4T$ the diffuse component of the radiation field will dominate near the front. This point time is reasonable, considering that, for the parameters in this test (see Sect. 9.3), it is equivalent to $\sim 1.6t_{\text{rec}}$, at which time recombinations have started to exert their influence.

A similar experiment was performed for an r^{-2} density distribution. As discussed in Appendix 11.A, one has to be careful to choose the source luminosity in such a way that the Strömgren radius has a finite value. The result for this equilibrium solution is depicted in the left panel of Fig. 11.3, showing the stellar, diffuse and total photon intensity versus the distance to the source. The position of the Strömgren radius is indicated by an extra line. Comparing these results to the analytical solutions in Fig. 11.1, bottom-left, we see that the agreement is very good indeed, again with the modification that the ionization front now has a certain thickness. Just like we did for the homogeneous density profile, we determined the diffuse versus stellar intensity ratios for various different points in time, leading up to the equilibrium solution. The results are depicted in the right panel of Fig. 11.3. The dashed line indicates the position of the ionization front, varying in time, as derived in Mellema et al. (2006). It is obvious that the diffuse radiation field is the dominant component, already at very early times.

We can conclude that the analytical work of Sect. 11.2, and the conclusions presented there, are valid, even though radiative transfer effects are ignored. The only modification apparent from numerical experiments that do perform a full radiative transfer treatment is the ability to resolve extended ionization fronts. As an extra bonus, the numerical approach enables us to follow the diffuse-stellar turnover point in time, showing that the general conclusions in Sect. 11.2 are valid very early on in time, when the equilibrium solution has not yet been reached for a long time.

11.4

Counter Effects

The results presented so far in this chapter are the best, or worst, case scenario, depending on whether or not you like diffuse photons. There are several effects that will diminish the effect of diffuse photons taking over, causing the turnover point r_{eq} to move towards R_S . Three of the more prominent ones are described in what follows.

11.4.1 Hard Photons

The analysis in this chapter was done assuming that the source spectrum is peaked just above the Lyman limit. It is much more realistic to assume that a substantial part of the luminosity will be in the higher frequency domain, in comparison with the diffuse photon spectrum being strongly peaked just above the Lyman limit threshold. The high-energy photons have a longer mean free path ($a(\nu) \propto \nu^{-3}$) than the diffuse photons, from which we expect that their contribution will in effect move r_{eq} towards R_S . To accurately model this, we would have to do a full radiative transfer treatment, but we can give an analytical estimate of this effect. We alter the analytical approach in Sect. 11.2 by not taking the source spectrum to be a delta function just above the Lyman limit, but at a certain higher frequency. As a result, we relax the condition that the medium is optically very thick to source photons, by which not $I_S(r)$ photons contribute to the locally available radiation field, but only a factor $0 \leq c \leq 1$ thereof. Effectively, we incorporate the differences in spectrum averaged ionization cross section, see also the κ factor in Eq.(11.4), with this factor c . Nothing changes in the equations except that the weighing factors will now be $cI_S/(cI_S + I_d)$ and $I_d/(cI_S + I_d)$. We solved the equations for the r^{-2} distribution for varying values of c and plotted the results in Fig. 11.4. One can see that the position of r_{eq} does change, but that the diffuse radiation field effect is still very noticeable, and does not disappear, even for a very hard source spectrum.

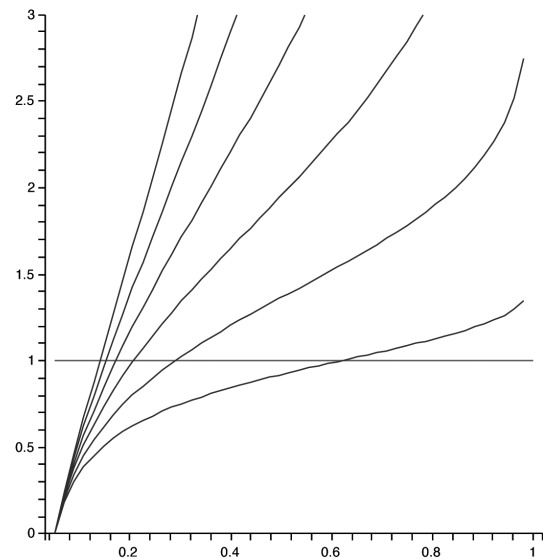


Fig. 11.4: Graphs of the functions $(I_d/I_S)(r)$ for the varying values (from top to bottom) $c = 1.0, 0.9, 0.8, 0.7, 0.6, 0.5$ in a r^{-2} density distribution. Overplotted is the line $I_d/I_S = 1$ which determines the radius above which the diffusion dominates.

11.4.2 Redshifting

The spectrum of diffuse recombination photons is sharply peaked just above the Lyman limit threshold. Consequently, any redshifting of the spectrum will result in diffuse photons that do not have enough energy to ionize a hydrogen atom. In a cosmological setting, this redshifting is due to the expansion of the Universe. We may estimate the effect hereof by estimating how much a diffuse photon will be redshifted when moving along one mean free path (Schaye 2005).

Given the thermal broadening scale v_{th} of a recombining plasma, an ionizing photon has to travel a distance L_z , before it gets redshifted away. Filling in typical values at redshift $z = 10$, we obtain

$$L_z = \frac{v_{\text{th}}}{H} = 8.2 \text{ kpc} \left(\frac{v_{\text{th}}}{10 \text{ km/s}} \right) \left(\frac{10}{1+z} \right)^{3/2}. \quad (11.15)$$

This distance L_z will equal the local mean free path for the following neutral hydrogen density:

$$n_{\text{HI}} = \frac{1}{A_0 L_z} = 6.3 \cdot 10^{-6} \text{ cm}^{-3} \left(\frac{10 \text{ km/s}}{v_{\text{th}}} \right) \left(\frac{1+z}{10} \right)^{3/2}, \quad (11.16)$$

in which we used the hydrogen ionization cross section at the Lyman limit, cf. Eq.(8.4). Dividing by the number density of neutral hydrogen gives the critical neutral fraction:

$$\frac{n_{\text{HI}}}{n} = 0.031 \left(\frac{10 \text{ km/s}}{v_{\text{th}}} \right) \left(\frac{1+z}{10} \right)^{3/2} \left(\frac{1}{1+\delta} \right), \quad (11.17)$$

in which δ is a measure for the local overdensity with respect to the mean.

From Eq.(11.17), we can conclude that for the EOR, i.e. $z \geq 6$, recombination radiation will only be important when the neutral fraction exceeds, say, 10%. This is assuming that the medium density is equal to its mean. The hydrogen gas and the sources will be distributed very inhomogeneously, however. Most sources will form in regions where the overdensity δ is very high, and Eq.(11.17) then predicts a much lower neutral fraction threshold. Moreover, there is the effect that the recombination rate scales with the square of the neutral hydrogen density, cf. Eq.(8.9), biasing the effect of recombination radiation to very near the source, where the density is highest. This effect is also apparent from the location of the turnover point r_{eq} being very near the source in an r^{-2} density profile.

Thus, given that most diffuse photons are produced very close to the source, we can conclude that the redshifting effect will not be very important in our analysis in this chapter. Still, we have to take into account that recombination radiation will not fill in shadows at scales exceeding L_z .

11.4.3 Diffusion Speed

When we determine the position and velocity of the ionization front by using the jump condition Eq.(11.1), we effectively ignore the speed of the particles themselves. The speed of the front is solely determined by the number of photons emitted by the source and the number of neutral atoms to be ionized at the position of the front. The approach of cosmological radiative transfer methods, discussed in Chapter 7, solving for each time step the quasi-static radiative transfer equation Eq.(7.7), amounts to the same thing: the speed of the individual particles is ignored, and time-dependent effects are only incorporated by changing the medium properties at each time step. We already discussed a minor drawback of this approach, showing that the ionization front propagation speed may exceed the speed of light very near the source. This approximation is valid in most cases, however. The light crossing times of the ionization bubbles is very small compared to the relevant time scales of front propagation, or, in other words, the speed of the stellar photons is so high that they reach the front almost immediately.

Once the diffuse photons enter the picture, the story may become different. Of course, each diffuse particle, once created, behaves exactly the same as a stellar photon, moving from

its point of creation in a straight line along one mean free path, until interaction. The overall field of diffuse particles may show a different behavior, though. Stellar photons are directed radially outwards from the source, moving towards the front always, whilst a recombination photon emitted isotropically near the front may move away from the front. When assuming a quasi-static approximation of radiative transfer, this effect is unimportant, because this photon can then just travel to the front on the other side of the HII region. The symmetry condition used in our analytic derivation then applies: diffuse photons moving in the opposite direction are exactly compensated for by photons doing the similar thing on the other side.

When we relax the quasi-static approximation, and do keep track of the finite speed of the individual particles, this effect might become important. It may impose a retardation effect of the diffuse photons with respect to the stellar ones. We can estimate this effect by recognizing that the collective behavior of the diffuse field can be described by a diffusion equation. Standard solutions of this equation predict Gaussian profiles expanding according to a diffusion coefficient $D = c\lambda/3$, in which c is the speed of light, and λ is the local mean free path. This entails that the diffuse photons effectively move at a reduced speed $c/3$, or, equivalently, that they move along a path that is effectively three times as long. Thus, because their overall speed is less, stellar photons may reach a neutral atom faster than their diffuse equivalents.

Still, as long as both the light crossing time and the diffusion time are much shorter than the relevant time scales of the front propagation, this effect will only be second-order.

11.5

Conclusions

What does all of this analysis entail? First and foremost, it shows that a large part of HII regions, in some cases even almost the whole volume, is diffusion dominated. This has a profound effect on the formation of shadows behind clumps within the HII region. A more quantitative analysis on this matter has to be done in forthcoming work, but one can immediately see that, even in the case of the only moderately diffuse homogenous medium distribution (Sect. 11.2.1), if a dense clump is put within the radius r_{eq} , the resulting shadow will be squeezed by the diffuse field in the outer region. This effect will be much stronger than in the analysis in Canto et al. (1998), in which only the locally produced diffuse photons, cf. Eq.(11.4), can move into the shadow region, whereas in our analysis also the overall diffuse radiation field will have a profound influence. This effect will have an impact on (cosmological) radiative transfer codes which use ray casting methods to delineate sharp shadows, and the resultant 21cm signatures of the HI regions in between the overlapping (re)ionization bubbles, which may be observed in the near future. Also the formation mechanism of knots in planetary nebulae (López-Martín et al. 2001; Rijkhorst 2005) may have to be altered.

Another effect is that the diffuse radiation field will influence the dynamics of the HII region. From the considerations in the previous section for the equilibrium solution, one can see that the only radiation component pushing against the front is the diffuse one, which is not mono-directional, but isotropic. Thus, the net flux of ionizing photons in the direction of the front is less. The symmetry condition, as used in Sect. 11.2.1, is only valid in the equilibrium (time-independent) solution, but in the time-dependent case the speed of photons diffusing from one side of the HII region to the other is finite, and, as a result, this effect may slow down the ionization front, by which the standard analytical solutions of Eq.(11.1), as

found in e.g. Spitzer (1978), will have to be modified to take into account that this diffusion speed is considerably smaller than the speed of light. If it turns out that this effect can not be neglected, it will have an influence on the standard analysis of how fast the first stars and/or quasars could have reionized the Universe (Cen & Haiman 2000; White et al. 2003; Wyithe & Loeb 2004; Yu 2005; Yu & Lu 2005). Even though Shapiro & Giroux (1987) have shown that in most realistic cosmological cases the Strömgren sphere is not reached, because the recombination time in a sparse medium is shorter than the Hubble time (for $z \leq 10$), we have argued that most recombination photons are produced near the source, where the medium is very high, and, hence, the recombination time very small. Thus, diffusion terms will have to be incorporated in the dynamics of ionization front propagation, even if they expand on cosmological scales.

Concluding, the usual on the spot approximation in which the diffuse radiation component is neglected can only be used to calculate the Strömgren sphere within an isotropic HI medium distribution. We have shown that the diffuse component can not be neglected, which will result in a drastic change of the morphology of anisotropic distributions, and which may change the speed of the ionization fronts.

Acknowledgements We would like to thank Vincent Icke and Garrelt Mellema for carefully reading the manuscript, Harm Habing for enlightening discussions, and finally Sijme-Jan Paardekooper for his mathematical insight. We are much obliged to the referee W. Steffen for his useful comments. Finally, we owe much gratitude to Will Henney who helped shape parts of this chapter through his sharp criticisms.

Bibliography

- Baker, J. G. & Menzel, D. H. 1938, ApJ, 88, 52
 Canto, J., Raga, A., Steffen, W., & Shapiro, P. 1998, ApJ, 502, 695
 Cen, R. & Haiman, Z. 2000, ApJ, 542, L75
 Hong, S. S. & Sung, H. I. 1989, Journal of Korean Astronomical Society, 22, 127
 Hummer, D. G. & Seaton, M. J. 1963, MNRAS, 125, 437
 López-Martín, L., Raga, A. C., Mellema, G., Henney, W. J., & Cantó, J. 2001, ApJ, 548, 288
 Mellema, G., Iliev, I. T., Alvarez, M. A., & Shapiro, P. R. 2006, New Astronomy, 11, 374
 Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493
 Osterbrock, D. E. 1989, Astrophysics of gaseous nebulae and active galactic nuclei (Mill Valley, CA: University Science Books)
 Rijkhorst, E.-J. 2005, Numerical Nebulae, PhD thesis, Leiden
 Rubin, R. H. 1968, ApJ, 153, 761
 Schaye, J. 2005, private communication
 Shapiro, P. R. & Giroux, M. L. 1987, ApJ, 321, L107
 Spitzer, L. 1978, Physical processes in the interstellar medium (New York: Wiley-Interscience)
 Strömgren, B. 1939, ApJ, 89, 526
 White, R. L., Becker, R. H., Fan, X., & Strauss, M. A. 2003, AJ, 126, 1
 Wyithe, J. S. B. & Loeb, A. 2004, Nature, 427, 815
 Yu, Q. 2005, ApJ, 623, 683
 Yu, Q. & Lu, Y. 2005, ApJ, 620, 31
 Zanstra, H. 1951, Bull. Astron. Inst. Netherlands, 11, 341

11.A

General Derivation of the Diffuse Radiation Field

The derivation of the diffuse versus stellar radiation field within a Strömgren sphere in Sect. 11.2 can be done in a much more general way, for every possible dimension of space, and every possible isotropic density distribution. Because it may benefit later analyses, we will do the same calculations in a general \mathbb{R}^d , in which $d \in \mathbb{N}$, for a general gas density distribution

$$n(r) = n_c \left(\frac{r_c}{r} \right)^p, \quad (11.18)$$

in which r_c is an inner cut-off, which is needed when $d/2 \leq p$, and $p \in \mathbb{R}$.

Pervasive to our derivations is the assumption of spherical symmetry, by which the Jacobian for integration over the r coordinate becomes $\xi(d)r^{d-1}$, in which $\xi(d) = 2\pi^{1/2}/\Gamma(d/2)$ is a geometrical factor. In a similar fashion as has been done in Sect. 11.2, one can define the number of recombinations to every level and the number of diffuse photons produced within a spherical shell of infinitesimal thickness dr at distance r to be

$$\begin{cases} N_{\text{rec}}(r) &= \xi(d)\alpha_A n_c^2 r_c^{2p} r^{d-1-2p} dr \\ N_{\text{diff}}(r) &= \xi(d)\alpha_1 n_c^2 r_c^{2p} r^{d-1-2p} dr, \end{cases} \quad (11.19)$$

Thus, the system of differential equation for $I_s(r)$ and $I_d(r)$ can be defined on the domain $r_c \leq r \leq R_S$ as

$$\begin{cases} \frac{dI_s(r)}{dr} &= -\xi(d)\alpha_A n_c^2 r_c^{2p} r^{d-1-2p} \frac{I_s(r)}{I_s(r)+I_d(r)} \\ \frac{dI_d(r)}{dr} &= -\xi(d)\alpha_A n_c^2 r_c^{2p} r^{d-1-2p} \frac{I_d(r)}{I_s(r)+I_d(r)} + \xi(d)\alpha_1 n_c^2 r^{d-1-2p} \end{cases} \quad (11.20)$$

The first term on the rhs of Eq.(11.20) denotes the total number of recombinations that have to be compensated for by either stellar or diffuse photons, the second term on the rhs of the equation for the diffuse component denotes the number of diffuse photons added along the way. The system is closed via the two initial conditions $I_d(r_c) = 0$ and $I_s(r_c) = S_*$. In all cases, S_* is chosen such that $I_s(r) + I_d(r) = 0$ at $r = R_S$, as is required in the classical Strömgren analysis.

Adding the two differential equations in Eq.(11.20) results in an equation for the total radiation field

$$\frac{d(I_s + I_d)(r)}{dr} = -\xi(d)\alpha_B n_c^2 r_c^{2p} r^{d-1-2p}. \quad (11.21)$$

This differential equation has three different regimes of solutions, depending on the dimension of space and the specific density profile. As such, we can distinguish three different cases: (A) $p > d/2$, (B) $p = d/2$, and (C) $p < d/2$. Each of these will be solved for separately in what follows.

11.A.1 Case A: $p > d/2$

In this particular regime, one needs an inner boundary at $r = r_c$, because there of a singularity present at $r = 0$. Depending on the choice of $I_s(0) = S_*$, it is even possible that $R_S \rightarrow \infty$. Solving Eq.(11.21) for these parameters, one obtains

$$(I_s + I_d)(r) = \frac{\xi(d)}{2p-d} \alpha_B n_c^2 r_c^{2p} \left[r^{-(2p-d)} - R_S^{-(2p-d)} \right]. \quad (11.22)$$

This equation can easily be used to determine which stellar photon rate S_* will cause the ionization front to keep expanding forever, never slowing down to zero velocity (i.e. $R_S \rightarrow \infty$). A relevant example hereof is the expansion of an ionization front in \mathbb{R}^3 in an r^{-2} density profile. For this configuration, the window of photon rates that will result in a finite Strömngren sphere turns out to be very narrow.

Filling in Eq.(11.22) into the original system Eq.(11.20) results in

$$\begin{aligned} I_s(r) &= (I_s + I_d)(r) - I_d(r) \\ &= \frac{\xi(d)}{2^{p-d}} \alpha_B n_c^2 r_c^{2p} \left[r_c^{-(2p-d)} - R_S^{-(2p-d)} \right]^{1-\frac{\alpha_A}{\alpha_B}} \left[r^{-(2p-d)} - R_S^{-(2p-d)} \right]^{\frac{\alpha_A}{\alpha_B}}, \end{aligned} \quad (11.23)$$

for which we can derive the radius at which the diffuse component of the radiation field starts to dominate,

$$r_{\text{eq}} = R_S \left[1 + 2^{-\alpha_B/\alpha_1} \left(\left[\frac{R_S}{r_c} \right]^{2p-d} - 1 \right) \right]^{-1/(2p-d)}. \quad (11.24)$$

Note that, for realistic parameters, $r_c \ll R_S$, by which Eq.(11.24) can be approximated by

$$r_{\text{eq}} = 2^{\alpha_B/[\alpha_1(2p-d)]} r_c, \quad (11.25)$$

which will approach r_c , when $p \gg d/2$. Thus, the density distributions in this category, which has the largest population of 'real life' density distributions, such as an r^{-2} distribution in \mathbb{R}^3 , have as a result an HII region which is fully dominated by the diffuse component.

11.A.2 Case B: $p = d/2$

In this domain of parameter choice, the singularity at $r = 0$ has still not vanished, so an inner boundary is still needed. A similar analysis as in the Sect. 11.A.1 results in

$$\begin{cases} (I_s + I_d)(r) &= \xi(d) \alpha_B n_c^2 r_c^d \ln \left(\frac{R_S}{r} \right) \\ I_s(r) &= \xi(d) \alpha_B n_c^2 r_c^d \left[\ln \left(\frac{R_S}{r_c} \right) \right]^{1-\frac{\alpha_A}{\alpha_B}} \left[\ln \left(\frac{R_S}{r} \right) \right]^{\frac{\alpha_A}{\alpha_B}} \\ r_{\text{eq}} &= R_S (r_c/R_S)^{2^{-\alpha_B/\alpha_1}}. \end{cases} \quad (11.26)$$

11.A.3 Case C: $p < d/2$

The singularity in $r = 0$ has vanished, so we may take $r_c \rightarrow 0$, if we like. Using this, we obtain,

$$\begin{cases} (I_s + I_d)(r) &= \frac{\xi(d)}{d-2p} \alpha_B n_c^2 r_c^{2p} \left(R_S^{d-2p} - r^{d-2p} \right) \\ I_s(r) &= \frac{\xi(d)}{d-2p} \alpha_B n_c^2 r_c^{2p} \left(R_S^{d-2p} \right)^{1-\alpha_A/\alpha_B} \left(R_S^{d-2p} - r^{d-2p} \right)^{\alpha_A/\alpha_B} \\ r_{\text{eq}} &= R_S \left(1 - 2^{-\alpha_B/\alpha_1} \right)^{1/(d-2p)}. \end{cases} \quad (11.27)$$

The homogeneous and r^{-1} density distributions in \mathbb{R}^3 belong to this class of parameters, and have solutions in the form of Eq.(11.27). Note that these solutions are also valid, when p becomes negative.

11.A.4 Overview

For later reference, we give an overview of the values of r_{eq} for several different combination of d and p . In case an inner boundary was needed, we chose $r_c/R_s = 0.05$.

| <i>Overview of values for r_{eq}/R_s</i> | | |
|--|-------------------|--------------------|
| | $d = 2$ | $d = 3$ |
| $p = 0$ | 0.82 | 0.88 |
| $p = 1$ | 0.38 | 0.67 |
| $p = 2$ | $0.087 = 1.74r_c$ | $0.14 = 2.80r_c$ |
| $p = 3$ | | $0.073 = 1.46r_c$ |
| $p = 4$ | | $0.0625 = 1.25r_c$ |

Refining Reionization Simulations

In order to make the computations more tractable, modern reionization simulations ignore several aspects of the intricate reionization process. In this chapter, we quantitatively study the impact of these aspects on the simulation results as a whole, thereby showing that reionization simulations need to be refined by including these effects. To accomplish this, we have coupled our radiative transfer method `SimpleX` to a cosmological SPH code, by which we were able to perform many different simulations for each separate aspect, and this for many different box sizes. We start with describing how the box size and effective resolution of the method influences the results, so that, having understood this, we are in a position to systematically study several physical effects that are often ignored. First, we show that including the effect of reheating has a profound influence on the reionization history, accelerating the process by globally reducing the clumping of the gas. We conclude by quantifying how including the diffuse radiation field alters the morphology of the HII regions, and the resultant 21-cm maps.

Refining Reionization Simulations
J. Ritzerveld, A. Pawlik, and J. Schaye
In preparation.

12.1

Introduction

Recent simulations of the Epoch of Reionization (EOR) underemphasize certain numerical and physical aspects that may have a profound impact on the overall result. These include the effects of reheating, and that of the diffuse radiation field. Because most of these have previously been neglected, it is interesting to see how they influence the physical mechanism of the process of reionization. In Chapter 10, we have specified which numerical ingredients are needed to successfully analyze the impact of these effects. An accurate hydrodynamics solver coupled to an efficient radiative transfer solver, both having a large adaptive range, is mandatory.

In this chapter, we present the results of coupling a modified version of the cosmological SPH code `GADGET2` (Springel 2005) to the adaptive cosmological radiative transfer code `SimpleX`, that was presented in Part II of this thesis. The SPH code is Lagrangian, so that it can adaptively follow the nonlinear evolution of the collapse of the baryonic matter, hereby resolving the structure of high density regions. The radiative transfer code `SimpleX` is

highly adaptive by construction, as well, and does not scale with the number of sources. The efficiency of this transfer method enables us to systematically study the parameter space of every single effect, by repeating the same experiment several times, but it also permits the inclusion of the diffuse radiation field, the importance of which was emphasized in Chapter 11. Moreover, as a combination, the codes allow for a systematic treatment of radiative feedback effects, such as reheating.

We start with giving a specification of the setup of the simulations in Sect. 12.2, showing some of the results of several representative reionization simulations that we have run, giving a global impression of the versatility of this unique combination of codes. Hereafter, we more systematically study the effects of resolution and box size on the star formation history and the reionization history as a whole in Sect. 12.3. Once these effects have been singled out, we can examine the radiative feedback effect of reheating, the results of which are presented in Sect. 12.4. We finish this chapter by studying the effect of the diffuse radiation field on the morphology of overlapping HII regions in Sect. 12.5.

12.2

Simulation Setup

For a realistic numerical simulation of the EOR, it is allowed, as explained in Chapter 8, to decouple the radiative transfer from the hydrodynamic evolution, post-processing the transport of photons at the end of each hydro time slice. This is the approach we shall adhere to here. The hydrodynamical evolution is provided by a variant of the cosmological SPH code `GADGET2`, the specifics of which are described in Schaye et al. (2006). The modifications with respect to the standard open source version of `GADGET2` are noteworthy, including new routines for the determination of hydrogen and helium cooling and heating, and a new Schmidt Law prescription for the computation of the star formation rates. The resultant code is very fast, which makes it ideal for running many different simulations, producing hydro output for several different box sizes, resolutions, and initial conditions.

12.2.1 Hydro Input

All the hydro input in this chapter is based on simulations runs with initial conditions at $z_{\text{start}} = 127$ that were created by displacing particles from a homogeneous, ‘glass-like’ distribution (White 1994) by a field with a power spectrum generated by `CMBFAST` (Seljak & Zaldarriaga 1996). The cosmological parameters were chosen to correspond to what was inferred from the latest 3-year WMAP observations presented in Spergel et al. (2006) (i.e. $[\Omega_0, \Omega_b, \Omega_\Lambda, h] = [0.25, 0.045, 0.75, 0.73]$).

To systematically study the effect of box size and resolution, all experiments were repeated for several different comoving box sizes: $L = 1.5625/h, 3.1250/h, 6.2500/h, 12.5000/h,$ and $25.0000/h$ Mpc, each with an increase of a factor 2. Except for those run for the resolution study in Subsect. 12.3.1, all simulations were run with a resolution of 256^3 SPH particles. This is low compared to the N-body simulations presented in Iliev et al. (2006) (3248^3) and Zahn et al. (2006) (1024^3), but our simulations do include hydrodynamics and our radiative transfer code is adaptive as well, by which the overall effective resolution of the transfer is many orders of magnitudes higher, as we will show later on. The effective mass resolution, i.e. the mass of each gas particle, is $M_{\text{gas}} = 2.63 \cdot 10^3 M_\odot, 2.11 \cdot 10^4 M_\odot, 1.69 \cdot 10^5 M_\odot,$

$1.35 \cdot 10^6 M_{\odot}$, $1.08 \cdot 10^8 M_{\odot}$, and the total number of sources at $z = 6$ is $4.4 \cdot 10^4$, $7.4 \cdot 10^4$, $7.1 \cdot 10^4$, $4.3 \cdot 10^4$, and $1.4 \cdot 10^4$, respectively.

The hydro simulations did not include feedback or metal enrichment, which is why we are free in Sect. 12.3.2 to rescale the star formation rates as we like. Cooling is included: a photodissociating background suppresses molecules, and when reheating is turned on (cf. Sect. 12.4), an extra UV background is added. The softening length is $1/25$ of the mean interparticle distance.

The output is generated by creating HDF5 files (see Sect. 8.2) for 60 different evenly spaced snapshots, starting at redshift $z = 40$ and ending at $z = 6$. The HDF5 files include the coordinates, masses and densities for each SPH particle. The star formation routines of this SPH code determine, based on a prescription that uses parameters like the local density, which particles evolve into a stellar particle. The mass and the stellar formation time of each of these stellar particles are stored in the same HDF5 files. Thus, we know, with the time-step resolution of the SPH code, when each star forms.

12.2.2 The SimpleX Lattice

The output for each snapshot serves as the input for our radiative transfer scheme. SimpleX was designed to read this specific input format (see Chapter 8 for more on that), and directly uses the SPH particle distribution as its own particle distribution. The adaptive random lattice, along which the transport of photons will take place, is generated by performing a Delaunay triangulation (cf. Chapter 3), using periodic boundary conditions. In this case, the number of particles is so large, that we need to split up the box, and compute the triangulation separately for each of the 8 subboxes (cf. Sect. 8.3).

The resultant unstructured lattice is highly adaptive. It has a huge range in resolution, with Delaunay lines lengths being as small as 10^{-5} of the width of the computational domain, mostly in high density regions, by which the effective radiative transfer resolution can reach values as high as 10^{15} ! This is a giant leap forward, compared to the 128^3 or 256^3 resolution as used in the more common radiative transfer methods. To illustrate this adaptive property, we refer to Fig. 12.1. The top panel depicts several zoom-ins of a 2D cut-out of the Delaunay lattice generated by the SPH particle distribution of one of the simulations ($L = 1.5626/h$ Mpc). More illustrative is a similar 2D cut-out, but now of a logarithmic plot of the volume of each Delaunay tetrahedron, which is shown in the bottom panel of that same figure. It is obvious that this property of our SimpleX radiative transfer method allows us to more accurately track the behavior of radiation and matter in highly dense, geometrically complex regions.

12.2.3 Source Luminosities

During the cosmological time spanned by each of the snapshots, several sources may turn on. The SPH code has determined the exact formation time of each of these sources, so that we have a catalogue of sources that gives us a precise indication of at which SimpleX time step the source should turn on.

The only thing that remains to be determined then is the luminosity of ionizing photons of the source as a function of their age and mass. To accomplish this, we use the stellar population synthesis by Bruzual & Charlot (2003). The IMF is an adjustable parameter of the model, so we adopt the parametrization by Chabrier (2003) of the single-star IMF in the Galactic disk. This was also used in the SPH simulations. As an input for these routines, we

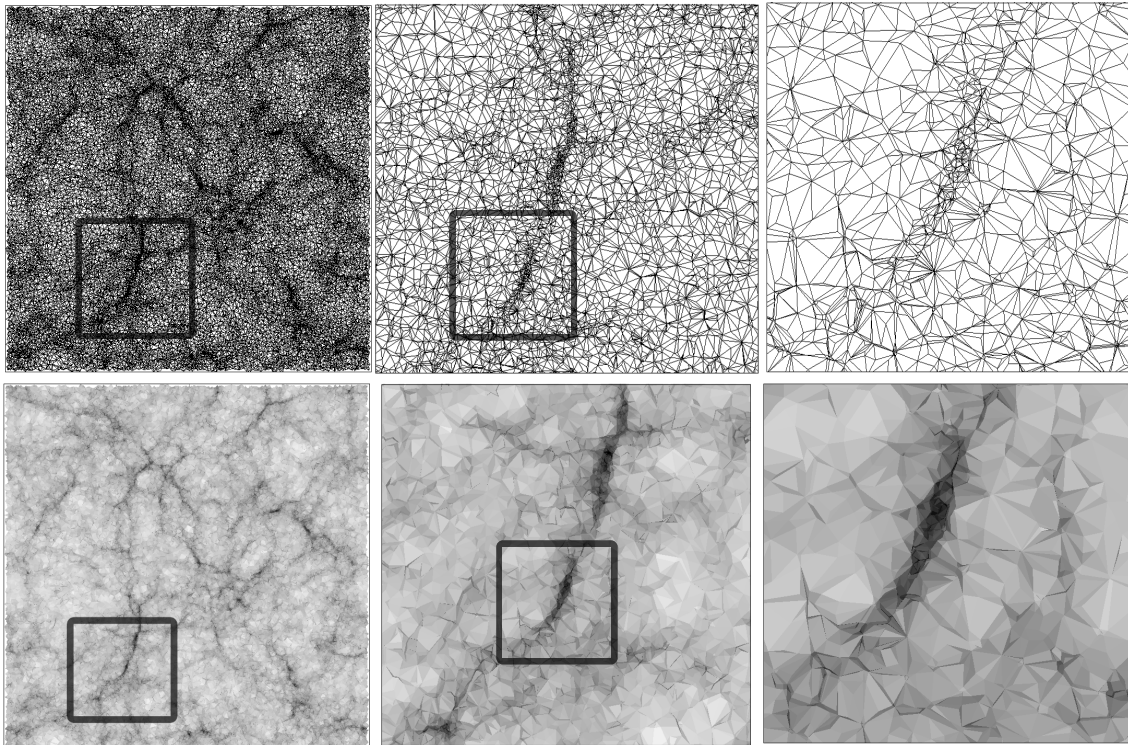


Fig. 12.1: Illustration of the highly adaptive nature of the SimpleX random lattice. Top: Zoom-ins of a 2D cut-out of the random lattice itself. Bottom: Same, but now for a 2D cut-out of a logarithmic plot of the volume of each Delaunay tetrahedron (darker is smaller). The lattice is based on a representative SPH particle distribution of resolution 256^3 in a computational box of width $L = 1.5625/h$ Mpc.

need the star’s mass, age, and their metallicity. The first two are given by the SPH code, the latter is a parameter we have to choose a priori. After testing several other values, we chose a metallicity of $Z = 0.0004$ as a fiducial value for the metallicity of the first sources. Once the source spectra are obtained, we can integrate the flux F_ν divided by the photon energy $h\nu$, using the ionization cross section as a weighing factor, to obtain the flux of ionizing photons.

We stress again that, in the simulations in this chapter, we solve for hydrogen ionization only, and we assume an escape fraction of unity, i.e. $f_{\text{esc}} = 1$, because the SPH method was designed to resolve the galaxies themselves, so that the actual escape fraction can be computed self-consistently.

12.2.4 The Transfer

Once the lattice has been constructed, and the sources have been located, the transfer can commence. The SimpleX transfer code tracks the evolution of the ionizing photons, as they move through the initially neutral hydrogen, blowing bubbles around each source.

Recombination is incorporated via a direct integration of the rate equations, as described in Chapter 8. To determine the local recombination rate, and the local optical depths, we need an estimate of the local density. The mass of a particle is given by the SPH code, so what needs to be determined is a measure for the volume of each particle. The most

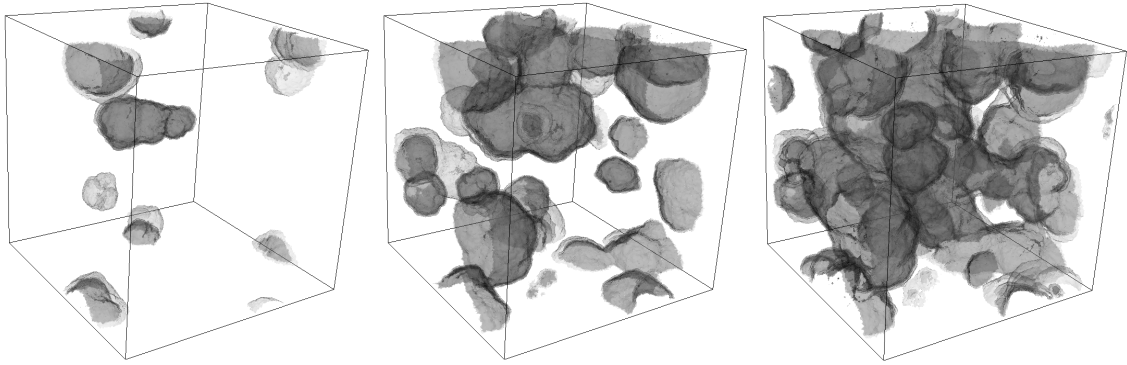


Fig. 12.2: Visual rendering of the HII distribution of a representative reionization simulation. Comoving box size is $L = 6.2500/h$ Mpc. Three different snapshots are depicted, and the gradual overlap of HII regions is apparent. Note that the effects of periodic boundary conditions is evident: bubbles disappear at one end of the domain and emerge on the other.

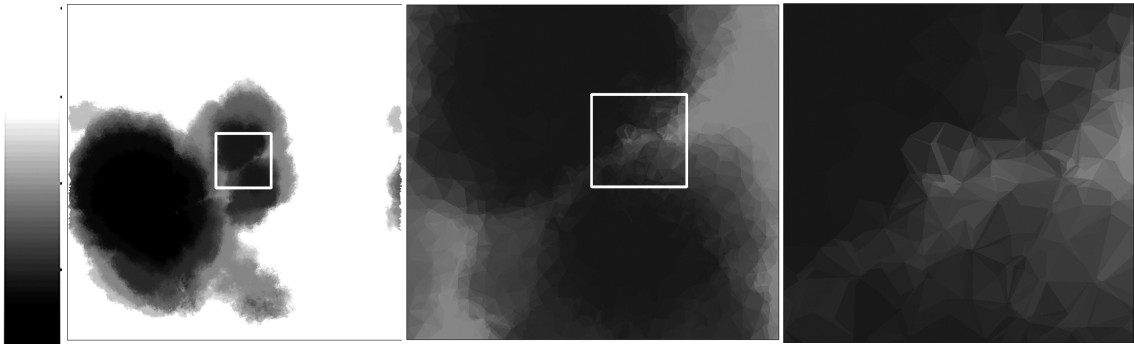


Fig. 12.3: Zoom-ins of a 2D slice of the HI distribution in a box with comoving width $L = 6.2500/h$ Mpc. Darker regions have a larger ionization fraction.

consistent approach is to use the volume of the Voronoi cell of each particle, and this is the volume we assign to each particle. Note that using m_i/ρ_i as a measure for the volume of each particle i does not conserve volume.

At the end of each snapshot, each grid particle has a certain ionization fraction, and we can make up a statistic of the global volume- or mass-averaged ionization fraction, to track the evolution of the process of reionization. Moreover, we can use these data to make a visualization of the distribution of HII regions, an example of which is depicted in Fig. 12.2, in which a visualization of the HII distribution at the end of three time slices is plotted. It is obvious that, as time progresses, the computational box is gradually filled, until full reionization has been reached. We can also make a similar zoom-in sequence as in Fig. 12.1, but now for the neutral hydrogen distribution. The result is depicted in Fig. 12.3, from which it is immediately apparent that using adaptive lattices will result in resolving inhomogeneities in the HI distribution even on very small scales.

This procedure can be repeated over and over again, for different box sizes, resolution, particle distributions, source luminosities, and such, and it is this setup of coupling our SimpleX radiative transfer code to the output of the SPH code that we will use to systematically study the impact of several effects on the global history of reionization.

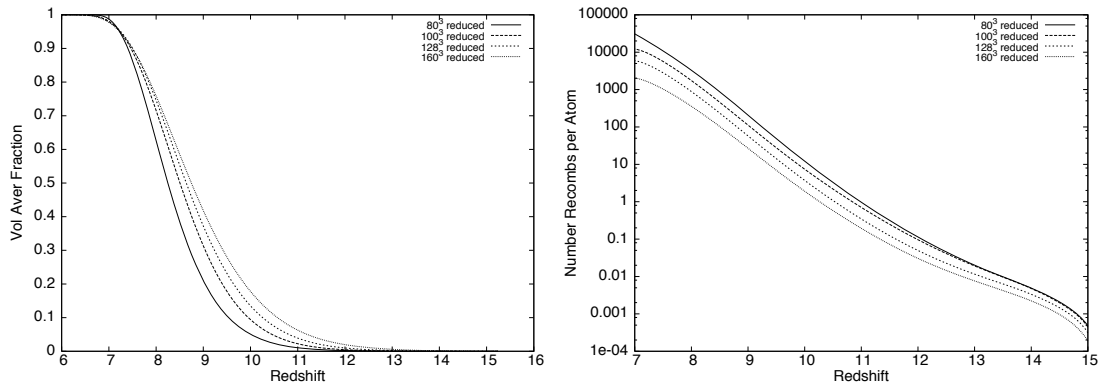


Fig. 12.4: Results for the reionization of a box of comoving width $L = 6.2500/h$ Mpc with SPH particle resolution 256^3 , for several effective radiative transfer resolution, $N_{\text{sub}} = 80^3, 100^3, 128^3$, and 160^3 . Left: Volume averaged ionization fraction as a function of redshift. Right: Total number of recombinations divided by the total number of atoms as a function of redshift.

12.3

Numerical Effects

When trying to study the effect of physical mechanisms, such as reheating or the contribution of diffuse photons, we first need to single out any numerical effect that may interfere with the overall result. These effects may be caused by the choice of box size and resolution, but also by choosing whether or not the boundary conditions are periodic. In a cosmological context, it is justifiably better to choose periodic boundary conditions, in order to circumvent boundary effects, but also to minimize cosmic scatter, so this is the approach we took.

That being settled, it still remains to be determined how the resolution and box size affect the global reionization history. If these effects are understood, and possibly eliminated, we can more accurately track the effects of the physical mechanisms.

12.3.1 The Effect of Resolution

We start with investigating the effect of resolution. To accomplish this, we choose just one specific box size, in this case the middle box size $L = 6.2500/h$ Mpc, and vary the effective `SimpleX` particle distribution for that one box. The SPH particle resolution is always 256^3 , as has been previously pointed out, so the trivial one-to-one mapping of the SPH particles to the `SimpleX` particles would result in an identical effective radiative transfer resolution.

We can, however, vary the radiative transfer resolution by dismissing the one-to-one mapping procedure, and, instead of that, taking random subsets of the SPH particle distribution. These random subsets have a predefined size with resolution $N_{\text{sub}} \leq 256^3$, and, because the particles are chosen randomly, the overall distribution will be representative of the original SPH distribution. To keep the star formation history fixed, all star particles are included. In order to conserve mass, though, one has to increase the mass of each particle by $N_{\text{gas}}/N_{\text{sub}}$, in which $N_{\text{gas}} + N_{\text{star}} = 256^3$. In most cases, $N_{\text{star}} \ll N_{\text{gas}}$.

Using the simulation setup as laid out in Sect. 12.2, we can follow the transport of ionizing photons, and determine the evolution of HII regions until reionization has been achieved, for several different effective transfer resolutions. The resolutions we have used are $N_{\text{sub}} = 80^3$,

100^3 , 128^3 , and 160^3 , each differing by a factor of roughly 2. The results for the global reionization histories for each effective resolution are plotted in Fig. 12.4. The left panel depicts the volume averaged ionization fraction of the particles as a function of redshift, i.e.

$$\langle x \rangle_{\text{vol}} = \frac{\sum_i x_i V_i}{\sum_i V_i}, \quad (12.1)$$

in which $x_i = N_{\text{HII}}^i/N^i$ is the ionization fraction of particle i , and V_i is its Voronoi volume. For this particular simulation, the SPH code predicts that the first source forms at $z = 15$, which is where the transfer code `SimpleX` can commence. The figures show a similar trend for each effective resolution: reionization starts relatively slowly, until, at an approximate redshift of $z = 10$, the box is rapidly filled with overlapping HII regions. At a redshift of $z = 7$, the box has been fully ionized.

Differences are also apparent. Although full reionization has been reached at $z = 7$ for all effective resolutions, the lines diverge somewhat for higher redshifts. One can see that higher effective resolutions will result in the box having a larger volume averaged ionization fraction, until full reionization. This effect can be understood by plotting the number of recombinations per atom (i.e. the total number of recombinations during one time slice divided by the total number of atoms) as a function of redshift, the result of which is depicted in the right panel of Fig. 12.4. This figure clearly shows that lower effective resolutions will result in more recombinations per atom¹, which is why it is harder to obtain the same ionization fraction. This is due to sampling effects: when using a sampling criterion based on the density of gas, as is used in methods such as SPH and `SimpleX` (cf. Chapter 6), there is a bias towards high density regions. The available points will mainly be put in dense regions, and sparse regions will be undersampled. Thus, the density averaged over all particles will be too high, overestimating the number of recombinations (which is very sensitive to the density, scaling with n^2). Higher resolutions will more accurately sample low density regions, eliminating this effect.

Indeed, one sees in the left panel of Fig. 12.4, that the volume averaged fractions are converging as the effective resolution is increasing. At an effective resolution of 160^3 , convergence has not yet been reached, though. We would like to stress that all simulations in this thesis were performed on basic desktop machines. This is why we stopped at this `SimpleX` particle resolution, because it is the maximum effective resolution that a simple desktop computer fitted with 4GB of memory can handle. We could have used more elaborate shared memory machines, but we did not find this useful, having globally understood the effect. Of course, if we were not just aiming for understanding several effects, but for a complete and accurate simulation of reionization itself, we would need to increase the resolution until convergence has been achieved.

The input for cosmological radiative transfer solvers is provided by either dark matter only simulations, or, as in our case, by more accurate hydro solvers. The maximum resolution at which the reionization simulation can be performed is dictated by the resolution of these input data. In almost all cases, the radiative transfer methods used for these simulations use a much lower effective resolution, so as to obtain a feasible operation count. From

¹Because our radiative transfer method is highly adaptive, the effective resolution of our method is many orders of magnitude higher than that of the more standard methods used in Iliev et al. (2006) and Zahn et al. (2006). Thus, the clumping factors, or, equivalently, the substructure of dense regions, is much more resolved. As a consequence, the recombination rates (scaling with n^2) we find will be higher than the ones in those papers.

the analysis in this section, we can conclude that the resolution of the radiative transfer method alone can heavily influence the overall reionization history, because the global density structure will not be resolved properly, so that the recombination rates will be either over- or underestimated. When aiming for realistic reionization simulations, it is therefore mandatory to use an effective radiative transfer resolution that is identical to the one of the input.

To eliminate the influence of this resolution effect on the rest of the simulations in this chapter, we chose one effective radiative transfer resolution, namely $N_{\text{sub}} = 128^3$. This number of grid points is large enough to resolve most of the relevant gas substructure, and small enough to limit memory usage, so that many different simulations could be performed.

12.3.2 Star Formation Histories

Once a fiducial SimpleX resolution has been chosen, we can expect the outcome of the results to change when changing box sizes. A fixed resolution with increasing box size results in a lower effective mass resolution, by which high density regions, which have a high degree of clumping, cannot be resolved properly. This will underestimate the recombination rate and change the morphology of the emerging HII regions, to name just a few consequences. The most notable consequence, however, is the effect of the box size on the star formation history. The SPH code described in Schaye et al. (2006) forms stars (i.e. converts SPH particles into source particles) according to a recipe that is based on several parameters, such as the local density. The higher the resolution, the better high density haloes are resolved, and the more sources are formed.

This behavior can be quantified by plotting the star formation rate as a function of redshift as determined by the SPH code, for every one of the five box sizes. The result is depicted in Fig. 12.5, in which the logarithm of the star formation rates, in units of solar masses per year per Mpc³, is plotted versus the redshift. It is readily apparent that the larger the boxsize, or, equivalently, the smaller the mass resolution, the lower the star formation rate. Moreover, the formation of the first stars is retarded, when the resolution is lowered: for the $L = 1.5625/h$ Mpc box, the first star forms at $z = 25$, whilst, for the $L = 25.0000/h$ Mpc box, the first star forms only at $z = 15$! This is a direct consequence of the resolution requirements that were already discussed in Sect. 10.2.3, and in Naoz et al. (2006).

In order to be able to systematically check the impact of physical effects on the history of reionization for various different box sizes, we need to eliminate this box size dependence of the star formation rate. The most trivial approach is to choose one fiducial star formation history, and scale all others to match that one. To accomplish this, we fitted the star formation rate data of each box size with an exponential function of redshift² (i.e. $\text{SFR}(z) = \exp(a-bz)$, in which a and b are constants). The five fitting functions are plotted in the same figure, and the exact functional forms are given in the legend. In the low redshift regime, the fitting function for the $L = 6.2500/h$ Mpc box most closely resembles the suggested star formation rate models for $z \geq 6$ in both Strolger et al. (2004) and Fardal et al. (2006), so we choose this fitting function as the fiducial one.

Thus, we can eliminate the effect of box size on the star formation history, by rescaling the star formation histories of all other box sizes to that of the fiducial one. This can be achieved by increasing the masses of each source particle by a factor $\text{SFR}_{6.25}(z)/\text{SFR}_L(z)$ for box size L . When implementing this, one has to be careful that for box sizes larger

²This fitting function is motivated by the star formation rate models used in Strolger et al. (2004), cf. their Eq.(5).

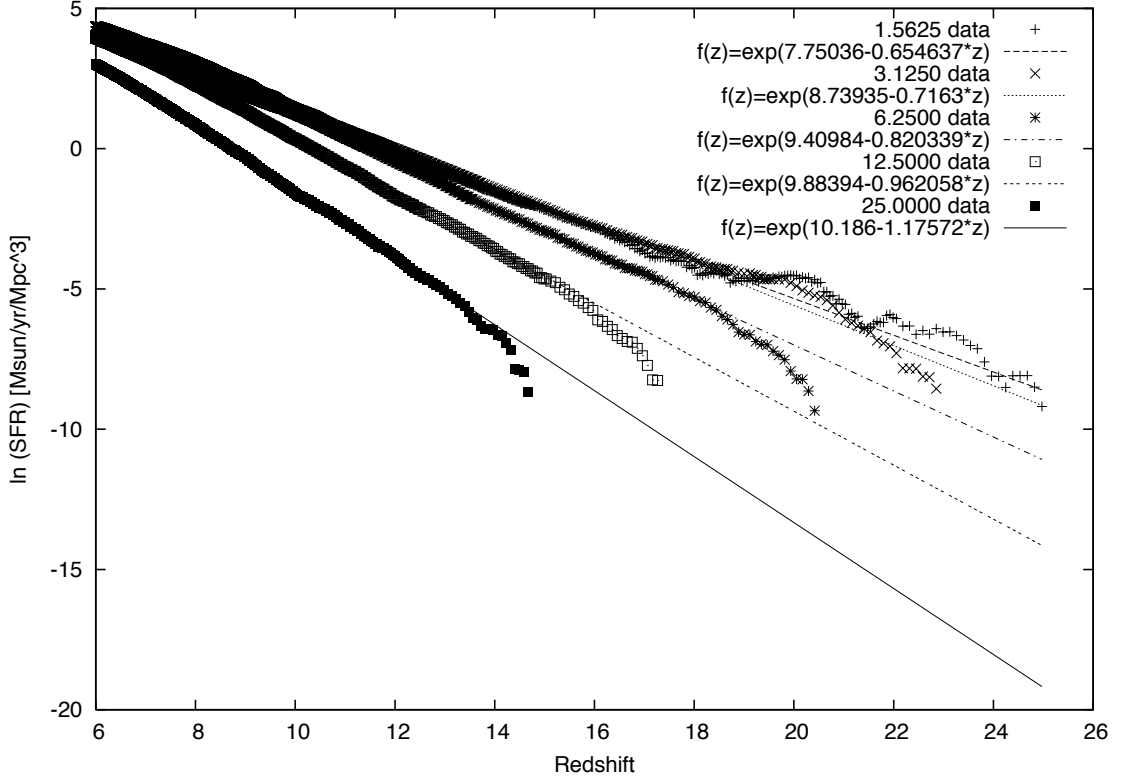


Fig. 12.5: Star formation rate data, in units of solar masses per year per Mpc^3 , as a function of redshift for several SPH simulations in the five different box sizes. Overlaid are the functions that were fitted to this data.

than $L = 6.2500/h$ Mpc, the first stars form later, so that these should be rescaled with a cumulative factor, i.e. the first source will be very massive to compensate for the retardation.

12.3.3 Results

When a fiducial radiative transfer resolution has been chosen, and the effect of differing star formation histories has been eliminated, we can compare the reionization histories results of the different box sizes. Several diagnostic results have been plotted in Fig. 12.6.

The top left panel depicts the volume averaged ionization fractions, cf. Eq.(12.1), as a function of redshift, for each of the boxes. As an alternative diagnostic, the top right panel shows the global mass averaged ionization fraction as a function of redshift, i.e.

$$\langle x \rangle_{\text{mass}} = \frac{\sum_i x_i m_i}{\sum_i m_i}, \quad (12.2)$$

in which $x_i = N_{\text{HII}}^i/N^i$ is the ionization fraction of particle i , and m_i is its mass. The volume and mass averaged fraction plots are nearly identical, and it is clear from both plots that all boxes have nearly identical reionization histories, except for a retardation factor: the largest boxes are reionized earlier. This retardation can be easily understood by recognizing that, although we have eliminated the effects on the star formation history, the size of the box still determines the mass resolution. The higher the mass resolution, the better dense regions are resolved, and the larger the local recombination rates. The recombination rate

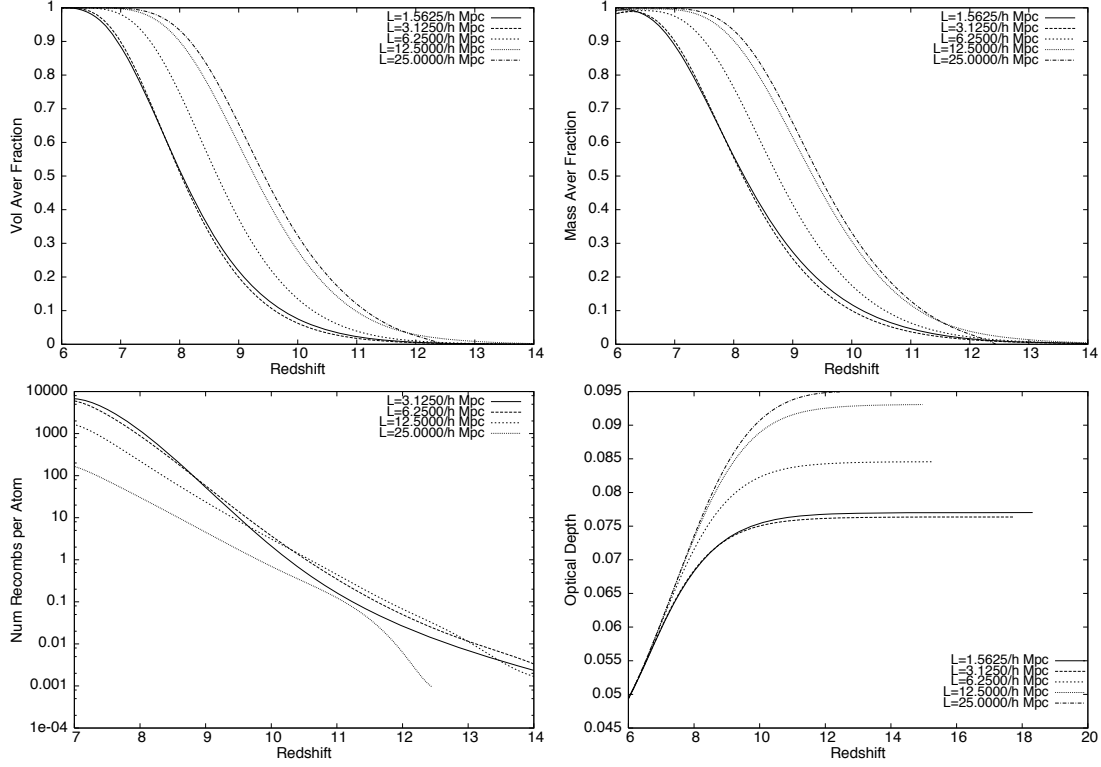


Fig. 12.6: Results of reionization simulations for five different box sizes. Top-left: Global volume averaged fractions as a function of redshift. Top-right: Global mass averaged fractions as a function of redshift. Bottom-left: Number of recombinations per atom as a function of redshift for each of the boxes. Bottom-right: Thomson optical depths, integrated from redshift z to us at $z = 0$, for each box.

is the most direct limiting factor of the speed of the ionization front. Thus, reionization is retarded, when more recombinations need to be compensated for. The bottom left panel of Fig. 12.6 depicts the global number of recombinations per atom as a function of redshift. It is difficult to compare the graphs for the different box sizes, because reionization commences at different times for each box. Still, it is clear that at, for example, $z = 12$, when the ionization fractions are still close together, the smallest boxes have the highest number of recombinations. This will immediately slow down the process of reionization with respect to the larger boxes.

This retardation effect is most aptly illustrated by determining the Thomson optical depth as a function as redshift, for each of the boxes. For a given reionization history, one can compute the mean optical depth along a line of sight between an observer at $z = 0$ and redshift z due to Thomson scattering by integrating

$$\tau(z) = c\sigma_{\text{T}} \int_z^0 dz' n_e(z') \frac{dt}{dz'}, \quad (12.3)$$

in which c is the speed of light, $\sigma_{\text{T}} = 6.65 \times 10^{-25} \text{ cm}^2$ is the Thomson scattering cross section, and $n_e(z)$ is the mean electron density at redshift z , given by

$$n_e(z) = n_{\text{H}}^0 (1+z)^3 \langle x \rangle_{\text{mass}}(z), \quad (12.4)$$

in which n_{H}^0 is the mean number density of hydrogen at present. The results of using SimpleX to reionize each box can be used to integrate Eq.(12.3) over each discrete time slice. This data unfortunately stops at $z = 6$, when reionization has ended. Fortunately, Eq.(12.3) has a closed analytical form, when assuming a constant $\langle x \rangle_{\text{mass}}$ between us and $z = 6$, $\Omega_0 + \Omega_{\Lambda} = 1$, and $X = 1$ (i.e. hydrogen only):

$$\tau(z) = \frac{2c\sigma_{\text{T}}\Omega_{\text{b}}\rho_{\text{crit},0}}{3H_0m_{\text{p}}\Omega_0} \langle x \rangle_{\text{mass}} \left\{ [\Omega_0(1+z)^3 + \Omega_{\Lambda}]^{1/2} - 1 \right\}, \quad (12.5)$$

in which we used Ω_{b} , Ω_{Λ} and Ω_0 , which are the total baryon, vacuum and matter densities in units of the critical density $\rho_{\text{crit},0}$. Moreover, H_0 is the Hubble parameter, and m_{p} is the mass of a proton. Assuming $\langle x \rangle_{\text{mass}} = 1$ for $z \leq 6$, we obtain, using the WMAP3 parameters, that $\tau(6) = 0.048$.

Using this machinery, we can plot the Thomson optical depths for each box size, and we refer to the bottom right panel of Fig. 12.6. The differences for each box size are more clear in this figure, because the Thomson optical depth is a cumulative quantity. The larger boxes have larger optical depths, because reionization commences earlier, so that more free electrons can contribute to the scattering optical depth from a certain redshift to us. Note that the optical depths all lie in the range $\tau \in [0.075, 0.095]$, which is in perfect agreement with the latest WMAP3 results (Page et al. 2006; Spergel et al. 2006), i.e. $\tau = 0.09 \pm 0.03$. This a very fortunate result, but should put in perspective: in this chapter, we are pointing out several effects that dramatically increase the uncertainties in the overall reionization results. Although these optical depth values may seem credible, an increase of effective resolution, either in the SPH code, or in the radiative transfer, may change the results drastically. Still, the optical depths for the $L = 1.56250/h$ and $3.1250/h$ Mpc are very close, and seem to have converged.

We conclude this section by stating that, although any effect on the star formation histories can be eliminated, the size of box still has a profound impact on the results of reionization simulations. The global reionization histories have the same profile, but they may show retardation. Any realistic simulation would therefore need to increase its effective SPH and radiative transfer resolution until the results for each box size have converged onto one fixed simulation outcome. Moreover, one needs to check the results for differing box sizes with a fixed effective mass resolution. For now, this ideal resolution is far beyond our computational capabilities, cf. Sect. 10.2.3, so that one has to be acutely aware of the limitations of current reionization simulations.

12.4

Reheating

Having studied the effects of several numerical aspects on the results of reionization simulations, we are in a position to more systematically study the impact of additional physical effects. The first of these will be that of reheating.

As explained more elaborately in Sect. 10.3.1, the mean free paths of the UV photons sharply increase when the HII regions start to overlap during the process of reionization. This effectively results in a rapid increase of the intensity of the UV background, which will interact with the intergalactic medium by photoheating it. The increase in gas temperature will suppress star formation, and it will also decrease clumping in high density regions, thereby lowering the local recombination rates.

Both of these effects will be studied in this section. Unfortunately, our transfer method has not yet been dynamically coupled to the hydro code, so a fully self-consistent radiation-hydrodynamic description of the radiative feedback has not yet been implemented. We can incorporate the effect of reheating, however, by externally turning on an ambient source of heating in the SPH code at a predesignated redshift. Referring to the top left panel of Fig. 12.6, one can see that for all box sizes the volume has been at least half ionized at $z = 9$. It is this redshift at which we will turn on the reheating. The SPH code described in Schaye et al. (2006) implements this reheating by giving a certain amount of energy to each particle over an extended redshift range starting sharply at the reheating redshift. Usually, the heating is distributed in the shape of a Gaussian, but since the particles take discrete steps, error functions are used. The parameters (amount of energy, sigma of Gaussian) for the hydrogen reionization heating are calibrated such that the temperature at the mean density of the Universe is 10^4 K (assuming a reionization redshift of $z = 9$).

Using this recipe, we are able to examine the effect of reheating on both the star formation rate and the clumping of the intergalactic medium. We will first only study the effect of the star formation rate by using the modified star formation rate, but hereby using the gas distribution that was produced in the simulation *without* reheating. The effect of clumping can only be studied by using the star formation histories of Fig. 12.5, i.e. those without reheating, in combination with the gas distribution as obtained from the SPH simulation *with* reheating. Once these two effects have been studied separately, we can examine their total effect by using the modified reheating star formation histories in combination with the reheated gas distribution.

12.4.1 Star Formation Rate Effect

When the temperature of the intergalactic medium increases, the Jeans mass follows this trend. It is therefore to be expected that fewer stars will form. This is indeed what we see when plotting the star formation rate as a function of redshift for SPH runs with and without reheating. Because we know that the overall results of the simulations depend heavily on the effective mass resolution, or, equivalently, the box size, we choose two of the box sizes, namely the $L = 3.1250/h$ and the $12.5000/h$ Mpc boxes, and study the effect of reheating on both. The star formation histories for both boxes are plotted in the top left and right panel of Fig. 12.7. The data that corresponds to the star formation histories without reheating is identical to the data of the corresponding box sizes plotted in Fig. 12.5. It is apparent from the figures that the star formation rate is suppressed, when reheating is turned on at $z = 9$.

The effect of this suppression on the process of reionization can be quantified by repeating the reionization simulations using the original gas distribution without reheating, but with the modified star formation histories. To accomplish this, we need to determine the fitting functions $\text{SFR}_L(z)$ to the modified star formation rate data for $z \leq 9$. The data and the fitting functions are shown in the bottom panels of Fig. 12.7. Once these new fits are known, we can again eliminate the numerical effects of the star formation history, as discussed in Sect. 12.3.2, by rescaling these new functions to the same fiducial fitting function $\text{SFR}_{6.25}(z)$ that was given in that same section. We take this fitting function, and not the one for the modified star formation history of the $L = 6.2500/h$ Mpc box, because that is most consistent with the scaling in the $z \geq 9$ regime.

The results of the new simulations are shown in Fig. 12.8. The top panels depict the volume averaged ionized fraction and the Thomson optical depth as a function of redshift

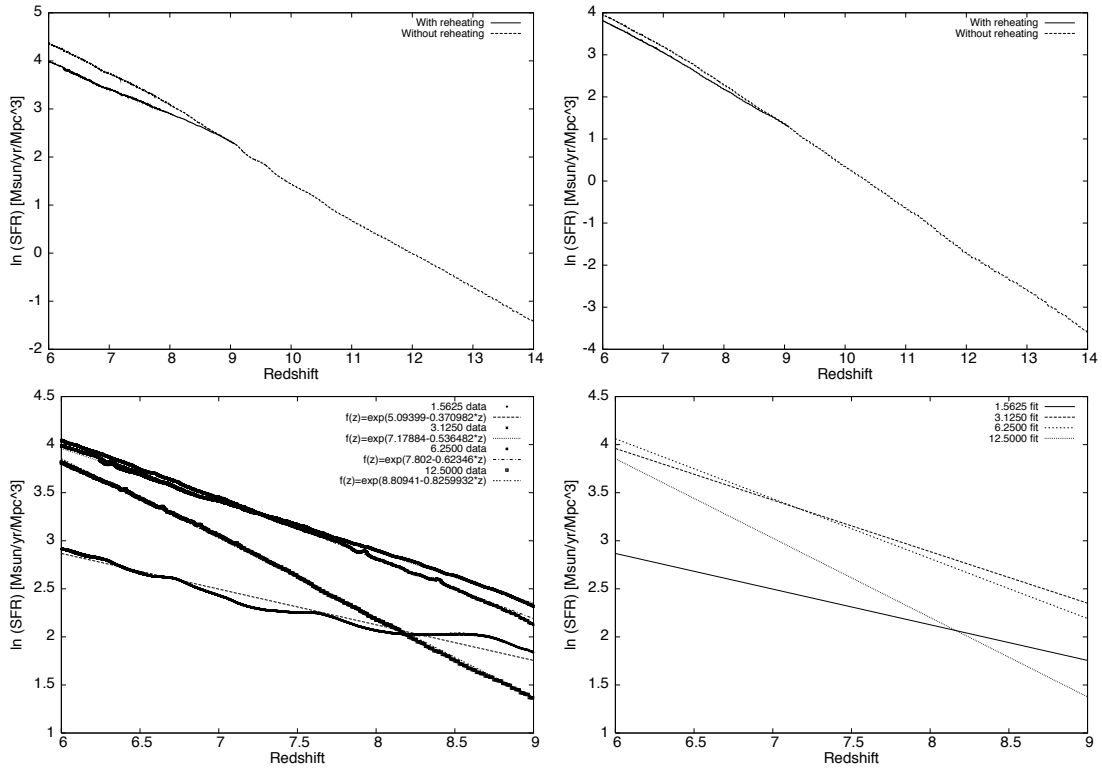


Fig. 12.7: Plots of the modified star formation histories as a result of reheating. Top: Star formation histories, in units of solar masses per year per Mpc^3 , as a function of redshift, for the $L = 3.1250/h$ and the $3.1250/h$ Mpc boxes. Results with and without reheating are both given. Bottom: The star formation history data with reheating for all boxes. The modified fitting functions are also given. Note the different redshift ranges in the top and bottom panels.

for the $L = 3.1250/h$ Mpc box, and the bottom panels do the same, but now for the $L = 12.5000/h$ Mpc box. All panels show the results with and without reheating. Note that the scaling of the modified star formation histories to the fiducial fitting function $\text{SFR}_{6.25}(z)$ will have as an effect that, for the $L = 3.1250/h$ Mpc box, the star formation rate is actually boosted relative to the one without reheating! This is why reionization is achieved earlier, as can be seen in the top left panel of Fig. 12.8. This is not the case for the $L = 12.5000/h$ Mpc box, for which the star formation history is still suppressed after rescaling.

It is immediately clear that, although the differences in star formation histories as depicted in Fig. 12.7 may seem severe, the overall results on the process of reionization are just marginally altered. We should note that the effect of reheating is underestimated, if H_2 coolers are important.

12.4.2 Clumping Effect

The effect of reheating on the clumping of the gas can be isolated by performing reionization simulations on the new gas distribution that is modified by reheating, but with the original star formation histories that were discussed in Sect. 12.3.2. The results hereof are depicted in Fig. 12.9. The left hand panels show the number of recombinations per atom, the volume averaged ionization fraction, and the Thomson optical depth, as a function of redshift, for

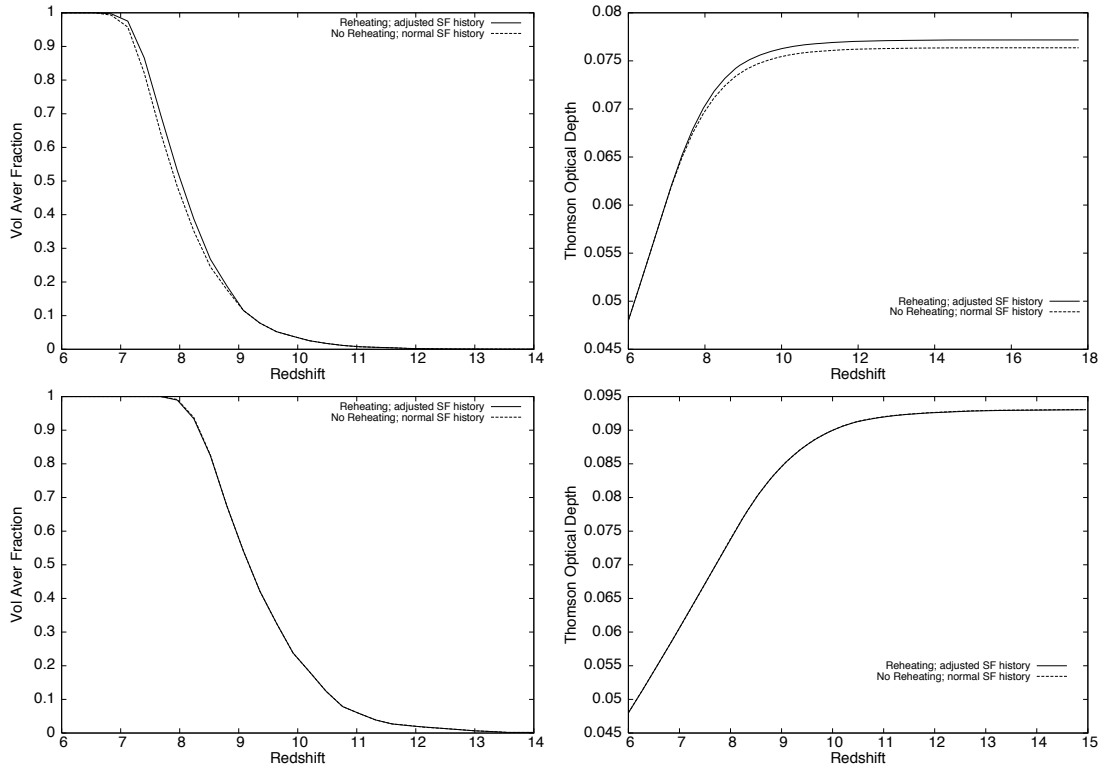


Fig. 12.8: Reionization results for modified star formation histories, but with original gas distributions. Top: Volume averaged ionization fraction and Thomson optical depth as a function of redshift for the $L = 3.1250/h$ Mpc box. Bottom: *Idem*, but now for the $L = 12.5000/h$ Mpc box. Both the results with and without reheating are depicted.

the box with comoving width $L = 3.1250/h$ Mpc. The right hand panels show the same, but now for the $L = 12.5000/h$ Mpc box. All panels show the results with and without reheating.

It is immediately apparent that, because clumping is reduced by reheating, high density regions are less dense, so that the number of recombinations is drastically lowered at reheating redshift $z = 9$. The speed of the ionization fronts is directly influenced by the local recombination rates, cf. Eq.(7.8), so a decrease in the number of recombinations will result in an acceleration in the process of reionization. This is indeed what we see in the plots of the volume averaged ionization fractions. The effect is less apparent in the larger of the two boxes, though. This has two reasons: first, reionization has already progressed further at the predefined reheating redshift for the larger box, so that reheating is turned on a bit too late; second, clumping is much more dominant in the smaller boxes, because the mass resolution is higher, so that any change in clumping due to reheating will have more drastic consequences. This difference in the impact of reheating is exemplified in the plots of the Thomson optical depths for both boxes. The process of reionization is accelerated, by which there are more free electrons at earlier redshifts. Hence, the Thomson optical depths are boosted for both of the boxes.

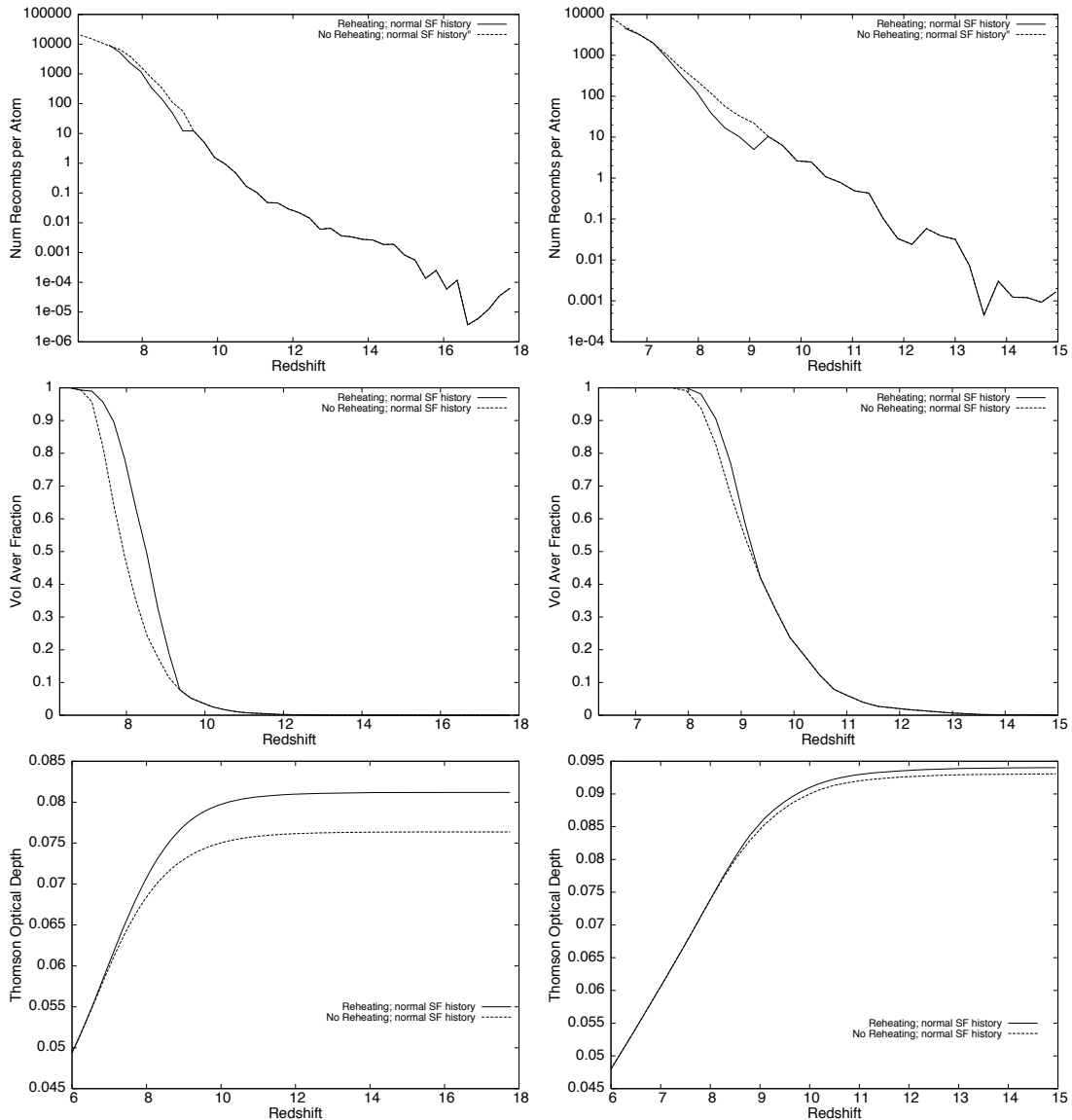


Fig. 12.9: Reionization results for reheated gas distributions, but with original star formation histories. Left: The number of recombinations per atom, the volume averaged ionization fraction and the Thomson optical depth as a function of redshift for the $L = 3.1250/h$ Mpc box. Right: Idem, but now for the $L = 12.5000/h$ Mpc box. Both the results with and without reheating are depicted. Note that both boxes have a different redshift range, because the smallest box forms stars earlier.

12.4.3 Total Effect

It is apparent from the results of separately studying the effect of reheating on the star formation histories and on the clumping of the gas, that it is the latter that has the most dominant effect on the history of reionization. Indeed, when we repeat the reionization simulations with both the modified star formation histories *and* the reheated gas distributions, we find that the results for the total effect of reheating, cf. Fig. 12.10, closely resemble those for the effect of clumping alone.

We can conclude from the results in this section that the physical effect of reheating has

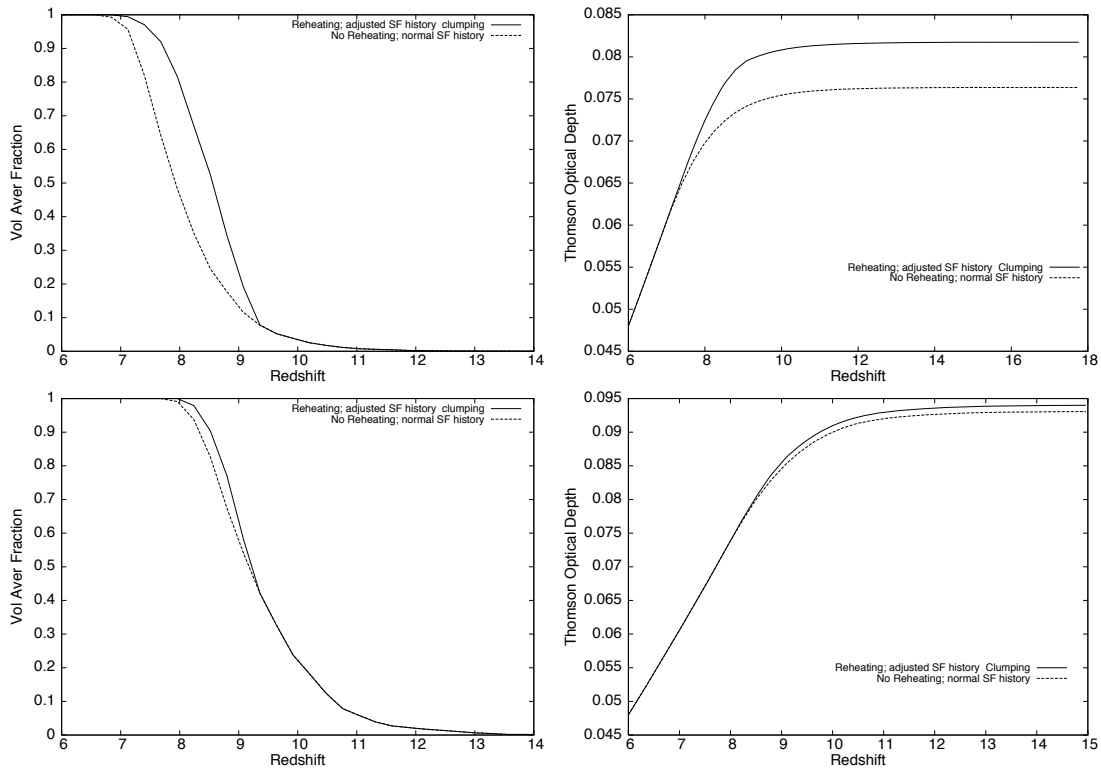


Fig. 12.10: Reionization results for both modified star formation histories and reheated gas distributions. Top: Volume averaged ionization fraction and Thomson optical depth as a function of redshift for the $L = 3.1250/h$ Mpc box. Bottom: *Idem*, but now for the $L = 12.5000/h$ Mpc box. Both the results with and without reheating are depicted.

a profound influence on the process of reionization, even for the larger boxes. Although the star formation history is altered by the suppression induced by the reheating, the decrease in clumping in highly dense regions most drastically influences the simulation results. It is well known that most photons are consumed in these dense regions, which is why any change there is bound to have a large impact. As an example, the Thomson optical depth for the reionization simulations in the $L = 3.1250/h$ Mpc box changes from $\tau = 0.075$ to $\tau = 0.082$.

Although the effect on the star formation history can be neglected to some extent, the effect of reheating on the clumping of the gas cannot be ignored: reheating accelerates reionization. This effect would be even more dramatic, when reheating would be implemented self-consistently, instead of turning it on by hand at a predesignated redshift. Most of the photons are produced in regions where the clumping is high, so that their heating effect immediately reduces clumping.

It is evident that any realistic reionization simulation *must* include the effect of reheating. This immediately rules out the use of dark matter only solvers in setting up reionization simulations, because these cannot capture the effect of reheating on the gas. Thus, hydrodynamics solvers are an indispensable ingredient in numerical studies of the epoch of reionization.

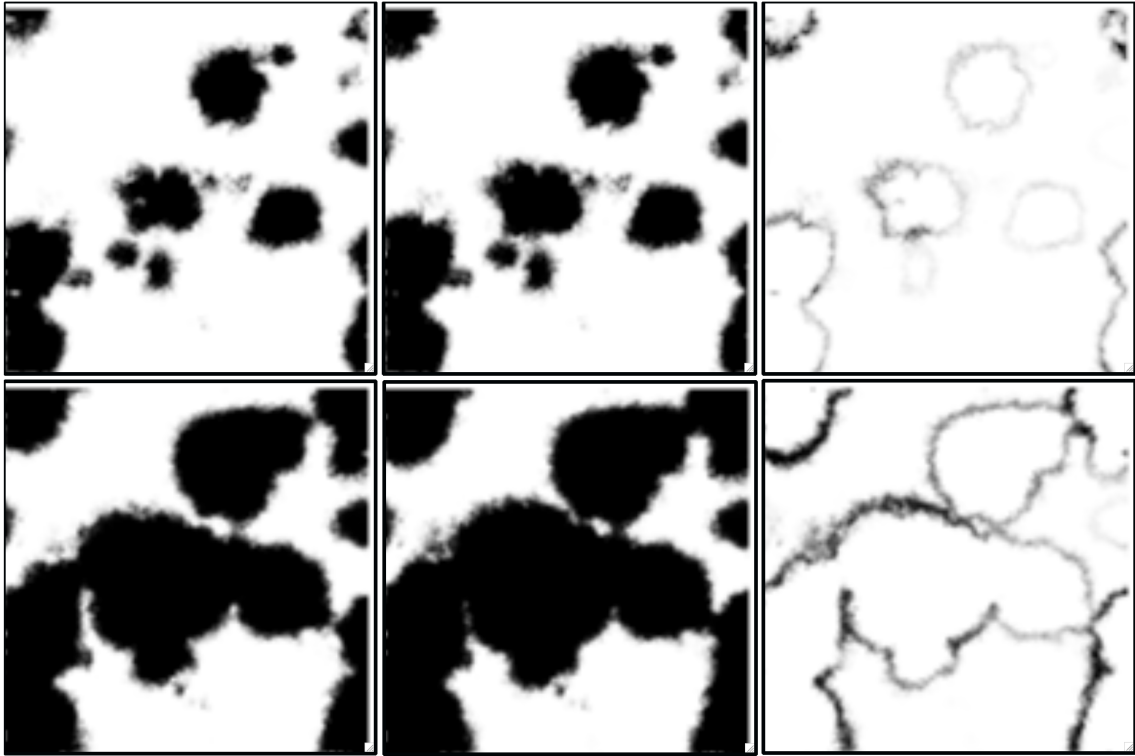


Fig. 12.11: Visualizations of the HII distributions for two reionization runs, one with and one without resolving the diffuse radiation field, for a box of comoving width $L = 1.56250/h$ Mpc. Depicted are 2D slices at redshifts $z = 9.08$ (top), and $z = 8.52$ (bottom). Left: Runs without diffuse photons. Middle: Runs with diffuse photons. Right: Difference plots, assuming Boolean ionization fractions (i.e. either 1 or 0); left is subtracted from middle.

12.5

Diffuse Photons

The final physical effect we have studied is that of the diffuse radiation field. In Chapter 11, we showed quantitatively that, for realistic cosmological gas density distributions, the diffuse photons dominate the radiation field inside HII regions. Moreover, only the diffuse photons push against the ionization fronts. It is therefore mandatory that a realistic treatment of the radiative transfer accurately incorporates this part of the total radiation field. Accomplishing this can be shown to be a formidable computational task: every parcel of gas that re-emits photons has to be added to the list of sources, so that, in principle, each of the 256^3 SPH particles can become a source. Most common cosmological radiative transfer methods scale with the number of sources, cf. Chapter 7, which is why including the diffuse radiation field is too demanding for these methods, and the usual on the spot approximation, shown to be faulty in the previous chapter, is invoked. Fortunately, the SimpleX transfer method does not scale with the number of sources, cf. Chapter 4, so that we are in the fortunate position that we can study the effect of these diffuse photons.

Including the diffuse radiation field will not affect the global energy budget, i.e. the number of photons emitted minus the number of recombinations will not change. This is why the position of the Strömgren sphere in an isotropic density distribution will remain

fixed, as explained in Chapter 11. Thus, the global diagnostics, such as the volume averaged ionization fractions, and the Thomson optical depths, presented in the previous sections, will remain unaltered when including diffuse photons. The effect of the diffuse radiation field can not be seen globally, but needs to be examined on a local scale. Indeed, it is the *morphology* of the pattern of HII regions as a whole that will change. This is very important for future 21-cm observations of the EOR: the observed maps can only be seriously compared to simulations when this effect is understood.

As a test case, we use as input an SPH simulation for a box with comoving width $L = 1.56250/h$ Mpc with the usual parameters (i.e. WMAP3 cosmology, 256^3 particles, etc.), and use `SimpleX` to solve for the reionization both with and without diffuse photons. We will use the results at three different representative redshifts to compare the structure of the resultant HII regions: $z = 9.64$, 9.08 , and 8.52 , at which the volume averaged ionization fractions are $\langle x \rangle_{\text{vol}} = 0.189$, 0.333 , and 0.543 , respectively. A visualization for the latter two redshifts is given in Fig. 12.11, showing a 2D slice of the HII distribution for a simulation without diffuse photons (left), and with diffuse photons (middle). When we assume that the ionized fractions can only be Boolean, i.e. 1 or 0, or, equivalently, black or white, we can make a difference plot between the results with and without diffuse photons, and we refer to the right panels of Fig. 12.11.

By visual inspection, we immediately see that the runs including diffuse photons produce HII regions that are a bit more extended, maybe even rounder, than the runs without. For comparison with future observations, this needs to be quantified. In this section, we will look at two possible options for doing this.

12.5.1 Power Spectrum

The most obvious choice for studying the structure of patterns in a computational box is the power spectrum. For a three-dimensional field, one can compute its Fourier spectrum, and use that to infer the 3D power spectrum. This 3D power spectrum is often collapsed into a one-dimensional one by averaging it over spherical bins centered on the origin of the frequency domain. This 1D spectrum based on a 3D density field is the most commonly used tool to study the distribution of HII regions in cosmological reionization simulations, see e.g. Iliev et al. (2006), Zahn et al. (2006), but also the semi-analytical approach in Furlanetto et al. (2004).

We have used this recipe to construct the power spectra of the reionization simulation with and without diffuse photons, for all three specified redshifts. The result is depicted in Fig. 12.12. The left panel shows the scale-free power $k^3 P(k)$ of the neutral fraction as a function of the wavelength mode k for the runs without diffuse photons, whilst the right panel shows the same for the runs that do include the diffuse radiation field. Note that, although values are shown for larger scales, the largest reliable wavenumber value for this box is $k = 2\pi/2L = 0.32 \cdot 2\pi h/\text{Mpc}$. Similarly, we have to be careful at the small scales: Poisson noise may affect the results there.

We can immediately observe two things. First, it is remarkable that there is more power at small scales than at large ones, because this somewhat contradicts what was found in Iliev et al. (2006) and Zahn et al. (2006). This is probably due to the fact that we use, in contrast with their methods, an adaptive mesh that can resolve much smaller scales. Second, we see that, as reionization progresses, the power at the large scales ($k \sim 0.5 h/\text{Mpc}$) increases. This is to be expected, because the initially small HII regions will grow and merge into larger

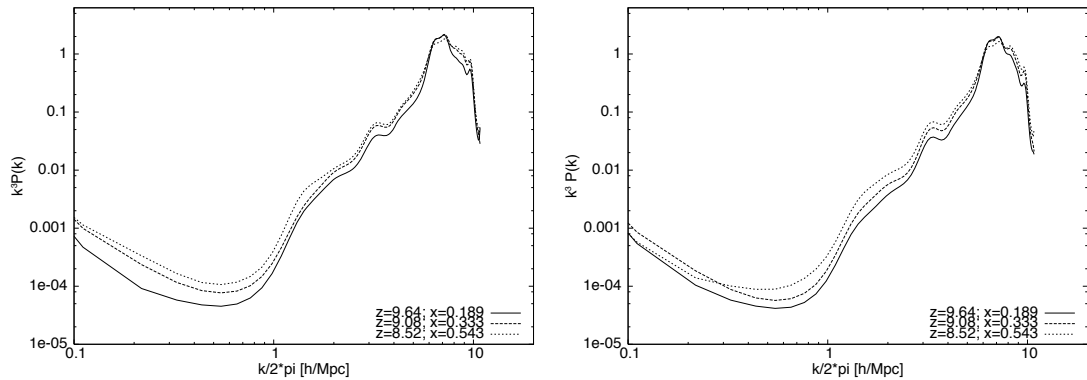


Fig. 12.12: Dimensionless power spectra $k^3 P(k)$ of the neutral fraction for two reionization runs, one without diffuse photons (left), and one with (right). In each figure, spectra are plotted for three different redshifts, $z = 9.64, 9.08,$ and 8.52 , each corresponding to an increasing volume averaged ionization fraction.

ones.

Although there is some minor difference between the spectra in the two different panels, nothing much can be inferred from that. Based on their power spectra, the results with and without diffuse photons are nearly identical. Thus, it is obvious that we need a more refined tool for analyzing the differences that could be already observed by visual inspection only.

12.5.2 Morphology

Although power spectra have proven their use in many different problems, their use is limited in the analysis of reionization simulations. Their major disadvantage: any information of the phase is immediately thrown away, when computing the norms of each wavelength mode. To give a more accurate description of the patterns that are to be observed in future 21-cm observations, we need an unambiguous method for quantifying the morphology of HII patterns.

There are several options for accomplishing this, the most elegant of which is the use of the *Minkowski functionals*. The Minkowski functionals (Hadwiger 1957) provide a complete set of measures of the topology of any structure, but are most often used to examine the morphological properties of a density field in \mathbb{R}^d . Their use in the context of cosmological reionization was suggested in Gleser et al. (2006), which describes the procedure for computing the Minkowski functionals for a three-dimensional field $u(\mathbf{x})$ in a box of volume V .

To accomplish this, one determines an excursion set F_ν which includes all the points that satisfy $u(\mathbf{x}) \geq \nu\sigma$, in which $\nu = u_{\text{thr}}/\sigma$ is the threshold value rescaled to the rms value $\sigma^2 = \langle u^2 \rangle$. The global morphology of $u(\mathbf{x})$ can then be described by calculating the Minkowski functionals as a function of u_{thr} , which defines an isosurface $S(\mathbf{x})$. There are four

of these functionals in \mathbb{R}^3 :

$$V_0(\nu) = \frac{1}{V} \int_V d^3x \Theta[\nu\sigma - u(\mathbf{x})], \quad (12.6)$$

$$V_1(\nu) = \frac{1}{6V} \int_{\partial F_\nu} d^2S(\mathbf{x}), \quad (12.7)$$

$$V_2(\nu) = \frac{1}{6\pi V} \int_{\partial F_\nu} d^2S(\mathbf{x}) [\kappa_1(\mathbf{x}) + \kappa_2(\mathbf{x})], \quad (12.8)$$

$$V_3(\nu) = \frac{1}{4\pi V} \int_{\partial F_\nu} d^2S(\mathbf{x}) \kappa_1(\mathbf{x}) \kappa_2(\mathbf{x}), \quad (12.9)$$

in which Θ is the Heaviside step function, and κ_1 and κ_2 are the principal curvatures (i.e. the inverse of the principal radii) at a point \mathbf{x} on the surface.

The first functional, $V_0(\nu)$, is the total volume of the regions with u above the threshold value $u_{\text{thr}} = \nu\sigma$. The other three functionals are computed by performing a surface integration over the boundary, ∂F_ν , of the excursion set. The functional $V_1(\nu)$ is a measure for the surface area of that boundary, $V_2(\nu)$ is the mean curvature over the surface, and $V_3(\nu)$ is the Euler characteristic χ . The Euler characteristic can be expressed in terms of the genus, g , of the regions, i.e. $\chi = 2(1 - g)$, where a sphere has genus $g = 0$, a torus has $g = 1$, etc.

Both the surface area and the integrated mean curvature functionals describe the shape of the regions at the threshold value. Pancake shaped regions, for example, are characterized by a large surface area and comparatively small mean curvature, whilst the opposite is the case for cigar shaped regions. These two functionals, $V_1(\nu)$ and $V_2(\nu)$, are therefore expected to be a perfect diagnostic for quantifying the differences between the pattern of HII regions in the results with and without diffuse photons.

Gleser et al. (2006) describe two efficient numerical methods for obtaining these four Minkowski functionals. As an extra bonus, they provide a ready-made source code package³ that implements both of these methods. This enables us to readily compute the Minkowski functionals of the reionization simulations with and without diffuse photons. As a scalar field $u(\mathbf{x})$, we choose the local neutral fraction $x(\mathbf{x}) = n_{\text{HI}}(\mathbf{x})/n(\mathbf{x})$. The results hereof are shown in Fig. 12.13.

The top panels depict the first Minkowski functional for the run without (left) and with the diffuse photons (right). The two figures are nearly identical, with the value of the functionals decreasing as time progresses. As the box is slowly ionized, the volume of the neutral hydrogen steadily decreases, so the evolution of this functional is not very surprising.

More striking are the differences when comparing the $V_2(\nu)$ and $V_3(\nu)$ functionals, depicted in the middle and bottom panel of Fig. 12.13, respectively. In the $n_{\text{thr}}/n_{\text{H}} \leq 1$ regime, which composes most of the volume (cf. the first Minkowski functional V_0), the $V_1(\nu)$ functional of the run with diffuse photons is consistently *lower* than for the run without diffuse photons. Moreover, the $V_2(\nu)$ functional for the run with diffuse photons is consistently *higher* than for the run without diffuse photons. This is a direct quantification of the difference in morphology of reionization with and without diffuse photons: when the diffuse radiation field is included, the global morphology of HII regions is more spherical.

Thus, it can be concluded that, although the use of diffuse photons will not alter global diagnostics that keep track of the energy budget, as is the case for other physical effects such as reheating, it will have a profound impact on the morphology of reionization. We

³Available at <http://physics.technion.ac.il/~lirong/>.

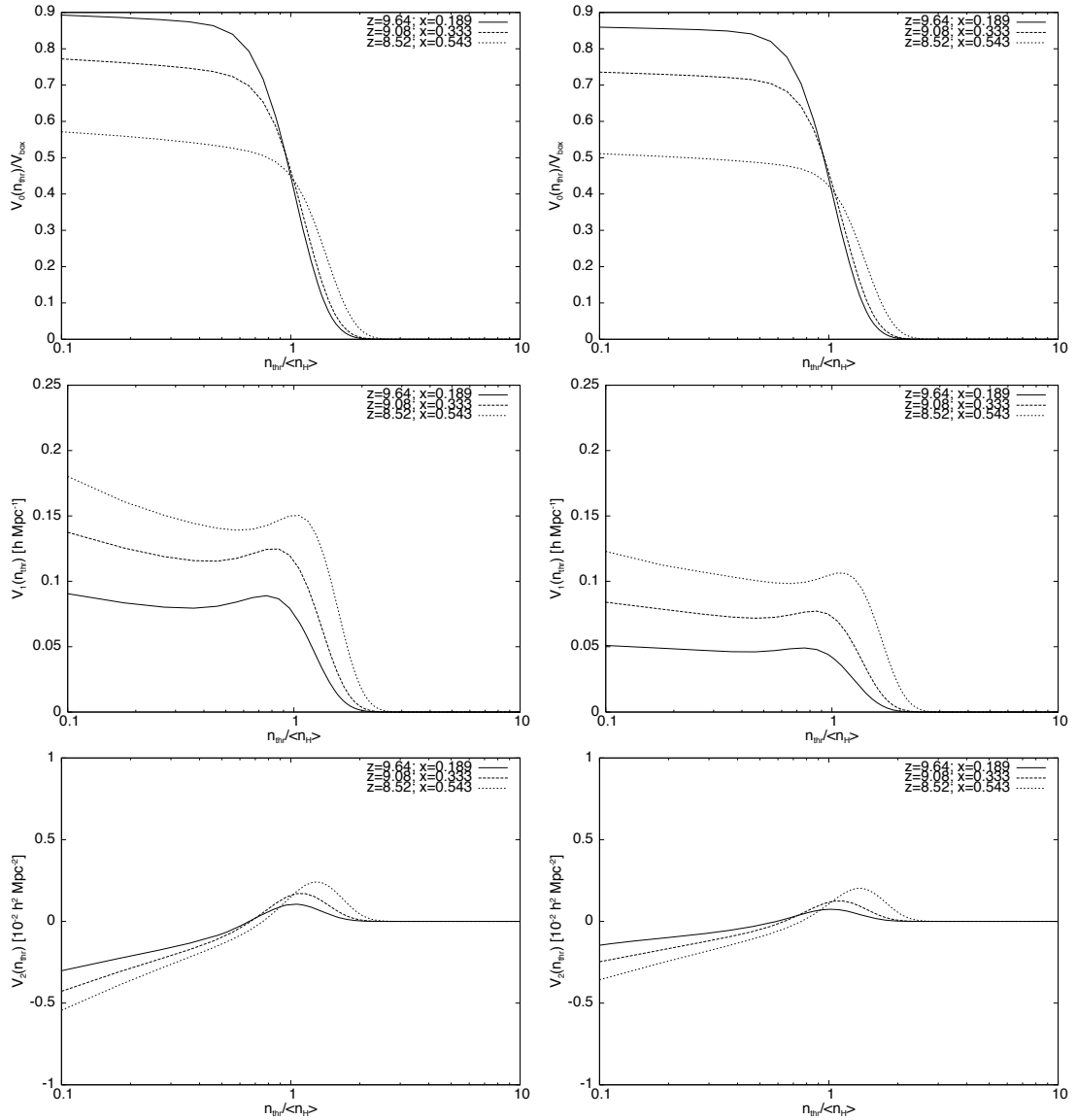


Fig. 12.13: The first three Minkowski functionals $V_0(\nu)$ (top), $V_1(\nu)$ (middle), and $V_2(\nu)$ (bottom), for reionization simulations without (left), and with diffuse photons (right). Each figure gives values for its functional at three redshifts $z = 9.64, 9.08,$ and 8.52 .

have shown quantitatively that the global shape of HII regions will be less pancake shaped, and more spherical. This is of great importance to future 21-cm observations of the epoch of reionization: the predicted 21-cm maps, based on numerical simulations, will have to be refined; and, conversely, when such maps are observed, conclusions based on these observations will be different when this effect is included. It is therefore mandatory, for any realistic reionization simulations that wishes to produce accurate 21-cm maps, to include the effect of the diffuse radiation field.

Conclusions

Simulations of the epoch of reionization are at the forefront of numerical astrophysics. Almost every imaginable physical ingredient has to be incorporated in one way or another to cover the full range of physical processes that take the Universe from a linear post-recombination phase to one that is able to form the first sources in collapsed haloes, and eventually to one that is filled with diffuse photons traveling through and interacting with the inhomogeneous intergalactic medium.

The numerical simulations that can successfully incorporate most of these processes will be very intricate, to say the least. In Chapter 10, we already pointed out that simulations that track the full range in resolution that is needed to accurately sample the pre-reionization density field are far beyond our current computational capabilities. Thus, in order to be able to perform at least a representative reionization simulation, several approximations are often invoked. Smaller box sizes and lower resolutions are used, hydrodynamical effects are ignored, and radiative transfer is treated in an approximate way, if it is included numerically at all.

In this chapter, we have quantitatively tested the validity of a variety of such approximations. The state-of-the-art cosmological SPH simulations, coupled to the adaptive cosmological radiative transfer code `SimpleX`, enabled us to determine the impact of several of these effects. After we have systematically checked the influence of box size and resolution on the star formation histories and the overall reionization results, we were in a position to directly measure the effect of two physical processes that are almost always ignored: the effect of reheating due to the UV background, and that of the diffuse radiation field.

The results showed that the effect of reheating cannot be ignored. The effect of this reheating on the star formation history, but, most of all, on the clumping of the gas, has as a consequence that the process of reionization is accelerated by a non-negligible factor. Thus, reionization is completed at earlier redshifts, and the mean Thomson optical depths computed are higher. Including the diffuse radiation field in the simulations produced results that differed from those without diffuse photons. We showed that not the power spectra but the Minkowski functionals can give an accurate and quantitative description of these differences: including diffuse photons will cause the HII regions, and thus also the observed 21-cm maps, to be less pancake shaped, and more spherical. Last but not least, we have shown that, even when all these physical effects are included, any simulation has to be rerun for many different box sizes and resolutions, in order to check whether or not convergence has been reached.

Thus, we can conclude that modern day reionization simulations need to be refined. Hydrodynamics methods have to be included to solve for non-negligible effects like reheating, and radiative transfer solvers must include diffuse photons in order to produce accurate 21-cm maps. Methods that use adaptive meshes are preferred, because large resolution ranges are mandatory. Only then can the results of any large scale reionization simulation be representative for the actual physical processes that took place during this important cosmological era.

Bibliography

- Bruzual, G. & Charlot, S. 2003, MNRAS, 344, 1000
Chabrier, G. 2003, PASP, 115, 763
Fardal, M. A., Katz, N., Weinberg, D. H., & Davé, R. 2006, ArXiv Astrophysics e-prints, astro-ph/0604534
Furlanetto, S. R., Zaldarriaga, M., & Hernquist, L. 2004, ApJ, 613, 1
Gleser, L., Nusser, A., Ciardi, B., & Desjacques, V. 2006, MNRAS, 370, 1329
Hadwiger, H. 1957, Vorlesungen über Inhalt, Oberfläche und Isoperimetrie (Berlin: Springer)
Iliev, I. T., Mellema, G., Pen, U.-L., et al. 2006, MNRAS, 369, 1625
Naoz, S., Noter, S., & Barkana, R. 2006, MNRAS, 373, L98
Page, L., Hinshaw, G., Komatsu, E., et al. 2006, ArXiv Astrophysics e-prints, astro-ph/0603450
Schaye et al. 2006, in preparation
Seljak, U. & Zaldarriaga, M. 1996, ApJ, 469, 437
Spergel, D. N., Bean, R., Dore, O., et al. 2006, ArXiv Astrophysics e-prints, astro-ph/0603449
Springel, V. 2005, MNRAS, 364, 1105
Strolger, L.-G., Riess, A. G., Dahlen, T., et al. 2004, ApJ, 613, 200
White, S. D. M. 1994, ArXiv Astrophysics e-prints, astro-ph/9410043
Zahn, O., Lidz, A., McQuinn, M., et al. 2006, ArXiv Astrophysics e-prints, astro-ph/0604177

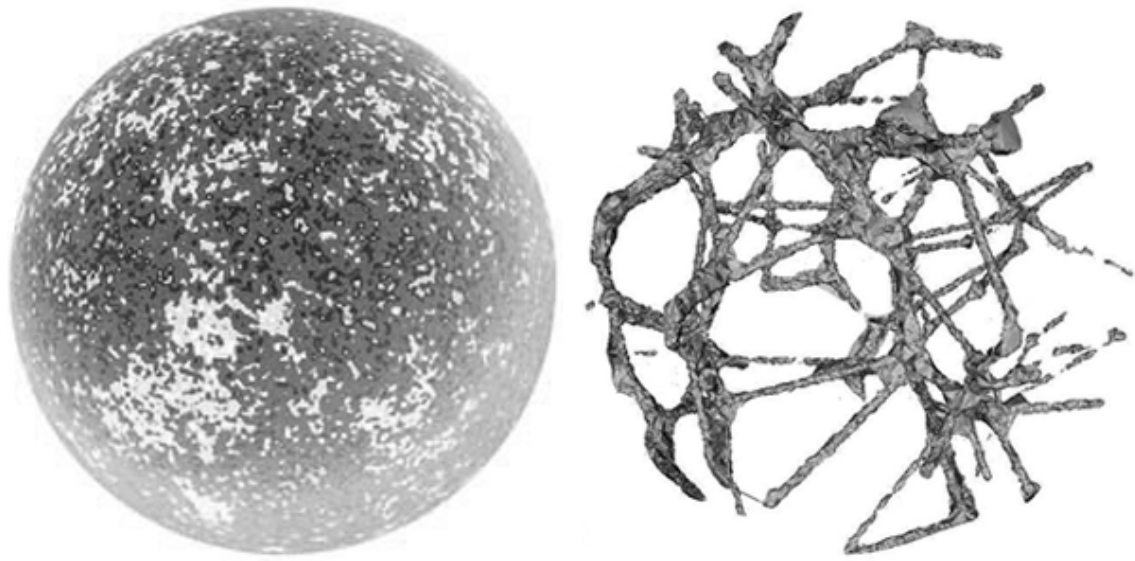
Samenvatting

Volgens de meest recente metingen is het heelal nu ongeveer 13.7 miljard jaar oud. Het lokale universum, zoals we dat kunnen zien met het blote oog op een heldere nacht, bestaat uit sterren en sterrenstelsels, en gas en stof. En, daartussen, heel veel lege ruimte.

Dit was ooit anders. Toen het nog jong was, was het heelal gevuld met een bijna volledig homogene oersoep, een plasma van protonen, neutronen en lichtdeeltjes (fotonen), waaruit alles wat we om ons heen zien is ontstaan. De oersoep was niet perfect glad: kleine rimpelingen, ontstaan aan het begin van de tijd, verstoorden het normaal gesproken rimpelloos oppervlak. Door de uitdijning van het heelal werd deze oersoep steeds verder verdund, totdat het op een bepaald moment, zo'n 400.000 jaar na de oerknal, zo transparant werd dat de lichtdeeltjes konden ontsnappen, en de protonen en electronen zich samen konden voegen tot neutraal waterstof. Die fotonen kunnen we nog steeds zien aan de hemel, als de nagloed van de oerknal. In het microgolf gedeelte van het spectrum van licht, zien we deze zogenaamde *kosmische achtergrondstraling*, die inmiddels door de verder uitdijning van het heelal is afgekoeld tot 2.73 graden boven het absolute nulpunt. Moderne precisie-experimenten hebben de temperatuur van de achtergrondstraling zeer nauwkeurig in kaart gebracht (zie Fig. 13.1, links). Deze kaart geeft een directe visuele representatie van de rimpelingen in de oersoep op het moment dat de fotonen ontsnapten. Opmerkelijk is dat deze rimpelingen slechts in de orde van een miljoenste graad zijn: de inhoud van het heelal was inderdaad nagenoeg glad.

We mogen blij zijn dat deze rimpelingen er waren. Onder de invloed van de zwaartekracht worden de minieme inhomogeniteiten in de loop van de tijd uitvergroet, doordat de materie zich samentrekt naar plaatsen met hogere dichtheid. Uiteindelijk ontstaat er een geometrische complexe verdeling van materie, die het meest doet denken aan een bijenkorf (zie Fig. 13.1, rechts). Deze structuur, waarneembaar aan de hemel als we de positie van sterrenstelsels uitzetten in een kaart, herbergt het ontstaan van de eerste sterren, planeten, en misschien zelfs wel leven.

Herakleitos' aforisme vat het het meest beknopt samen: *Παντα ρει και ουδεν μενει*. Veranderend van een gladde soep naar een sponsachtige grote schaal structuur, de inhoud van het heelal is immer in beweging. Elke keer als er een gradient is, in druk of potentiaal, komen baryonische en donkere materie deeltjes in beweging. Fotonen en andere massaloze deeltjes zijn constant in beweging, verplicht zich te houden aan de lichtsnelheid. Als dit natuurlijke proces niet bestond, dan zou het heelal nog steeds in dezelfde toestand zijn als die gezien kan worden in de achtergrondstraling. Het transport van deeltjes is daarom *het* essentiële ingrediënt voor het begrijpen hoe het heelal van toestand verandert.



Figuur 13.1: Links: Een kaart van de temperatuur van de kosmische achtergrondstraling. De oersoep is glad, op enkele zeer minieme rimplingen na. Credit: NASA/WMAP Science Team. Rechts: De inhoud van het huidige heelal heeft een heel andere structuur, dat veel weg heeft van een bijenkorf of een spons.

Transport Theorie

De beweging van een deeltje kan gevolgd worden door het oplossen van zijn bewegingsvergelijking. Het resultaat is een geodetisch pad, langs welke het deeltje zal bewegen totdat een interactie (annihilatie, verstrooiing, dissociatie, etc.) plaats heeft. Helaas is het zo dat de meeste deeltjes zich in groepen voortbewegen. Een druppel water bestaat uit zeer veel afzonderlijke moleculen (in de orde van het getal van Avogadro, $\sim 10^{23}$), en een laserstraal bestaat meestal uit minstens zoveel fotonen. Dit zorgt dat het probleem enorm veel gecompliceerder wordt: we moeten niet alleen de bewegingsvergelijking van elk deeltje afzonderlijk oplossen, we moeten nu ook rekening houden met het feit dat de deeltjes onderling wisselwerking zullen hebben. De dynamische geschiedenis van elk afzonderlijk deeltje zal verwerkt worden in een complex geheel van verwoven paden, waardoor het bijna onmogelijk wordt om het gedrag van een groep deeltjes te begrijpen door het pad van elk afzonderlijk deeltje op te lossen, ook al is dat de meest precieze manier van aanpak.

Natuurkundigen, geconfronteerd met dit enorme probleem, hebben een gedeeltelijke oplossing gevonden: in plaats van het beschrijven van het gedrag van elk afzonderlijk deeltje, kan men ook als doel hebben om het collectieve gedrag van de groep als geheel te beschrijven, door uit te zoomen van het microscopische niveau naar een mesoscopische of macroscopische. De bewegingsvergelijkingen voor de individuele deeltjes zullen hierdoor vervangen worden door vergelijkingen voor gemiddelden over het systeem, grootheden die de fysische eigenschappen van de groep als geheel beschrijven. Deze vergelijkingen kunnen voorkomen in veel verschillende vormen, afhankelijk van het specifieke probleem. Het collectieve gedrag van gasdeeltjes bijvoorbeeld wordt beschreven door de Euler of Navier-Stokes vergelijkingen, en het gedrag van een groep fotonen die beweegt door de atmosfeer kan beschreven worden door de stralingstransport vergelijkingen.

De algemeen wiskundige context waarbinnen het gedrag van groepen deeltjes kan worden beschreven staat bekend onder de naam *statistische fysica*. De collectieve beweging van deeltjes, wellicht door een bepaald medium, kan worden beschreven door de enorm veelzijdige *transport theorie*. Beide dezer onderwerpen zijn hoekstenen van de moderne natuurkunde: ze zijn bijvoorbeeld gebruikt om het gedrag van gassen, vloeistoffen, plasma's en straling te beschrijven. Hun principes blijken veel universeler te zijn. Elk systeem dat ontleed kan worden als een groep van afzonderlijke elementen die wisselwerken via simpele regels kan beschreven en begrepen worden met hetzelfde gereedschap. Of het nu de beweging van verkeer op snelwegen is, de uitwisseling van geld op de beurs, de stroom van data op het internet, of zelfs de beweging van scholen vis, het zijn de principes van transport theorie die gebruikt kunnen worden om het specifieke proces te modelleren, of zelfs te begrijpen.

Stralingstransport

Een van de eerste transport problemen dat bestudeert is, is het transport van lichtdeeltjes door een medium, waarmee wisselwerking kan plaatsvinden. Dit zogenaamde stralingstransport is nog steeds een van de meest ingewikkelde en computationeel veeleisende problemen in de moderne natuurkunde. De getransporteerde eenheden zijn fotonen, deeltjes die bewegen met de maximaal mogelijk snelheid. Hierdoor bereiken ze vele delen van de ruimte, waarbij wisselwerking kan optreden, voordat ze op een bepaalde plaats aankomen. Dit maakt het probleem extreem niet-lokaal, waardoor zijn oplossing verre van triviaal is. De fotonen hebben een bepaalde plaats, reizen in een bepaalde richting, en hebben een bepaalde frequentie, en dit alles is tijdsafhankelijk. Dit maakt het probleem zeven dimensionaal: een enorme opgave.

De toepassingen zijn veelzijdig, variërend van laserfysica, tot het begrijpen van atoom-bom detonaties, tot zelfs het creëren van realistisch ogende 3D animaties. In de astrofysica is stralingstransport een van de meest essentiële ingrediënten voor het beschrijven van kosmologische processen. De vorming van kosmische objecten, zoals sterrenstelsels en sterren, wordt beïnvloed, soms zelfs gedomineerd, door stralingseffecten. Het is daarom noodzakelijk om dit specifieke transport theoretische proces mee te nemen in elk kosmologisch onderzoek.

Numerieke Methodes

Het is zeer moeilijk, soms zelfs onmogelijk, om de vergelijkingen die transport theoretische problemen beschrijven analytisch op te lossen. Gesloten oplossingen komen enkel voor in zeer specifieke, geometrisch en fysisch simpele, problemen. Het is daarom nodig om numerieke methodes te gebruiken om een aanwijzing te krijgen voor de echte oplossing, als die al bestaat. Het grootste deel van de bestaande numerieke methodes gaan uit van de macroscopische vergelijkingen die het transport probleem beschrijven, en proberen deze op te lossen door de vergelijking te discretiseren. In stralingstransport methodes, bijvoorbeeld, worden de eigenschappen van het medium, waardoor de fotonen zullen reizen, in vakjes opgedeeld. Door het gebruik van eindige differentiatie methodes worden oplossingen verkregen.

De meeste van deze methodes gebruiken een regelmatig rechthoekig computationeel grid als uitgangspunt voor deze discretisatie procedure. Deze keuze heeft niets te maken met het fysische probleem, maar is puur artificieel. Het blijkt ook dat, als je dit soort grid kiest,

er onwenselijke artifacten in de oplossing optreden: rotatie- en translatiesymmetrie wordt gebroken. Ook als de verdeling van het medium erg inhomogeen is, zullen regelmatige grids of te grof zijn om gebieden met veel detail op te lossen, of ze zullen veel te verfijnd zijn in bijna homogene gebieden. De methodes die van dit soort grids gebruik maken kunnen heel nauwkeurig zijn, maar in de meeste gevallen blijken ze zeer ingewikkeld en computationeel zeer veeleisen te zijn.

Een Nieuwe Methode

In dit proefschrift presenteren we een nieuwe methode die gebruikt kan worden om stralingstransport problemen op te lossen, maar ook andere processen waarin groepen deeltjes door een medium wordt getransporteerd. De methode verschilt radicaal van bestaande methodes in twee opzichten.

Ten eerste lost het niet de macroscopische, uitgezoomde, differentiaalvergelijkingen op die het dynamische gedrag van groepen deeltjes als geheel beschrijven. In plaats daarvan, gaat het terug naar het oorspronkelijke mesoscopisch perspectief van individuele deeltjes bewegend door het medium van een interactie naar de volgende.

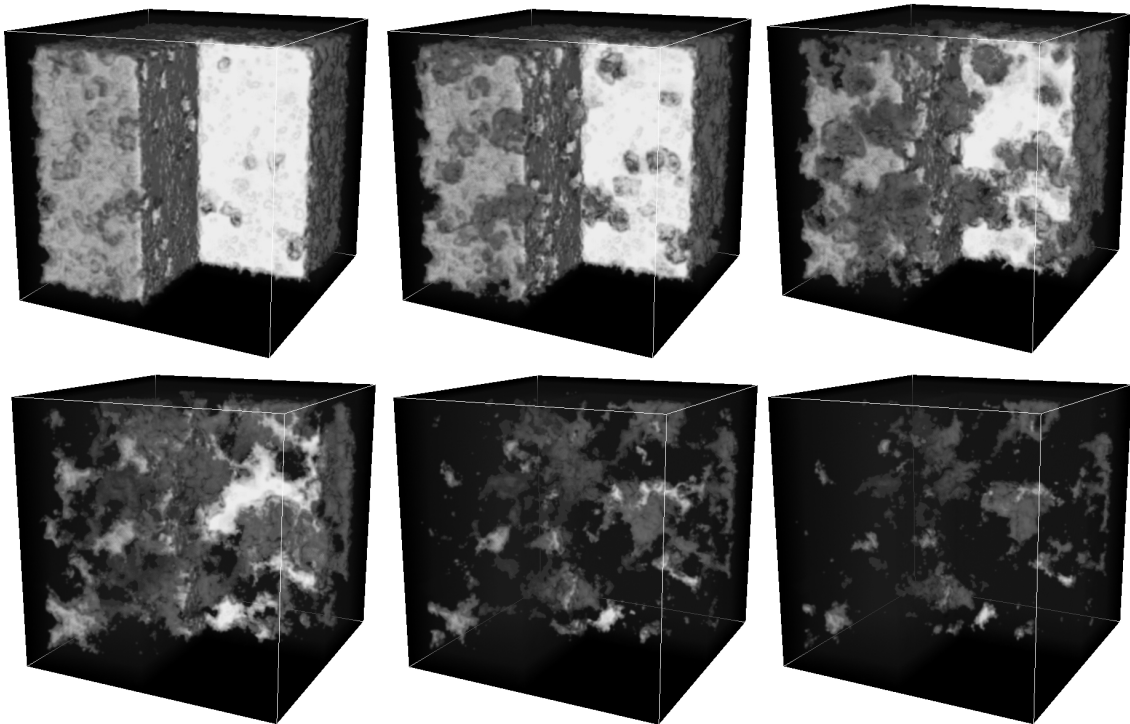
Ten tweede gebruikt het een computationeel grid dat drastisch anders is dan de gebruikelijke. Een punt proces wordt gebruikt om een verdeling van punten te genereren dat een directe representatie is van het medium. Gebaseerd op deze verdeling wordt een grid geconstrueerd door toepassing van een tessellatie procedure. Het resultaat is een graaf waarlangs transport kan plaatsvinden: een directe weerspiegeling van het werkelijke Markov proces van deeltjes die bewegen van een wisselwerking naar de volgende. We kunnen laten zien dat, door deze specifieke keuze van het opzetten van een grid, de lengtes van de lijnen van de graaf correleren met de lokale vrije weglengte van het medium. De keuze van de tessellatie procedure zorgt er ook nog voor dat zowel rotatie- als translatiesymmetrie wordt behouden.

Het resultaat is een methode dat gebruik maakt van een grid dat zeer fysisch is, en homogeen vanuit het perspectief van de deeltjes zelf: elke lijn, of die nou heel lang of kort is, is identiek voor het getransporteerde deeltje, namelijk een vast aantal vrije weglengtes. De methode heeft een groot bereik in resolutie vanwege zijn adaptieve eigenschappen, en het is ook vele malen sneller dan andere methodes. Die schalen met het aantal bronnen, en die van ons niet.

De methode combineert technieken uit verschillende takken van de wetenschap, en heeft, door de generieke opzet, vele toepassingen. We besteden daarom in dit proefschrift relatief veel tijd om deze zaken uiteen te zetten. Uit alle mogelijke toepassingen hebben we een specifieke gekozen: het tijdperk van kosmologische reionisatie.

Het Eerste Licht Trianguleren

Voornamelijk door het transport van deeltjes verandert het heelal van toestand. Uitgaand van een bijna homogene verdeling, voegen groepen baryonische en donkere materie deeltjes zich samen tot een sponsachtige grote schaal structuur, onder invloed van de zwaartekracht van de initiële fluctuaties. In de meest dichte gebieden van deze structuren kunnen de



Figuur 13.2: Visualisatie van een van onze reionisatie simulaties. Fotonen, uitgestraald door de eerste sterren, zorgen ervoor dat het neutraal waterstof (wit en ondoorzichtig) wordt geïoniseerd (zwart en doorzichtig). Naarmate de tijd verstrijkt, bedekken de individuele HII gebieden het volledige heelal: het heelal is transparant en het reionisatie tijdperk is beëindigd.

omstandigheden dusdanig zijn dat het aanwezig gas genoeg kan samentrekken en koelen om de vorming van de eerste sterren mogelijk te maken.

Deze bronnen leverden de eerste nieuwe voorraad fotonen, en hun impact op de materie in het heelal is verregaand. Omdat ze zeer zwaar zijn, vormen de eerste sterren fotonen die energetisch genoeg zijn om het waterstof, dat eerder was gevormd, opnieuw te ioniseren. Rondom elke ster worden gebieden geïoniseerd gas (HII gebieden) geblazen, die groeien in de tijd. Dit proces eindigt met het moment dat de HII gebieden overlappen, waardoor het heelal opnieuw transparant wordt voor ioniserende straling. Het proces begint aan het eind van de zogenaamde *Dark Ages*, en staat bekend als *het Tijdperk van Reionisatie*. Het is een van de hot topics in de moderne kosmologie.

Het is onmiddellijk duidelijk dat, als we dit tijdperk willen simuleren, transport methodes onontbeerlijk zijn. De collectieve beweging van gas en donkere materie deeltjes kan gesimuleerd worden met het gebruik van kosmologische hydrodynamica methodes, en het transport van ioniserende straling behoeft een stralingstransport methode. De hydrodynamica methodes zijn vrij efficiënt; de stralingstransport methodes vormen helaas het knelpunt in de huidige reionisatie simulaties: ze schalen met het aantal bronnen. Reionisatie wordt in gang gezet door zeer vele bronnen, die zeer inhomogeen zijn verdeeld. Bestaande transport methodes zijn hierdoor computationeel zeer duur.

Onze nieuwe methode schaalt niet met het aantal bronnen, en gebruikt een grid dat zich aanpast aan het medium. Met een implementatie van deze nieuwe methode zijn we daarom

in de fortuinlijke positie om systematisch verschillende aspecten van dit kosmologisch tijdperk te kunnen bestuderen. De resultaten van deze analyse worden beschreven in dit proefschrift. Een voorbeeld van een van onze reionisatie simulaties wordt gegeven in Fig. 13.2.

Overzicht

Dit proefschrift is opgedeeld in drie delen: definitie, implementatie, en toepassing.

Deel I

Het eerste deel geeft een algemene introductie van onze nieuwe transport methode, waarbij de nadruk wordt gelegd op zijn veelzijdigheid. We beginnen met een algemene introductie van transport theorie in Hoofdstuk 2, de nadruk leggend op verschillende aspecten die we nodig hebben om onze nieuwe methode uit te leggen. Hierna, in Hoofdstuk 3, bespreken we de eigenschappen van het adaptieve grid dat door onze methode wordt gebruikt. Als dat gedaan is, zijn we in de positie om onze nieuwe methode te beschrijven in Hoofdstuk 4. In Hoofdstuk 5 leggen we nadruk op de veelzijdigheid van de methode door in te gaan op zijn overeenkomsten met andere takken van de wetenschap, waarbij we mogelijke toepassingen aanstippen. We eindigen dit deel van het proefschrift met een kwantitatieve analyse van de nauwkeurigheid van deze adaptieve grids in Hoofdstuk 6.

Deel II

Het tweede deel van het proefschrift beschrijft een specifieke implementatie van de meer algemene methode, speciaal bedoeld voor kosmologisch stralingstransport. We geven allereerst een algemene introductie van de huidige status van kosmologisch stralingstransport in Hoofdstuk 7. Vervolgens beschrijven we in Hoofdstuk 8 hoe onze nieuwe methode werd geïmplementeerd als een C++ pakket met de naam `SimpleX`, dat specifiek werd ontwikkeld om gebruikt te worden in de context van kosmologisch stralingstransport. Dit pakket werd vergeleken met verscheidene andere methodes in een internationaal project, de resultaten waarvan gepresenteerd worden in Hoofdstuk 9.

Deel III

Het derde en laatste deel van dit proefschrift beschrijft de resultaten van het gebruik van de kosmologische stralingstransport implementatie, `SimpleX`, voor het doen van kosmologische reionisatie simulaties. Omdat het stralingstransport het knelpunt in reionisatie simulaties was, werden verscheidene fysische aspecten van het proces genegeerd, met als doel dat de simulaties werkbaar werden. We geven in Hoofdstuk 10 een beknopte introductie van reionisatie simulaties, waarbij we hun numerieke en fysische voorwaarden beschrijven, en we bespreken een aantal eigenschappen die meestal werden genegeerd. Een van deze aspecten is de invloed van het diffuse stralingsveld. In Hoofdstuk 11 laten we kwantitatief zien dat diffuse fotonen in veel gevallen het stralingsveld in HII gebieden domineren, zodat hun invloed zeker niet kan worden genegeerd. We eindigen dit proefschrift met resultaten van het gebruik van `SimpleX` voor het doen van veel verschillende reionisatie simulaties. In Hoofdstuk 12 laten we zien dat verscheidene numerieke en fysische aspecten een grote invloed hebben op het hele reionisatie proces, en dus meegenomen *moeten* worden.

Curriculum Vitae

Op 14 februari 1981 kwam ik in Maastricht via een keizersnede ter wereld, de navelstreng topologisch interessant om mijn nek gewikkeld. Het zuurstoftekort dat ik toen opliep is wellicht een gedeeltelijke verklaring voor het boek dat u op dit moment in handen heeft. Het ging echter helemaal mis toen mijn moeder me op mijn vijfde vertelde hoe ver het daarboven door ging in de ruimte. Ik geloofde er geen snars van, en was vastbesloten: ik ging sterrenkunde studeren. Na een luchtig intermezzo van voetballen en aan computers prutsen op de Montessori basisschool, en tafelfoetballen en poolen op het Sint Maartenscollege, beide te Maastricht, vertrok ik in 1998 naar Leiden om eens kwantitatief te checken wat er van waar was.

Het bleek al snel nodig de monotonie van tentamens scoren te doorbreken, in afwachting van zelf onderzoek doen. Enerzijds vond ik mijn heil in het recreatief tafeltennissen, het competitief poolen, en de horeca Leiden spekken, anderzijds kon ik mijn hart ophalen aan zeer uitdagende nevenwerkzaamheden. Bij Vodafone in Maastricht ontwikkelde ik applicaties voor het GSM-netwerk, bij ACI Worldwide te Gouda werkte ik als cryptograaf aan smartcard toepassingen, en bij het ingenieursbureau IKM, eveneens te Gouda, ontwikkelde en implementeerde ik mathematische modellen van een verscheidenheid aan zaken.

In 2001 begon ik aan onderzoek in de Theoriegroep aan de Sterrewacht Leiden. Ik ben er sindsdien blijven plakken. Onder leiding van prof. dr. Vincent Icke deed ik een klein onderzoek naar analytische en numerieke oplossingen voor potentiaalstroming in open sterclusters: een verkapt Landau en Lifshitz literatuuronderzoek. Hierna, in 2002, startte ik met mijn afstudeeronderzoek onder leiding van dr. Erik-Jan Rijkhorst en prof. dr. Vincent Icke. Het idee indertijd: "Wellicht kunnen we stralingstransport doen op Voronoi grids!". Dit prima uitgewerkt projectplan culmineerde uiteindelijk in een afstuderen in februari 2003, en in dit proefschrift 4 jaar later.

In die vier jaar promoveren bezocht ik conferenties in o.a. Chicago, Santa Barbara en Toronto, en gaf ik colloquia aan de meeste instituten in Nederland. Ik was assistent bij het Interacademiaal College 2005, en bij het college Sterren, en tegelijkertijd kon ik mijn coachingscapaciteiten bijschaven als afstudeerbegeleider van drs. Tim van Kempen en drs. Franco Maschietto. Ik kon mijn didactische ei kwijt in twee lezingen in het kader van de collegereeks "Het Levend Heelal", over informatietheorie en cryptografie enerzijds en over kansen in het leven anderzijds, en ik gaf in bijna alle uithoeken van Nederland publieke lezingen over de moderne numerieke astrofysica. In 2006 was ik medeorganisator van de Lorentz Center workshop "The World a Jigsaw: Tessellations in the Sciences", die Voronoi-experts uit alle takken van de wetenschap bij elkaar bracht.

Nu dit proefschrift er is, is het tijd om verder te gaan. Vanaf maart 2007 ga ik aan de slag als Quantitative Risk Analyst bij ABN Amro te Amsterdam. Zelfde wis- en natuurkunde, andere toepassing, en dresscode.

Nawoord

Er zijn velen die direct, maar vooral indirect hebben meegeholpen aan de totstandkoming van dit proefschrift. De vrije, niet hiërarchische structuur van de Theoriegroep is succesvol gebleken als een ideale broedplaats voor innovatieve ideeën, waarbij kruisbestuiving in velerlei opzicht tot baanbrekend werk leidde. Christianne, Erik-Jan, Franco, Huib, Inga, Inti, Jaron, Marijke, Nathan, Paul, Peter, Rowin, Siard, Sijme-Jan, Tim en Yvonne, het was een genoegen om met jullie van gedachten te hebben mogen wisselen. Jan-Pieter, ik zal met zeer veel interesse je werk blijven volgen; ik weet zeker dat er veel moois zal volgen. Erik-Jan, je was een ideaal afstudeerbegeleider: we wisten allebei niet goed waar we mee bezig waren.

Hoe meer andere faculteiten ik bezocht, hoe meer ik bewondering kreeg voor de informele sfeer op de Sterrewacht. Soms leek het wel alsof het echte werk in de Kaiser lounge werd gedaan. Ik hoop dat deze sfeer met de recente expansie behouden zal blijven. Een ander benijdenswaardig aspect van de Sterrewacht: Aart, David, Erik en Tycho, jullie kunsten ben ik bij nog geen enkel bedrijf tegengekomen. Tycho, ik denk dat jij onze gadget-wedstrijd hebt gewonnen. Ook ben ik het secretariaat zeer erkentelijk: Jeanne, Kirsten, en Liesbeth, bedankt voor al jullie hulp! Pierre-Marie and Paola, it was good to have you as roommates.

Het belangrijkste element voor een succesvol promotietraject is goede afleiding. Ik dank de horeca Leiden en Amsterdam voor al die keren dat ik me 's ochtends met schorre stem hardop afvroeg waar ik al dat geld had uitgegeven. Bas, bedankt dat je het vol hield om al die jaren te 9-ballen met een snookerspeler! De DK-club, Edo, Franco, Koen, Marlon, Remco, Reinier, Rowin, en Siard: het is immer een eer om in jullie gezelschap, met een Monte Christo tussen de lippen, een glas Moët et Chandon te heffen. Volgende keer in New York!

Rogier, jij hebt mijn respect afgedwongen met de manier hoe je omgaat met het feit dat een lichaam niet altijd in de pas loopt met de snelheid van de geest. Je bent een van de meest intelligente en geestige mensen die ik ken, en daarbij een enorm goede vriend. Ik hoop dat we in de toekomst nog veel mooie steden kunnen gaan bekijken.

Marlon, onze levenspaden kruisten in de Ardennen, meer dan 13 jaar geleden. Hoewel we niet altijd volledig in fase hebben gelopen, heeft onze vriendschap zich immer verdiept. Van achter op mijn rug de Kurfürstendamm afgaan tot bij FoodXL soulfood halen als het even tegenzit. Ik heb respect voor hoe je je de laatste jaren door alle tegenslagen heen hebt gebokst, en ben trots op een vriend zoals jij.

Lucy, wat heerlijk dat je nog net de laatste loodjes hebt meegepikt. Ik geniet er nu al van dat we de rest samen gaan invullen. Lach nog eens naar me, mooie vrouw...

Mam, het is een groot goed om iemand achter je te hebben die je immer en onvoorwaardelijk steunt. Je liefde, wijsheid en vechtlust hebben mij gebracht waar ik nu ben. Deze is voor jou.

