Radiation-coupled Primordial Chemistry in Simulations of The First Galaxies



Sean Barrett

St Edmund Hall

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2024

Thank you Mum, Dad, and Issy, for keeping me going.

Statement of Originality

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

Acknowledgements

Above all, I thank my supervisors, Julien Devriendt and Adrianne Slyz, for their patience, and determined support. I can never thank you both enough, and I have been supremely fortunate to be your student.

Harley Katz, you have my gratitude for hours upon hours of useful conversations, and your unfiltered advice.

My academic brothers, Sergio, Marius, and Curro, thank you for your camaraderie and assistance.

To my old friends, Jamie, Ollie, George, Divya, Dana, Lauren, Alex, Kate, Dan, James, and Jamie *(another one!)*, thank you for being wonderful. Without you all, I'd have nothing.

To my new friends, colleagues and mentors at First Light, who witnessed the final two years of this DPhil: Tom (and Sammi!), Sangi, Adam, Martin, Sam, Jumana, Nic, Mat, James, Jon, Dave, Nathan, and Nick, thank you for making me feel at home.

My family have provided their unconditional support and love for my entire life. Words will never be enough. Thank you, Mum, Dad, and Jason.

Issy, you have no idea how much you've helped me do this. Thank you for being the best friend anyone could ever ask for.

Abstract

In galaxies in the primordial universe, before any stars have formed, gas is almost entirely composed of neutral hydrogen and helium. As it accumulates into clouds under gravity, this gas necessarily heats up, and, in the absence of cooling processes, would eventually reach a stable state in which the cloud produces enough pressure to support itself. In order to ever collapse to form the first stars, this gas must radiatively cool.

In a purely atomic hydrogen and helium mixture, the cooling rate is strongly suppressed below temperatures of $\sim 10^4$ K, as the bulk of atom-atom collisions in the gas are not energetic enough to excite a hydrogen atom out of its ground state, and only the tail of the Maxwell-Boltzmann distribution contributes. However, the primordial hydrogen, and its isotope deuterium (present in small but significant amounts after The Big Bang), can form molecules H_2 and HD, which have vibrational and rotational modes that are able to be excited at much lower energies than the ground state of hydrogen. Thus, by forming these molecules, primordial gas is able to cool itself and collapse to form the first stars.

The chemistry involved in primordial H_2 and HD formation is rich, especially in the highly ionised, strongly irradiated, gas around the first stars. Modelling this chemistry, coupled to the radiation field in radiation-hydrodynamics simulations is expensive and challenging, requiring the advancement of the numerical solution of a sparse system of time-dependent ~ 50 ODEs in every mesh element for every time step. This computational cost is prohibitively large, so most studies neglect some or all of the time-dependence, and the effect of radiation, instead assuming collisional equilibrium concentrations of the reactants of H_2, or even H_2 itself, in order to calculate the molecular radiative cooling rate.

The validity of such approximations in the environment around the first stars has not been comprehensively proven. Indeed, one might expect that the gas around these stars is far from collisional equilibrium, as it is subject to

a strong ultraviolet radiation field. To investigate this, I have developed a fully non-equilibrium, radiation-coupled photochemical solver, coupled to the radiation-hydrodynamics code Ramses-rt (Rosdahl et al., 2013), that solves the photochemical system involved in the formation of H_2 and HD, so that their cooling rates may be calculated *self-consistently*.

I use my new code to demonstrate that molecular, star-forming gas around the first stars is far from collisional equilibrium (in both the concentration of H_2, and its reactants), employing a suite of high-resolution cosmological zoom simulations of two dwarf galaxies. I show that a full treatment of the primordial chemistry is necessary to properly model the formation of these stars, and the early evolution of their host objects — which are strongly influenced by the energetic supernovae and intense illumination they produce.

Contents

1 Introduction **1**

1.1 ΛCDM Cosmology . . . 3

1.1.1 CDM: Cold Dark Matter . . . 7

1.1.2 Λ: Dark Energy . . . 9

1.1.3 Inflation . . . 11

1.2 Structure Formation . . . 14

1.3 The First Stars and Galaxies . . . 22

1.3.1 Primordial Chemistry . . . 22

1.3.2 Population III Stars . . . 25

1.4 Reionisation . . . 27

1.5 Thesis Outline . . . 30

2 Numerical Methods **31**

2.1 RAMSES-RT . . . 31

2.1.1 Gravity . . . 32

2.1.1.1 Exact Particle-Particle Force Calculation . . . 33

2.1.1.2 Barnes-Hut Algorithm . . . 33

2.1.1.3 Particle Mesh . . . 35

2.1.1.4 P^3M and Tree-PM . . . 39

2.1.2 Hydrodynamics . . . 40

2.1.2.1 The Riemann Problem . . . 42

2.1.2.2 Numerical Stability . . . 46

2.1.2.3 Higher-Order Methods . . . 48

2.1.3 Adaptive Mesh Refinement . . . 51

2.1.4 Coordinates . . . 53

2.1.5 Alternative Hydrodynamics Schemes . . . 54

2.1.5.1 Smoothed-particle Hydrodynamics . . . 55

2.1.5.2 Moving Mesh Methods — Arepo . . . 60

2.1.5.3 Mesh-Free ALE — Gizmo . . . 64
2.1.6 Radiative Transfer . . . 68
2.1.6.1 Equations of Radiative Transfer . . . 68
2.1.6.2 Numerics . . . 72
2.1.6.3 Update Steps . . . 76
2.1.7 Sub-grid Physics . . . 77
2.1.7.1 Star Formation . . . 77
2.1.7.2 Radiative Feedback . . . 79
2.1.7.3 Supernovae . . . 80
2.2 Thermochemistry . . . 84
2.2.1 Thermal Evolution . . . 84
2.2.2 Cooling Processes . . . 85
2.2.3 Heating Processes . . . 88
2.2.4 Cosmological UV Background . . . 90
2.2.5 Equilibrium Thermochemistry . . . 91
2.2.6 Non-equilibrium H and He Photochemistry (Ramses-rt) . . . 91
2.2.7 Semi-equilibrium H_2 Photochemistry . . . 95
2.2.8 Fully Non-Equilibrium H_2 and HD Photochemistry . . . 98
2.3 Halo Finding . . . 106

3 Testing the Non-equilibrium Chemistry 109
3.1 Strömgren Spheres . . . 109
3.1.1 Tests in Static Atomic Hydrogen . . . 115
3.1.1.1 Test 1: Monochromatic Source in Static Gas at Constant Temperature . . . 115
3.1.1.2 Test 2: 10^5 K source with Thermal Evolution . . . 119
3.1.2 Tests in Non-Static Atomic Hydrogen . . . 120
3.1.2.1 Test 3: 10^5 K Source with Hydrodynamics . . . 123
3.1.3 Tests in Molecular Hydrogen . . . 127
3.1.3.1 H_2 Self-Shielding . . . 130
3.1.3.2 Test 4: 4.3×10^4 K Black-Body Source in Static H_2 at Constant Temperature . . . 133
3.1.3.3 Test 5: Comparison of Self-Shielding Prescriptions . . . 138
3.2 Spherical Cloud Collapse . . . 139
3.2.1 One-zone Model . . . 140
3.2.2 Comparison with 3D Simulations . . . 144

3.2.2.1 Forcing the Density Evolution 147

4 Non-equilibrium Primordial Chemistry in Simulations of Dwarf Galaxies in the Early Universe 151

4.1 Introduction 151
4.2 Methods 154
4.2.1 Cosmological Simulations 154
4.2.2 Halo Finding 158
4.3 Results 159
4.3.1 Properties of the Simulated Haloes 160
4.3.1.1 Final Thermochemical State 166
4.3.1.2 Stellar Mass – Halo Mass Relation 169
4.3.1.3 Metal Enrichment and Chemical Evolution 171
4.3.2 Star Formation and Supernova Feedback 175
4.3.3 Chaos and the Stochasticity of Star Formation 181
4.3.3.1 Comparing Simulations Under Different Physics Prescriptions 186
4.3.4 Non-Equilibrium H_2 Chemistry 188
4.4 Discussion 194
4.4.1 Comparison of Simulated Galaxies Under Different Physics Prescriptions 194
4.4.2 Comparing Simulations of a Chaotic System 196
4.4.3 Importance of Non-Equilibrium, Radiation-Coupled Primordial Chemistry 196

5 Conclusions 198

Bibliography 201

List of Figures

1.1 Energy content of the universe. 11

1.2 *Top*: Schematic showing the comoving Hubble radius vs. time. During the inflationary epoch, $(aH)^{-1}$ decreased at least 10^{28} fold, allowing the comoving ~ 1 m patch[a] that grew to become the observable universe to be in causal contact pre-inflation. *Bottom*: Diagram showing particles on a comoving grid in the universe before and after inflation, demonstrating that a decreasing comoving Hubble radius (circled) allows particles that were once in causal contact pre-inflation to no longer be able to communicate with each other, as they pass out of the Hubble radius during inflation. Blue ringed particles are the same pre- and post-inflation. Both panels are based on figures from Dodelson (2003). 13

1.3 Illustration of the tree-like structure of halo mergers. A resultant halo is formed by the merging of smaller ones, which tend to form earlier. The width of the 'branches' represent the halo masses. Time increases downwards, and t_0 marks the present time. This figure is adapted from (Lacey & Cole, 1993). 21

1.4 Radiative cooling rates per H nucleus vs. temperature for collisional ionisation and excitation (both from Cen 1992) in atomic hydrogen gas with an ionised fraction of 0.00031 (approximately the fraction after recombination) and proton density of $n_H = 1.0\,\mathrm{cm}^{-3}$, and per molecule for H_2 (Galli & Palla, 1998) and HD (Lipovka et al., 2005) in the optically thin limit. The steep cut-off of atomic cooling below $T \lesssim 10^4\,\mathrm{K}$ means that even small amounts of H_2 or HD can cool much more effectively than pure atomic gas at low temperature. 24

1.5 Evolution of temperature vs. number density for protostellar clouds at different metallicities, calculated with a one-zone model where chemicals are evolved according to a primordial network, and density evolves according to a free-fall law. Blue dashed lines mark the possible temperature and density of Jeans-stable clouds at the labelled masses. The red dashed line shows where the cloud becomes optically thick, at which point radiative cooling becomes inefficient. Lower metallicity clouds are able to remain hotter while they collapse, allowing clouds to remain Jeans-stable at higher masses, instead of fragmenting. This figure is from (Omukai et al., 2005). 25

1.6 Observer-frame spectra of 19 quasars, sorted by redshift. Quasars with $z > 6$ show strong Gunn-Peterson troughs below the Lyman-alpha wavelength (left of the peak), but as redshift decreases past 6, features — the Lyman-alpha forest — begin to appear in some of the spectra in this region, indicating that there is no longer a dense column of neutral hydrogen along the line of sight. This figure is from (Fan et al., 2006). 29

2.1 An example of a quadtree (2D equivalent to an octree) built around particles such that each exclusively occupies its own cell (Barnes & Hut, 1986). 34

2.2 *The Riemann 'Fan':* an example of a possible solution to a Riemann problem, exhibiting the key features. Lines indicate the spacetime trajectory of waves in the gas. *Left:* the *rarefaction wave* spreads out as its characteristics have increasing speed from right to left. *Middle:* the *contact discontinuity* (or entropy wave) separates the post-rarefication and post-shock gas. *Right:* the *shock* (or compression wave) is a sharp front, as the characteristics have decreasing speed from left to right, so they converge. . 44

2.3 Illustration of a graded Octree mesh (left), and the graph structure of the associated nodes (right), with the root node at the top (Wikipedia contributors, 2023). 52

2.4 A Rayleigh-Taylor instability between two fluids simulated on an ALE mesh (Hirt et al., 1974) (evolving from **left** to **right**). The rotational flow has tangled the mesh to the point of failure. 61

2.5 Arepo simulation of a Kelvin-Helmholtz instability (with the Voronoi mesh edges overlaid on the bottom half of each panel). Unlike with SPH, the instability is well-resolved here. 62

2.6 Partitioning of volume around 'particles' in GIZMO with a cubic spline partitioning function (*left*), Voronoi moving mesh methods (*middle*), and SPH (*right*). It is clear here that the kernel functions of SPH are not a true partitioning of volume, and the covering of the space is subject to numerical Monte Carlo noise. 65

2.7 Photon number density for two directional sources facing each other, simulated by RAMSES-RT (from Rosdahl et al. (2013)). The beams meet in the middle and unphysically collide, spreading out, with little directionality, from the centre. *Top:* using HLL to calculate inter-cell fluxes. *Bottom:* using GLF to calculate inter-cell fluxes; the beam diffuses much more strongly here than with HLL. 74

3.1 Equilibrium ionised gas fractions $\chi(r)$ (calculated by solving Equation 3.17 with the RK45 solver (a 4^{th} order explicit Runge-Kutta scheme for ODE integration) from the SCIPY package (Jones et al., 2001)) around a star with $\dot{N} = 5 \times 10^{48} s^{-1}$ in homogeneous neutral hydrogen with number density $n = 10^{-3}$ cm^{-3}, for three different ionisation cross sections, with the Strömgren radius $r_S = 5.4$ kpc marked by the vertical solid line. 3.007×10^{-18} cm^2 is the average cross section of photons emitted by a 10^5 K black-body between 13.6 eV and 24.59 eV (a close match to the situation in simulations shown later in this chapter); 6.3×10^{-18} cm^2 is the cross section of monochromatic 13.6 eV photons; 10^{-16}cm^2 is chosen arbitrarily to illustrate that the solution tends toward a step function (Strömgren's solution) as the cross section increases. The red dash-dotted line marks the position of the front as defined by $\chi(r_I) = 0.5$. 114

3.2 *Test 1:* Equilibrium ionised (dashed) and neutral (solid) spherically-averaged gas profiles produced by my modified version of RAMSES-RT (black), the public version of RAMSES-RT (red, labelled RT13), and the analytic prediction (blue). The Strömgren radius r_S is marked by the vertical blue dotted line. The numerical codes predict a slightly sharper front than the analytic result, but otherwise agree very closely with each other. 116

3.3 *Test 1:* Ionisation front radius r_I as a function of time, for my modified version of RAMSES-RT (black), the public version (red), and the prediction of Equation 3.19 (blue dashed). The analytic prediction lags behind the numerical result when the front is not sharp, if the front position r_I is defined such that $\chi(r_I) = 0.5$. If another ionisation threshold were chosen, r_I could be moved arbitrarily to any point on the extended front. The solution of Equation 3.19 converges to the Strömgren radius, marked by the horizontal blue dotted line. The slight difference between the two codes may be attributed to the difference in the numerical scheme used in the thermochemical step. Small differences in the ionisation profile in the front may lead to relatively large differences in the position r_I, since the front is steep, so the discrepancy observed here is not severe. 118

3.4 *Test 2:* Non-equilibrium ionised (dashed) and neutral (solid) spherically-averaged gas profiles produced (without hydrodynamics) by my modified code (black), and the public version of RAMSES-RT (red, labelled RT13), at 10 Myr (thinnest), 100 Myr and 500 Myr (thickest). The Strömgren radius r_S is marked by the vertical blue dotted line. The difference between my code and RAMSES-RT is barely visible in the ionised fraction, but appears to be more significant in the neutral fraction due to the use of a log scale; these differences are of the order $\sim 10^{-3}$, and are more strongly present in the rapidly ionising region, rather than the tail extending into the neutral gas. 120

3.5 *Test 3:* Non-equilibrium ionised (dashed) and neutral (solid) spherically-averaged gas profiles produced (with hydrodynamics) by my modified code (black), and the public version of RAMSES-RT (red, labelled RT13), at 10 Myr (thinnest), 100 Myr and 500 Myr (thickest). The Strömgren radius r_S is marked by the vertical blue dotted line. 124

3.6 *Test 3:* Radius of the ionisation front vs. time for my modified version of RAMSES-RT, and the public version. Analytic solutions are plotted in blue for the R-type and D-type regimes, with a horizontal dotted line marking the Strömgren radius, assuming a 10^5 K medium. 125

3.7 *Test 3:* 2D maps of f_H and ρ at $t = 500$ Myr. As predicted by the analysis in Section 3.1.2, the hydrodynamic shock (clearly visible in the ρ map, where gas has been swept up by the shock) has propagated ahead of the I-front. . 126

3.8 Schematic of dissociation front (outer border of the hatched region) and ionisation front (outer border of the dotted region) propagation around a star in an initially pure molecular hydrogen medium. The dotted, hatched and clear regions represent ionised, neutral atomic, and molecular hydrogen respectively. The Strömgren radius is labelled by r_S, and the equilibrium radius for the dissociation front is labelled by r_{H_2}. **Top:** when $t < t_{\mathrm{rec,\,H}}$, the dissociation and ionisation fronts have not yet passed the Strömgren radius, so neutral hydrogen (produced by the dissociation front) is rapidly ionised, only existing in a thin shell between the two fronts. **Bottom:** when $t > t_{\mathrm{rec,\,H}}$, the ionisation front does not progress beyond r_S, and the dissociation front carries on, leaving a large shell of neutral hydrogen behind it; the dissociation front eventually comes to rest at r_{H_2}. . 128

3.9 *Test 4:* 2D slices of mass fractions of ionised, neutral and molecular gas at 1500 Myr. Here $t_{\mathrm{rec,H}} < t < t_{\mathrm{rec,H}_2}$, so the system resembles the schematic in the bottom part of Figure 3.8, with the ionisation front in equilibrium at r_H, and the dissociation front still moving out to r_{H_2}. 135

3.10 *Test 4:* Ionised, neutral and molecular gas profiles at 1500 Myr. Since $t_{\mathrm{rec,H}} < t < t_{\mathrm{rec,H}_2}$, the ionisation front has reached equilibrium at the Strömgren radius (marked by the vertical blue dotted line), but the dissociation front is still moving out toward r_{H_2} (vertical red dotted line). The profile of $f_{\mathrm{H}_2}(r)$ rises sharply from $r = 0$ as the concentration of neutral hydrogen (from which H_2 can form) increases, but beyond $r = r_H$ the gradient is reduced as $f_H(r)$ is saturated (and thus so is the H_2 formation rate, which is $\propto f_H n_H^2$) and the equilibrium concentration of H_2 is determined almost entirely by the balance of the destruction by Lyman-Werner photons against the formation rate on dust. 136

3.11 *Test 4:* Positions of the ionisation (red) and dissociation (blue) fronts vs. time, and the analytic predictions (dashed). A horizontal red dotted line marks the Strömgren radius, and a horizontal blue dotted line marks r_{H_2}. The simulated ionisation and dissociation fronts do not converge exactly to r_H and r_{H_2}, since, in the system with molecular hydrogen, there are secondary reactions that contribute to the destruction of neutral hydrogen (particularly the formation of H_2 on dust), and the destruction of H_2 (particularly collisions with electrons and neutral hydrogen) that have not been accounted for in the calculation of the equilibrium radii (which balance only the primary formation channels against photodestruction). . . . 137

3.12 *Test 5:* Profiles of ionised, atomic and molecular gas for simulations with three different prescriptions for H_2 self-shielding. **Black:** No self-shielding, **Blue:** Gnedin et al. (2009) with $N_{H_2} = n_{H_2}\Delta x/5 \times 10^{14}\,\mathrm{cm}^{-2}$, **Red:** Nickerson et al. (2018) with $S = 400$. 139
3.13 **Top:** Temperature vs. number density (of hydrogen nuclei) for the one-zone model of Omukai (2000) at various metallicities (in solar units). The straight dashed line marks the threshold beyond which the cloud becomes opaque to its thermal continuum radiation emission. Dotted lines show the dust grain temperatures at each metallicity. **Bottom:** Temperature vs number density at various metallicities for the one-zone model with my network and cooling suite. 143
3.14 Temperature against density of the one-zone model with my network and a free-fall density evolution (dashed), and the core of the 3D cloud (solid), for various metallicities (in solar units). 145
3.15 Core density evolution for the 3D cloud (coloured solid lines) for various metallicities (in solar units), the one-zone free-fall model (black solid line), and the version of RAMSES-RT of Kimm et al. (2017) (dashed) with $Z = 0$. The free fall time is marked by the vertical dotted line. 146
3.16 Temperature vs. density of the core of the 3D cloud (solid) and the one-zone model with a density evolution consistent with the 3D simulation (dashed), for various metallicities (in solar units). 148
3.17 Mass fraction of molecular hydrogen vs. density of the core of the 3D cloud (solid) and the one-zone model with a density evolution consistent with the 3D simulation (dashed), for various metallicities (in solar units). . 149

4.1 Maps of the average gas density along the line of sight (weighted by density) for each of the physics prescriptions (excluding III), for each set of ICs (Halo 1 and Halo 2). The 8 most massive subhaloes are marked by red circles, with the size of the circle (arbitrarily) proportional to the virial radius. Clusters of stars are marked by white symbols at their centre of mass, with adjacent numbers indicating the count of star particles (where solitary particles are not marked with a count). 162
4.2 Maps of the average gas temperature along the line of sight (weighted by density) for each of the physics prescriptions (excluding III), for each set of ICs (Halo 1 and Halo 2). 163

4.3 Maps of the average hydrogen ionisation fraction along the line of sight (weighted by density) for each of the physics prescriptions (excluding III), for each set of ICs (Halo 1 and Halo 2). 164
4.4 Maps of the average molecular hydrogen mass fraction along the line of sight (weighted by density) for each of the physics prescriptions (excluding III and V), for each set of ICs (Halo 1 and Halo 2). 165
4.5 Gas phase histograms in (ρ, T) space for each of the physics prescriptions (excluding III), for each set of ICs (Halo 1 and Halo 2). 167
4.6 Gas phase histograms in (ρ, x_{H_2}) space (where x_{H_2} is the gas mass fraction of H_2) for each of the physics prescriptions (excluding III), for each set of ICs (Halo 1 and Halo 2). 168
4.7 Maps of the average HD mass fraction along the line of sight (weighted by density) for Prescription I, for each set of ICs (Halo 1 and Halo 2). . . . 169
4.8 Gas phase histograms in (ρ, x_{HD}) space (where x_{H_2} is the gas mass fraction of HD) for Prescription I, for each set of ICs (Halo 1 and Halo 2). 169
4.9 Stellar vs. halo masses of subhaloes that contain stars (plotted over many snapshots) in each physics suite for Halo 1 (**left**) and Halo 2 (**right**). Each subhalo is identified (consistently between runs, across all figures) with a different marker style on a separate line. The fit of Behroozi et al. (2019) to the observed Stellar Mass – Halo Mass relation at $z = 0$ is shown by the dashed line. 170
4.10 Mass-weighted average hydrogen ionisation of the simulation zoom region with each physics suite for Halo 1 (**left**) and Halo 2 (**right**). As the NewHorizon prescription does not track non-equilibrium ionisation, this is calculated after the fact by assuming collisional ionisation equilibrium (equal recombination and collisional ionisation rates) given the gas temperature in the snapshot. 172
4.11 **Top:** Total mass of metals in the zoom region for Halo 1 (**left**) and Halo 2 (**right**) for each of the physics prescriptions. **Bottom:** Mass-weighted average IGM metallicity for Halo 1 (**left**) and Halo 2 (**right**). Here, the IGM is taken to be the volume inside the zoom region, excluding the spheres within the virial radii of each of the 8 most massive progenitor subhaloes. 173

4.12 **Top:** Total metal mass in subhaloes in Halo 1 (**left**) and Halo 2 (**right**) for each of the physics prescriptions. **Bottom:** Median mass-weighted subhalo metallicity for Halo 1 (**left**) and Halo 2 (**right**). Of the 8 most massive subhaloes, only those with stars are included in the average. Shaded regions are bounded by the upper and lower quartiles. 174
4.13 H_2 mass fractions in the 8 most massive subhaloes for each physics prescription with non-equilibrium H_2 formation, in Halo 1 (**left**) and Halo 2 (**right**). Shaded regions are bounded by the upper and lower quartiles. . . 175
4.14 Mass of population III (**top**) and II (**bottom**) stars formed per year in each physics suite for Halo 1 (**left**) and Halo 2 (**right**). 176
4.15 Initial masses of all population III (**top**) and II (**bottom**) star particles formed in each physics suite for Halo 1 (**left**) and Halo 2 (**right**). Pop. III particles represent individual stars, whereas Pop. II particles represent a population of equivalent mass, so the bottom panel does not represent a physical IMF, but rather the initial representative masses of populations. The initial mass of a Pop. II particle is an integer multiple of the minimum mass, where this integer is sampled from a Poisson distribution such that the expected value of the particle mass is consistent with the mass given by the Schmidt law in the star formation model (Section 2.1.7.1). 178
4.16 Probability density function of metallicities of stars (**left**) and counts of star particles in each metallicity bin (**right**; as in the left panel, but no longer dividing by bin width) formed in each physics suite for Halo 1 (**top**) and Halo 2 (**bottom**). In Prescriptions that model the formation of Pop. III stars (I and II), these are contained in the leftmost bin. 179
4.17 Histograms showing the probability density for the comoving separation r of pairs of Pop. II (**bottom**) and Pop. III stars (**top**) in Haloes 1 (**left**) and 2 (**right**) in the latest snapshots for each run. 180
4.18 Probability density function of gas density at supernovae sites (**left**) and star formation sites (**right**) in each physics suite for Halo 1 (**top**) and Halo 2 (**bottom**). 181
4.19 Maps of the average gas density, temperature, hydrogen ionisation fraction, and molecular gas mass fraction, along the line of sight (weighted by density) for each of the additional runs with Prescription II of Halo 1 at $z = 9.87$. Clusters of stars are marked by white symbols at their centre of mass, with adjacent numbers indicating the count of star particles (where solitary particles are not marked with a count). 183

4.20 Mass of stars formed per year in multiple simulations of Halo 1 with Prescription II. **Left:** simulations with identically-seeded PRNGs, with differences occurring only due to accumulated non-deterministic rounding-errors in parallel reduction operations. **Right:** simulations with differently-seeded PRNGs. 185
4.21 Histogram of hydrogen ionisation fraction of gas in cells in subhaloes of simulations with Prescription I just after the formation of the first star, with the measured value on the left axis, and the collisional equilibrium value on the bottom axis. The dashed line passes through the points where the measured values would be equal to the collisional equilibrium values. . 189
4.22 Histogram of the mass fraction of H^- (**top**), H_2^+ (**middle**) and H_2 (**bottom**) of gas in cells in subhaloes of simulations with Prescription I just after the formation of the first star, with the measured value on the left axis, and the collisional equilibrium value on the bottom axis. The dashed line passes through the points where the measured values would be equal to the collisional equilibrium values. 191
4.23 Histogram of the mass fraction of H^- (**top**), H_2^+ (**middle**), and H_2 (**bottom**) of gas in cells in subhaloes of simulations with Prescription I, with the measured value on the left axis, and the collisional equilibrium value, with the measured value of the free electron number density substituted into all other reaction rates, on the bottom axis. The dashed line passes through the points where the measured values would be equal to the collisional equilibrium values. 192
4.24 Histogram of the mass fraction of H_2 of gas in cells in subhaloes of simulations with Prescription I, with the measured value on the left axis, and the collisional equilibrium value, with the measured value of the concentration of H^-, H_2^+ and free electrons substituted into the H_2 formation and destruction rates, on the bottom axis.The dashed line passes through the points where the measured values would be equal to the collisional equilibrium values. 193

Chapter 1

Introduction

In 1924, Edwin Hubble published[1] his finding that the Andromeda 'Nebula' was, in fact, not inside the Milky Way but a whole galaxy of its own, using Cepheid variable stars as standard candles to measure its distance from Earth. It was found to be hundreds of kiloparsecs away — far beyond the then-overestimated radius of the Milky Way of ~40 kpc. These observations were made with the 2.5 m Hooker Telescope, which Hubble then used to show that not only did many other galaxies exist, but that they are mostly moving away from the Milky Way, at speeds (and thus redshifts) proportional to their distances — this is *Hubble's Law* (Hubble, 1929a), which says that *the universe is expanding*.

Since then, increasingly larger and more sophisticated telescopes have been able to observe many more galaxies in much greater detail across the electromagnetic spectrum (seeing galactic gas emit from radio to X-ray rest-frame energies), and look further into the cosmic past. The 2.4 m Hubble Space Telescope (HST) has been used to confirm the detection of a galaxy at redshift $z = 11.1$ (Oesch et al. 2016; later found to be closer, at $z = 10.6$ — Bunker et al. 2023), identifying the Lyman-break in its spectrum — redshifted from the ultraviolet into the near-infrared. This was the most distant confirmation to date before the launch of the 6.5 m James Webb Space Telescope (JWST) at the end of 2021. JWST is capable of identifying many more high redshift galaxies than the HST, with a larger mirror (affording greater sensitivity), a wider field of view, the ability to capture light further into the infrared, and more spectral resolution (with the NIRSpec instrument). Indeed, the JADES survey (the JWST Advanced Deep Extragalactic Survey;

[1] First in *The New York Times*, but eventually also in the peer-reviewed *Astrophysical Journal* (Hubble, 1929b)

Eisenstein et al. 2023) that is being carried out on JWST, by the end of its observation period, was predicted to have discovered thousands of galaxies with $z > 6$, tens with $z > 10$, with the possibility of identifying ~ 1 in this field alone with $z > 13$ (Steinhardt et al., 2021; Williams et al., 2018), enabling better measurements of the evolution of the high-redshift UV luminosity function (which informs, for example, which sources are responsible for reionising the universe). JADES is responsible for the most distant galaxy detection to date, at $z = 14.32$ (Carniani et al., 2024).

Despite the wealth of observations of galaxies (with many more already being provided by JWST) back into the first couple of billion years of the universe, there remain fundamental limitations of what can be discovered by observation alone. Macroscopic properties of galaxies (their mass, luminosity, stellar content etc.) evolve on timescales of at least tens of millions of years, so the evolution of a single galaxy cannot be tracked within even thousands of consecutive human lifetimes. Additionally, galaxies in the earliest stages of formation are extremely difficult to observe, as a result of their high redshift (which reduces their surface brightness proportionally to $(1+z)^{-4}$), so increasingly larger telescopes see diminishing returns in improving the redshift at which they can obtain spatially resolved images (at a given magnitude).

This leaves theoretical galaxy evolution with the problem of having to explain a population of observed galaxies without being able to directly see the time evolution of any single one. Therefore, the success of a 'complete' model of galaxy evolution must be judged by whether it can predict results that are statistically consistent with the observed population across all redshifts. Developing such a model is a difficult problem, as the physical processes involved (including gravity, magnetohydrodynamics, radiative transfer, stellar evolution and chemistry) couple together in a complex, highly-nonlinear system that exhibits behaviour over a large range of distance scales, from the size of stars to the largest clusters of galaxies. Such a system can only be solved computationally, but this is not without difficulties. One of the greatest challenges is in capturing the dynamic range: simultaneously resolving gas dynamics on the scale of protostellar clouds and the scale of clusters of galaxies would be extremely computationally demanding (far beyond the

memory capacity and compute performance of even the largest modern supercomputers). Therefore, limited by available computing power, cosmological simulations of galaxies must each be more narrow in scope, focusing on, for example, trying to predict the key observable properties of a high number of galaxies in a very large simulated volume (e.g. Horizon-AGN (Dubois et al., 2014), IllustrisTNG (Nelson et al., 2019), Simba (Davé et al., 2019), EAGLE (Schaye et al., 2015)), simulating a smaller volume of objects but with higher resolution (e.g. SPHINX (Rosdahl et al., 2018), FIREbox (Feldmann et al., 2022)), or investigating single objects in great detail (e.g. Kimm et al. 2017, Hopkins et al. 2018, Latif et al. 2022, Katz et al. 2023).

This chapter contains a summary of the key ingredients of theoretical galaxy evolution, focusing on the formation of the very first objects in the Dark Ages of the universe. A description of the numerical methods used to combine and solve these models follows in Chapter 2.

1.1 ΛCDM Cosmology

Cosmology is concerned with the universe (and its entire history) on very large scales. It describes the spacetime background in which all matter exists, and the original state of that matter, so it is the foundation of galaxy evolution. Modern models of the universe are very successful and highly constrained by precise observations of the cosmic microwave background (CMB) (Planck Collaboration et al., 2020) and surveys of supernovae, galaxies and gravitational lensing (Abbott et al., 2022a; Heymans et al., 2021). Thus, competing models tend to differ only in the finer details where observations are less constraining, e.g. the fundamental nature of dark matter. The simplest of these that succeeds in describing the observed cosmos is called the Λ Cold Dark Matter (ΛCDM) Model, which is the basis for modern galaxy formation theory.

In ΛCDM, matter in the universe is made up of two components: baryons — the familiar matter we can see — and *dark matter* — which interacts very weakly with baryonic matter and radiation. This dark matter has been hypothesised to exist, though never directly detected, in order to explain excess non-radiatively-interacting mass implied by galaxy

cluster kinematics, galaxy rotation curves, and the CMB power spectrum (discussed further in Section 1.1.1). It is assumed to be 'cold' (as 'hot' dark matter would reduce power at large scales in the galaxy mass spectrum in a manner inconsistent with observations), meaning that it has sub-relativistic speed, but its fundamental composition is not well-constrained. If CDM is made up of thermally-produced weakly interacting massive particles (WIMPs), then its low speed is a result of having a relatively high mass. However, it is also hypothesised that CDM could be made up of macroscopic, non-luminous objects (so-called massive astrophysical compact halo objects — MACHOS) such as primordial black holes, which may contribute to either some or all of CDM, especially if they were able to form with masses in the 'asteroid range' of $10^{17}\,\mathrm{g} < M < 10^{23}\,\mathrm{g}$ (Carr et al., 2021).

The Λ in ΛCDM refers to the *cosmological constant*: a coefficient in Einstein's field equations (thus present in Equation 1.1) that causes the accelerating expansion of the universe. Einstein originally believed that Λ was necessary to allow the universe to remain static, but since Hubble's discovery that galaxies were receding with a velocity proportional to their distance (Hubble, 1929a), it has been known that the universe is not static, but *expanding*, so Λ was dismissed and considered to be zero. However, it was later discovered that this expansion is accelerating, via distant observations of type 1a supernovae (discussed further in Section 1.1.2), that required the revival of Λ.

The evolution of the spacetime of the universe, assuming homogeneity and isotropy (i.e. taking the axiomatic *cosmological principle*), can be described by the Friedmann equations (using $c = 1$ throughout; derived from Einstein's field equations; Misner et al., 1973)

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3}, \tag{1.1}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}, \tag{1.2}$$

where a is the *scale factor* that quantifies the expansion of space — it is the ratio of physical distances to comoving (coordinate) distances, that do not change as the universe grows — ρ is the energy density of matter and radiation, p is the pressure, Λ is the aforementioned cosmological constant, and $k = -1, 0, 1$ is the *curvature parameter* that describes

how the intrinsic geometry of the universe (its 'shape') affects its evolution. $k = -1, 0, 1$ correspond to *open*, *flat* and *closed* universes respectively, and though the most recent CMB observations are consistent with a flat universe (Planck Collaboration et al., 2020), flatness relies on the total energy in the universe summing to be exactly equal to a critical value. Thus, it is impossible to rule out either positive or negative curvature without either compelling theoretical reasons to prefer a flat universe (e.g. *Inflation*, proposed by Guth (1981), discussed later in Section 1.1.3, which is a theoretical solution to this apparent fine-tuning problem of flatness, amongst other problems), or more informative observations that show a preference toward one side or the other of the critical value.

If we assume that the universe is flat ($k = 0$) and that $\Lambda = 0$ we recover the definition of the *critical density*:

$$\rho_c(t) = \frac{3H^2(t)}{8\pi G}, \tag{1.3}$$

where we have defined the *Hubble parameter* $H(t) \equiv \frac{\dot{a}}{a}$, and ρ_c is the density, in a universe of pure matter and radiation, that leads to a flat geometry at time t. It is convenient to define ratios of other quantities to ρ_c, called *density parameters*:

$$\Omega_m \equiv \frac{\rho_m}{\rho_c}; \quad \Omega_r \equiv \frac{\rho_r}{\rho_c}; \quad \Omega_k \equiv \frac{k}{a^2H^2} \equiv \frac{\rho_k}{\rho_c}; \quad \Omega_\Lambda \equiv \frac{\Lambda}{3H^2} \equiv \frac{\rho_\Lambda}{\rho_c}, \tag{1.4}$$

where ρ_m is the matter density, ρ_r is the radiation density, and we also define effective energy densities ρ_k and ρ_Λ. This allows us to write the first Friedmann equation in a more elegant form

$$1 - \Omega_k = \Omega_m + \Omega_r + \Omega_\Lambda, \tag{1.5}$$

in which it can be seen that $\Omega_m + \Omega_r + \Omega_\Lambda = 1$ in a flat universe, and the dependence of the curvature on the content of the universe is explicit.

By assuming the universe is composed of a multi-component zero-viscosity fluid (matter, radiation and dark energy — with each component having a different pressure-density relationship) and considering its energy conservation (see e.g. Dodelson, 2003), we can separate the effect of matter and radiation (since radiation applies a pressure, whereas

matter does not at cosmological densities), and find the dependence of each term in Equation 1.5 on the scale factor a. However, I present a simpler, informal argument here. Intuitively, we expect the density of matter to decrease in an inverse cubic manner as the linear scale of the universe increases, i.e. $\Omega_m \propto 1/a^3$, as the total mass is conserved. Similarly, the photon number density will scale the same way, however, as the photon wavelength λ is stretched proportionally to a, and the photon energy $E \propto 1/\lambda$, we expect the radiation density parameter to obey $\Omega_r \propto 1/a^4$. Thus, without formal calculation, we have

$$\Omega_m = \frac{\Omega_{m,0}}{a^3}; \quad \Omega_r = \frac{\Omega_{r,0}}{a^4}, \tag{1.6}$$

where the subscripts 0 indicate that the quantity is taken to have its value at the present day t_0, and the definition $a(t_0) \equiv 1$ is applied. k and Λ are constant, so the a dependence of their associated density parameters is clear from their definitions. Then we can write the first Friedmann equation as

$$H^2 = H_0^2 \left(\frac{\Omega_{k,0}}{a^2} + \frac{\Omega_{m,0}}{a^3} + \frac{\Omega_{r,0}}{a^4} + \Omega_\Lambda \right). \tag{1.7}$$

Already, we can begin to construct a qualitative history of the universe without explicitly solving this equation. As the radiation energy density has the strongest inverse dependence on a, it dominates the total energy up to the time when $\Omega_r = \Omega_m$ which is at $a_{eq} = \Omega_{r,0}/\Omega_{m,0}$, corresponding to a *redshift* of $z_{eq} = 3387$ (according to the latest CMB measurements from Planck Collaboration et al., 2020) — referring to how much the wavelength of light is increased by cosmological expansion when emitted from a certain time in the past and observed today, defined as

$$z \equiv \frac{\lambda_{\text{obs}} - \lambda_{\text{em}}}{\lambda_{\text{em}}} = \frac{1}{a} - 1. \tag{1.8}$$

In the radiation-dominated epoch, the scale factor evolves as $a \propto t^{\frac{1}{2}}$. Next comes the epoch of matter dominance, which lasts until $a_{de} = \sqrt[3]{\Omega_{m,0}/\Omega_\Lambda}$ (a redshift of $z_{de} = 0.28$), in which the scale factor goes as $a \propto t^{\frac{2}{3}}$. Beyond this, the cosmological constant dominates

the effective energy density, and the scale factor increases exponentially with time.

1.1.1 CDM: Cold Dark Matter

Despite never being detected directly, the evidence for the existence of dark matter is multifold and difficult to explain otherwise. Though alternative models exist, dark matter may be considered the simplest that is compatible with the evidence.

The first of this evidence is that the velocities of galaxies in clusters imply a much greater gravitational mass than can be explained by the stellar and gas content. This was first noted by Zwicky (1933) in observations of the Coma Cluster, whose conclusion is held up by modern work (e.g. in the weak-lensing analysis by Okabe et al., 2010b), despite the somewhat inaccurate initial cluster mass estimation from its luminosity. Both strong (Okabe et al., 2010a) and weak (Richard et al., 2010) gravitational lensing methods find that there is a large excess of gravitational matter compared to the luminous matter in other local clusters too.

Second, observations of the Bullet Cluster show gas separate from the stellar and non-luminous content inferred by weak lensing maps. The Bullet Cluster contains two sub-clusters of galaxies that have just collided. The stars in the galaxies within the sub-clusters are effectively non-collisional, and so they passed through each other, remaining in separate clumps, but the hot gas components of the sub-clusters collided, causing them to lag behind their original hosts in the centre between the two sub-clusters. The X-ray emission from this hot gas was observed by Chandra (Markevitch et al., 2002), being found to make up ~85% of the visible baryonic mass of the cluster. However, weak lensing analysis (Clowe et al., 2004) shows that most of the cluster mass is within the galaxy sub-clusters, rather than the hot X-ray halo, implying that the non-luminous component is also weakly-collisional. Analysis of the Bullet Cluster by Markevitch et al. (2004) places an upper bound on the dark matter collisional cross section to mass ratio σ/m of $1\,\mathrm{cm^2 g^{-1}}$; the equivalent quantity for atomic hydrogen gas is approximately $\pi(2a_0)^2/m_H = 2.1 \times 10^8\,\mathrm{cm^2 g^{-1}}$ (where the radius of hydrogen is taken to be the Bohr radius $a_0 = 5.291 \times 10^{-9}$ cm), so dark matter self-interacts many orders of magnitude less strongly than gas.

Third, the rotation curves in spiral galaxies exhibit a similar need for mass in addition to what is accounted for by their visible content. Spectroscopic observations (e.g Rubin et al., 1980; de Blok et al., 2008) find that stars in spiral galaxies continue to rotate quickly around the centre as the radius of their orbit increases, that is, the rotational velocity curves of spirals are flat, rather than the Keplerian prediction of a $1/\sqrt{r}$ falloff outside the disc. This indicates that not only is there more mass than is accounted for by the luminous matter, but also that it is distributed in a more extended way than the visible galactic gas and stars. Furthermore, Ostriker & Peebles (1973) used N-body simulations to show that spherically symmetric haloes of undetected matter, exceeding the *mass and radius* of the stellar disc, were necessary to stabilise observed structures in galaxies.

Perhaps the most compelling evidence for the existence of dark matter is found in the CMB, as it rules out simple modifications to gravity that could otherwise explain galaxy and cluster dynamics without dark matter (e.g. Modified Newtonian Dynamics — MOND; Milgrom, 1983). In the very early universe, hot, ionised baryonic matter was strongly coupled with radiation, allowing pressure waves to propagate very quickly. These waves originated from the collapse of matter into overdense regions, where baryons were able to 'bounce' back to an extent (due to an effective pressure from its radiation-coupling), but the dark matter, which does not exert any pressure, simply accrues in the overdensity. Dark matter thus acts to suppress the expanding phase of oscillations (waves travelling away from overdensities) and enhance compression. These so-called *acoustic* oscillations manifest in the CMB as anisotropies, where the angular scale of peaks in the power spectrum are determined by the series of wavenumbers corresponding to extrema of the oscillations: odd modes are maxima (compressive), even are minima (expansive). So, odd modes consist of one more compression than there are expansions, and thus are enhanced relative to even modes (which compress then expand the same number of times) by the presence of dark matter. This leads to the ratio of the odd and even peaks being highly sensitive to the baryon-dark matter ratio, and indeed this signal has been detected with high precision in modern CMB observations (Planck Collaboration et al., 2020), finding $\Omega_m = 0.31$ and the baryonic density parameter $\Omega_b = 0.049$ both with percent-level accuracy, meaning *dark matter makes up 85% of all matter in the universe.*

Although little is known about the nature of dark matter, constraints can be put on its kinetic state. Dark matter particles are unable to coalesce into structure on scales below their free-streaming length (which increases with speed). Thus, if dark matter is 'hot', there would be a very strong damping of high wavenumbers in the matter power spectrum. This suppression is not observed in the universe, so hot (relativistic) dark matter can be ruled out. Intermediate mass 'warm' dark matter is not ruled out, with a rest-frame mass-energy above 2.96 keV allowed by SDSS measurements of the *Lyman-alpha forest* (Baur et al., 2016), assuming the particles are thermally-produced WIMPs. The Lyman-alpha forest refers to the absorption features in the spectra of distant quasars, caused by neutral hydrogen in galaxies along the line of sight, where the absorbers are at different redshifts, causing the Lyman-alpha line to occur at different observed wavelengths, leaving a so-called 'forest' of absorption lines in the spectrum. This can be used to probe the density of structure on intermediate scales, and thus constrain the free-streaming length of dark matter particles. Warm dark matter models have more subtle effects on structure formation (compared to hot vs. cold dark matter), and are not considered for the purposes of this thesis.

1.1.2 Λ: Dark Energy

The interpretation of the cosmological constant Λ has a rich history. Einstein first conceived of it as a mathematically-allowed term in his field equations that enabled a steady-state universe, then going on to describe this as his greatest "blunder" after the Big-Bang model of the universe came to be accepted, eliminating the 'need' for Λ for decades. However, once it was possible to measure the distance to galaxies at higher redshift by using type 1a supernovae as standardisable candles, it was discovered that the expansion of the universe has accelerated (Riess et al., 1998), which is only possible if $\Lambda > 0$ (when total matter and radiation are conserved). These observations were incompatible with $\Lambda \leq 0$ at the 3σ level, and ruled out a flat universe of pure matter ($\Omega_m = 1$) with 7σ significance. Λ behaves as another part of the multi-component fluid of the universe, with the strange property of having *negative* pressure in proportion to its density.

In the interpretation that Λ represents the effect of some form of exotic matter or

field (i.e. a component of the energy-momentum tensor, rather than a term in the fundamental laws of spacetime), it is called *dark energy*, analogously to dark matter. It has been thought that the quantum vacuum energy may contribute to, or be the source of, this dark energy, but the prediction from the standard model is that the vacuum energy exceeds that required by the cosmological constant by *hundreds of orders of magnitude*! Indeed, by the standards of particle physics, the energy density of dark energy is minuscule, but unchanging as the universe expands, and demonstrates a need for drastically new theoretical models if its origin is to be explained.

In modern cosmology, various techniques enable the precise measurement of Ω_Λ. The previously mentioned acoustic oscillations in the early universe cause anisotropies in the CMB temperature, but also in the primordial density field, that imprint a preferred scale for galaxy separation that acts as a statistical standard ruler. That is, the two-point correlation function between the positions of galaxies on the sky has a peak at the comoving scale of acoustic oscillations, as a function of redshift to take into account the expansion of the universe. This means observations of many galaxies grouped by redshift (e.g. Abbott et al., 2022b) allow measurement of the evolution of the Hubble parameter over time, which is a strong function of Ω_Λ at redshifts below 1. Combining these *baryon acoustic oscillation* (BAO) measurements with supernova distances (Abbott et al., 2022a) allows a precise, CMB-independent constraint on Ω_Λ.

Observations of the CMB also depend on the value of the cosmological constant. Although Ω_Λ is not dominant in the early universe, the CMB must pass through the more local universe to reach us, and can be affected in a number of ways (Planck Collaboration et al., 2016), as it is subject to both accelerating expansion of the universe and interactions with structure along the line of sight. Changes in the matter density between the last scattering surface (where the CMB is emitted from) and Earth cause the CMB to be gravitationally redshifted, in what is called the *integrated Sachs-Wolfe effect*. The distance to the last scattering surface is also sensitive to the expansion history, where dark energy causes a shift in the CMB power spectrum peaks. Planck Collaboration et al. (2020) find $\Omega_\Lambda = 0.6889 \pm 0.0056$ from a combination of CMB and BAO observations. This means that (as shown in Figure 1.1) over 68% of the content of the universe is in the form of dark

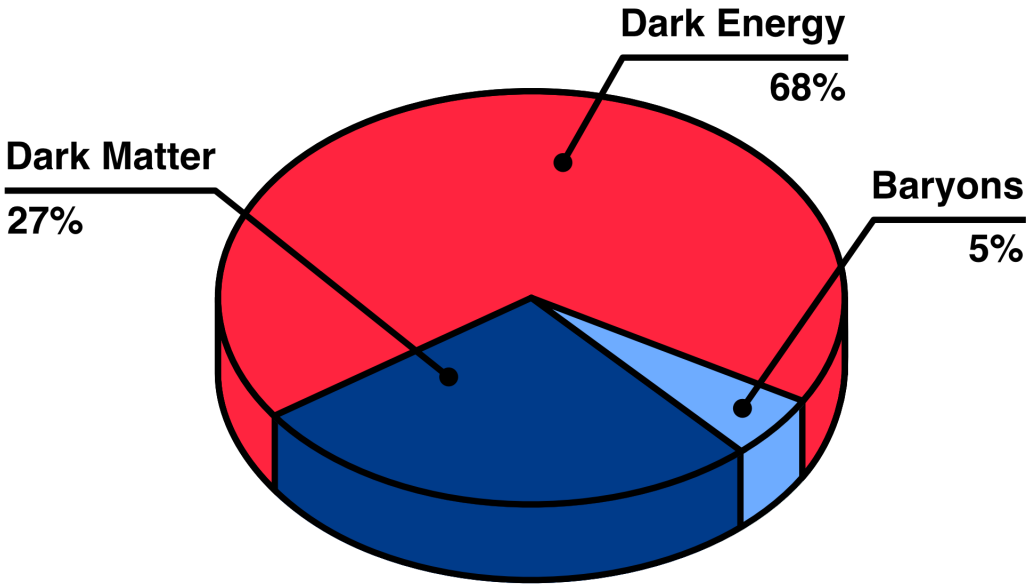


Figure 1.1: Energy content of the universe.

energy, 27% is dark matter, and only the remaining 5% is baryonic — 95% of the universe is *dark* in a very fundamental sense, and has so far never been directly detected!

1.1.3 Inflation

Some properties of the universe are not fully explained by the ΛCDM model of cosmology. Firstly, the energy density of the universe is observed to be very precisely consistent with $|\Omega_0 - 1| = 0$, i.e. a flat universe. This implies that the initial density of the universe was extremely precisely tuned, as $|\Omega - 1|$ is amplified by $\sim 10^{62}$ between the Planck epoch ($< 10^{-43}$ s after the Big Bang) and today (determined by evolving the Friedmann equations backwards in time through the epochs of matter and radiation domination). This is the *flatness problem.*

Secondly, the CMB has the same temperature in every direction (averaged over large enough scales that the anisotropies do not contribute, and accounting for the dipole due to the peculiar velocity of the solar system), which implies that, at some point before the formation of the CMB, all matter in the universe that was involved in the formation of the observable CMB was in thermal equilibrium. This conflicts with ΛCDM, as, in a universe with strictly the contents that we infer to exist today, regions of the CMB

would be causally disconnected from others, so there would have been no way to achieve thermal equilibrium across the whole sky. This can be demonstrated by comparing the angular diameter distance from the present to the last scattering surface at $z_{\rm rec} \approx 1100$ (i.e. the size of the observable universe at $z_{\rm rec}$)

$$d_A(z_{\rm rec}) = \frac{1}{1+z_{\rm rec}} \int_0^{z_{\rm rec}} \frac{dz}{H(z)} \approx 12.61\,{\rm Mpc}, \tag{1.9}$$

(assuming a near-flat cosmology after $z_{\rm rec}$) and the physical size of the *horizon* (the maximum distance light could have travelled since the beginning of the universe) at the time of CMB formation

$$d_{\rm hor} = \frac{1}{1+z_{\rm rec}} \int_{z_{\rm rec}}^{\infty} \frac{dz}{H(z)} \approx 0.25\,{\rm Mpc}. \tag{1.10}$$

Thus, only regions within 1.1° patches of the CMB should have ever been in causal contact with each other, so there would be no reason to expect that the entire CMB would be in thermal equilibrium, or indeed that the observable universe would be apparently homogeneous on large scales. This is the *horizon problem*.

Guth (1981) proposed a solution to these problems: an *inflationary* universe. Such a universe undergoes a period of extremely fast expansion at very early times ($t < 10^{-32}$ s, while the temperature of the universe exceeded $k_{\rm B}T \sim 10^{15}$ GeV). In order to solve the horizon problem, this period of inflation must have allowed the entire observable universe to originate from the same patch of causally-connected space. This can be understood by considering the comoving Hubble radius $(aH)^{-1}$ — the maximum comoving distance photons can travel from *now*, beyond which regions cannot *currently* be in causal contact, but may have been in the past. We require that the comoving Hubble radius must have been larger before inflation than it is today by a factor $\sim 10^{28}$ (Dodelson, 2003), so that the comoving volume containing the observable universe was in causal contact before the formation of the CMB (demonstrated in Figure 1.2). However, in a non-inflationary radiation/matter-dominated ΛCDM universe, $(aH)^{-1} \equiv 1/\dot{a}$ *increases*[2], so new physics is

[2]The scale factor goes as $a \propto t^{\frac{1}{2}}$ during radiation domination, then $a \propto t^{\frac{2}{3}}$ during matter domination, so $\dot{a}$ decreases with time up until at least $z_{\rm de} \approx 0.28$, long past the formation of the CMB and most of the observable universe.

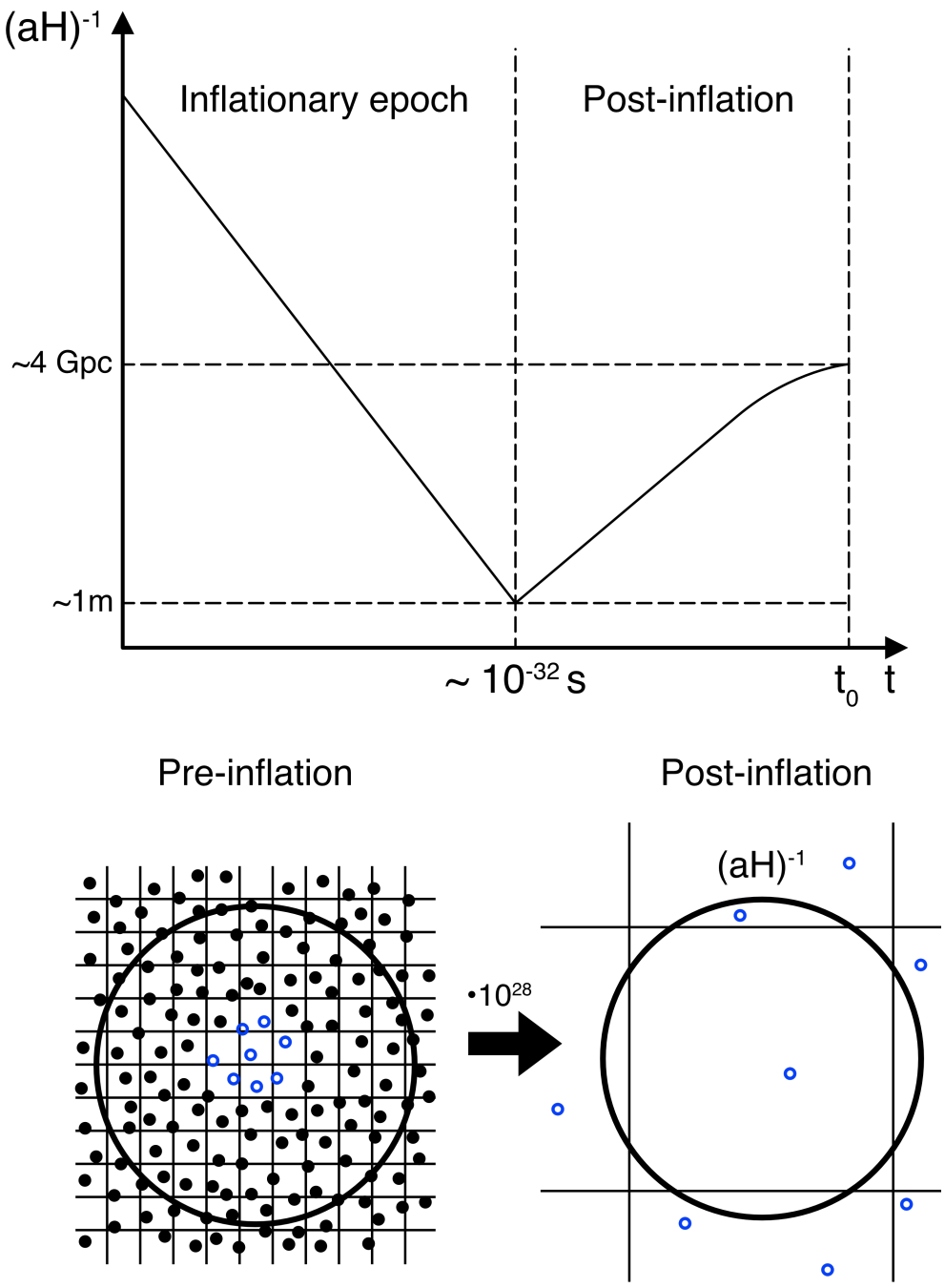


Figure 1.2: *Top*: Schematic showing the comoving Hubble radius vs. time. During the inflationary epoch, $(aH)^{-1}$ decreased at least 10^{28} fold, allowing the comoving $\sim$ 1 m patch[a] that grew to become the observable universe to be in causal contact pre-inflation. *Bottom*: Diagram showing particles on a comoving grid in the universe before and after inflation, demonstrating that a decreasing comoving Hubble radius (circled) allows particles that were once in causal contact pre-inflation to no longer be able to communicate with each other, as they pass out of the Hubble radius during inflation. Blue ringed particles are the same pre- and post-inflation. Both panels are based on figures from Dodelson (2003).

[a]This is calculated by extrapolating backwards in time from the present $(aH)^{-1}$, assuming radiation dominance $a \propto t^{\frac{1}{2}}$, so that $a \approx 10^{-26}$ at the end of inflation.

required to allow this. Thus, during inflation, it is required that

$$\frac{d}{dt}(aH)^{-1} < 0 \tag{1.11}$$

$$\implies \frac{d}{dt}(aH) > 0 \tag{1.12}$$

$$\implies \frac{d}{dt}\left(a\frac{\dot{a}}{a}\right) > 0 \tag{1.13}$$

$$\implies \frac{d^2a}{dt^2} > 0. \tag{1.14}$$

That is, the expansion of the universe must have been accelerating during this epoch. An elegant way to explain this is that there exists a quantum scalar field, associated with an *inflaton*, that was excited initially, contributing to the content of the universe in the same way as a cosmological constant (but not the same Λ that is inferred today, associated with dark energy), causing exponential expansion during inflation, but 'rolled' towards its ground state (Linde, 1982; Albrecht & Steinhardt, 1982), eventually contributing negligibly, halting inflation and beginning the epoch of radiation-dominance.

This hypothesised exponential expansion would also damp deviations in the curvature $|\Omega - 1|$ in the pre-inflation universe by a factor 10^{28}, explaining the apparent flatness of the universe. Thus, two cosmological problems can be solved with the same simple model, making inflation a compelling explanation.

1.2 Structure Formation

After the epoch of radiation-dominance and the formation of the CMB, the baryonic plasma decouples from the radiation (which has a sufficiently-reduced energy density from the expansion of the universe), cooling and recombining into neutral gas. The random density fluctuations present in the CMB at its formation began to gravitationally collapse into sheets, filaments, and haloes of the cosmic web. At first, the density fluctuations are small enough that the dynamics are, to a very good approximation, linear, and baryonic pressure, magnetic fields, and streaming velocity can be ignored. In this regime,

the matter (with density ρ, velocity $\vec{u}$, in gravitational potential ϕ) evolves according to

$$\dot{\rho} + \vec{\nabla} \cdot (\rho \vec{u}) = 0, \tag{1.15}$$

$$\dot{\vec{u}} + (\vec{u} \cdot \vec{\nabla})\vec{u} = -\vec{\nabla}\phi, \tag{1.16}$$

$$\nabla^2 \phi = 4\pi G \rho, \tag{1.17}$$

which are, respectively, conservation of mass, conservation of momentum, and the Poisson equation. Between redshifts $z_{\mathrm{eq}} = 3387$ and $z_{\mathrm{de}} = 0.28$, the universe is matter-dominated, so the term associated with Λ is omitted from the Poisson equation above. Changing to comoving coordinates $\vec{x}' = \vec{x}/a, \vec{u}' = \vec{u} - \dot{a}\vec{x}'$, then linearising about the average density $\bar{\rho}$ and taking the Fourier transform, the dynamics of the *density contrast* $\delta = (\rho - \bar{\rho})/\bar{\rho}$ on scales with wavenumber $\vec{k}$ is described by

$$\ddot{\delta_k} + 2H\dot{\delta_k} - 4\pi G \bar{\rho} \delta_k = 0, \tag{1.18}$$

which has solution

$$\delta_k = C_1 t^{2/3} + C_2 t^{-1}, \tag{1.19}$$

where C_1, C_2 are integration constants. The solution is composed of a growing mode and a decaying mode, where the growing mode dominates given enough time. This describes the evolution of the density field in the matter-dominated universe until δ grows to be too large and the linear approximation no longer holds.

Birkhoff's theorem (first proved by Jebsen 1921) states that a spherically symmetric solution to the Einstein field equations in a vacuum must be *asymptotically flat* and *static*, and is thus uniquely the Schwarzschild solution. Analogously to the shell theorem in Newtonian gravity, this means that the metric around any spherically symmetric system is independent of the interior mass profile, and that the interior metric of a spherical shell is *Minkowski* (flat). A corollary of this is that the dynamics of a sphere embedded in a homogeneous universe depend only on the contents of the sphere, not its surroundings.

Thus, we can treat such a sphere as evolving separately to the universe, according to the same Friedmann equations but with different density parameters. An overdense homogeneous sphere (in a matter-dominated universe with $\Lambda = 0$) behaves as an independent closed sub-universe (first considered by Gunn & Gott, 1972), and serves as a useful model in understanding galaxy formation. The radius r at time t of such a sphere obeys the parametric solution[3]

$$r = A(1 - \cos\theta), \tag{1.20}$$

$$t = B(\theta - \sin\theta), \tag{1.21}$$

$$\theta \leq 2\pi, \tag{1.22}$$

where the strictly positive constants A and B obey $A^3 = GMB^2$ and the conserved mass of the sphere is $M = 4\pi\rho_0 r_0^3/3$ (the subscript 0 indicating initial values at $t = 0$). In this solution the sphere expands initially but turns around at $\theta = \pi$ and then collapses into a singularity at $\theta = 2\pi$. Of course, the collapse into a singularity is not realistic. Eventually the particles inside the sphere gain enough kinetic energy that their random motions prevent further collapse and equilibrium is reached (which happens within ∼10% of a typical particle orbit time, in a process called *violent relaxation*; Lynden-Bell 1967). Thus, the system is said to be *virialised*, with radius $r_{\rm vir}$ and (time-averaged) kinetic energy T related to the (time-averaged) potential energy V according to the virial theorem $T = -V/2$ (neglecting external pressure on the surface of the system, and the effect of Λ). The total energy is then $E = V + T = V/2$ when virialised. Assuming the kinetic energy of the system is negligible at the radius of turnaround $r_{\rm max}$, we have $V(r_{\rm max}) = V(r_{\rm vir})/2$ by conservation of energy. Using the Newtonian potential $V \propto -1/r$, we see that the virial radius is $r_{\rm vir} = r_{\rm max}/2$, which occurs at $\theta = 3\pi/2$. Here the *virial density contrast* is $\Delta_{\rm vir} \equiv \delta(r_{\rm vir}) = 18\pi^2 - 1 \approx 177$, which provides a convenient way of defining $r_{\rm vir}$ in inhomogeneous systems where the radius is less sharply defined: it is the minimum radius of a sphere such that the density contrast of the matter contained within is greater than

[3]Note that, to first order, this solution is $\delta \propto t^{2/3}$ as expected from Eq. 1.19.

$\Delta_{\rm vir}$.

Now we seek to describe the formation of structure of the entire universe in a statistical manner, rather than a single idealised object. Press & Schechter (1974) developed a technique to predict the population of collapsed structures as a function of mass, which was then developed more rigorously by Lacey & Cole (1993), whose approach we summarise here. The density contrast just after the Big Bang is very close to being a Gaussian random field, so it can be fully described by the initial power spectrum

$$P_i(k) \propto k^n, \tag{1.23}$$

where n is the *spectral index*. Before and during recombination, processes act to alter the power spectrum, including, for example, Silk damping (Silk, 1968), in which the diffusion of photons reduces the power of fluctuations on small scales. These processes are accounted for in the *transfer function* T, which multiplies the primordial power spectrum to give the post-recombination spectrum

$$P(k) = P_i(k)T^2(k). \tag{1.24}$$

The transfer function leaves scales much larger than the horizon at the time of matter-radiation equality unchanged. These are the initial conditions from which we now proceed. Lacey & Cole (1993) first consider the homogeneous spherical collapse as described above, in which a spherical overdensity collapses to a singularity in time $t_{\rm coll} = 2\pi B$, after turning around at $t = t_{\rm coll}/2$. By comparing the early-time growth of this sphere to that of an unperturbed region of the background universe, and then extrapolating in the linear regime (Equation 1.19), the linear behaviour for spherical overdensities that collapse at $t_{\rm coll}$ is found to be

$$\delta_c(t) = \frac{(12\pi)^{\frac{2}{3}}}{10}\left(\frac{t}{t_{\rm coll}}\right)^{\frac{2}{3}}. \tag{1.25}$$

The cosmological density contrast field $\delta(\vec{x}, t)$ can be extrapolated to a fixed time t_0 in the same way, labelling this as $\delta(\vec{x}) \equiv \delta(\vec{x}, t_0)$ henceforth. Then, $\delta_c(t)$ can be considered a critical threshold that decreases with time, such that if $\delta(\vec{x}) \geq \delta_c(t)$, a region around $\vec{x}$

will have collapsed by time t at the latest. Now we seek to assign regions to haloes of mass M, by smoothing the extrapolated density contrast field with a window function $W_M(r)$. The smoothed field is the integral over all space

$$\delta(\vec{x}; M) \equiv \int W_M(|\vec{x} - \vec{y}|)\ \delta(\vec{y})\ d^3\vec{y}, \tag{1.26}$$

which can be written (by the convolution theorem) as the inverse Fourier transform of the product of the Fourier transforms of the density contrast $\delta_{\vec{k}}$ and the window function $\hat{W}_M$

$$\delta(\vec{x}; M) = \int \hat{W}_M(|\vec{k}|)\ \delta_{\vec{k}}\ e^{i\vec{k}\cdot\vec{x}}\ d^3\vec{x}. \tag{1.27}$$

The smoothed field can be relabelled by its variance

$$S(M) \equiv \sigma^2(M) = \left\langle \delta(\vec{x}; M)^2 \right\rangle, \tag{1.28}$$

which, for a scale free power spectrum $\left\langle \delta(\vec{x}; M)^2 \right\rangle \propto k^n$, is proportional to $M^{-(n+3)/3}$, and a k-space 'top-hat' chosen as the window function

$$\hat{W}_M(k) = \begin{cases} 1 & k \leq k_s(M), \\ 0 & k > k_s(M), \end{cases} \tag{1.29}$$

where $k_s(M) \propto M^{-1/3}$ is the wavenumber associated with the mass scale. This allows the exploitation of the fact that Gaussian random fields are *scale-free*: the contribution to the integrated density contrast from an infinitesimal shell in k-space is statistically independent of other shells. Thus, for any given realisation of the random field, the density contrast at the origin $\delta(S) \equiv \delta(\vec{x} = 0; S)$ will be a *Brownian* random walk in S: the change in $\delta(S)$ over an interval $[S, S + \Delta S]$ depends only on ΔS, not S.

The density of particles undergoing Brownian motion obeys a diffusion equation (Einstein, 1905), but here there is a mathematically equivalent scenario where the 'particles' are trajectories $\delta(S)$, S is the 'time' variable, and δ is the 'space' variable. Thus, the number density Q of trajectories $\delta(S)$ per unit S obeys

$$\frac{\partial Q}{\partial S} = D\frac{\partial^2 Q}{\partial \delta^2}, \tag{1.30}$$

where the factor D can be computed using Einstein's formula for the coefficient in the diffusion equation

$$D = \frac{1}{2\tau}\int_{-\infty}^{\infty} \Delta^2 P(\Delta) d\Delta, \tag{1.31}$$

where $P(\Delta)$ is the probability that a particle is displaced by Δ in a time interval τ. Without loss of generality, we can replace Δ with $\delta(S)$ and τ with S — i.e. we consider the change in δ over an interval S, beginning at the origin with $S = 0, \delta = 0$; the choice of initial values are irrelevant (due to the statistical independence of shells of matter in S or k space). Thus, using the fact that $\left\langle \delta^2 \right\rangle \equiv S$ and $\langle \delta \rangle = 0$, we have

$$D = \frac{1}{2S}\int_{-\infty}^{\infty} \delta^2 P(\delta) d\delta = \frac{\left\langle \delta^2 \right\rangle}{2S} = \frac{1}{2}, \tag{1.32}$$

in which we realise the convenience of the choice of parameter S. Now we seek the fraction of matter in the universe that belongs to haloes of mass M. This is the fraction of trajectories $\delta(M)$ (out of all possible realisations of the density field) that exceed $\delta_c(t)$ at a mass scale M, but are below it at any scale larger than this. Since the variance S decreases as M increases, this can be found by imposing a 'barrier' boundary condition on the system, where trajectories begin at $S = 0$ (infinite mass), and are absorbed by the barrier if $\delta(S) > \delta_c(t)$. The solution under these conditions (normalised such that the integral over δ is 1) is

$$Q(\delta, S; \delta_c) = \frac{1}{\sqrt{2\pi S}}\left[\exp\left(-\frac{\delta^2}{2S}\right) - \exp\left(-\frac{(\delta - 2\delta_c(t))^2}{2S}\right)\right] \tag{1.33}$$

(Chandrasekhar, 1943), where the second term is obtained by the method of images, cancelling out the density on the other side of the barrier. Then the probability density function for matter occupying a halo with density contrast variance S is $f_S(S; \delta_c)$, which is equal to the flux of trajectories crossing the barrier boundary (the fraction absorbed per S interval)

$$f_S(S;\delta_c) = -\frac{\partial}{\partial S}\int_{-\infty}^{\delta_c} Q\, d\delta, \tag{1.34}$$

which (applying Equation 1.30) is

$$\begin{aligned} f_S(S;\delta_c) &= -\left[\frac{1}{2}\frac{\partial Q}{\partial \delta}\right]_{-\infty}^{\delta_c} \\ &= \frac{\delta_c}{(2\pi)^{1/2} S^{3/2}} \exp\left(-\frac{\delta_c^2}{2S}\right). \end{aligned} \tag{1.35}$$

Finally, the halo mass function $\frac{dn}{dM}$ can be obtained by transforming the occupancy fraction in S to be in M, multiplying by the average density of the universe ρ_0 (to get the total mass occupying haloes with mass M), and dividing by the halo mass (to get the total number of haloes with mass M)

$$\begin{aligned} \frac{dn}{dM} &= \frac{\rho_0}{M} f_S(S,\delta_c)\left|\frac{dS}{dM}\right| \\ &= \sqrt{\frac{2}{\pi}}\frac{\rho_0}{M^2}\frac{\delta_c(t)}{\sigma(M)}\left|\frac{d\ln\sigma}{d\ln M}\right| \exp\left(-\frac{\delta_c(t)^2}{2\sigma^2(M)}\right). \end{aligned} \tag{1.36}$$

Given that $\sigma^2(M) \propto M^{-(n+3)/3}$, this is

$$\frac{dn}{dM} \propto M^{(n-9)/6} \exp\left(-\frac{\delta_c(t)^2}{2\sigma^2(M)}\right), \tag{1.37}$$

which agrees quite well with the observed mass spectrum below the mass cut-off ($\lesssim 10^{14}\,\mathrm{M}_\odot$ at $z = 0$), but is too steep at the high mass end, needing a softening of the exponential cut-off to fit well (Böhringer et al., 2017; Driver et al., 2022). An arguably more important and fundamental prediction of this model is that *small structures collapse before large ones*, that is, structure formation in the universe is a *bottom-up* process. This is exhibited in the variance of the smoothed density contrast $\sigma^2(M) \propto M^{-(n+3)/3}$, which is larger at smaller masses, so larger fluctuations are more probable, which then collapse more quickly as they exceed $\delta_c(t)$ at earlier times. Larger haloes are thus formed hierarchically, by the *mergers* of smaller ones (as in Figure 1.3).

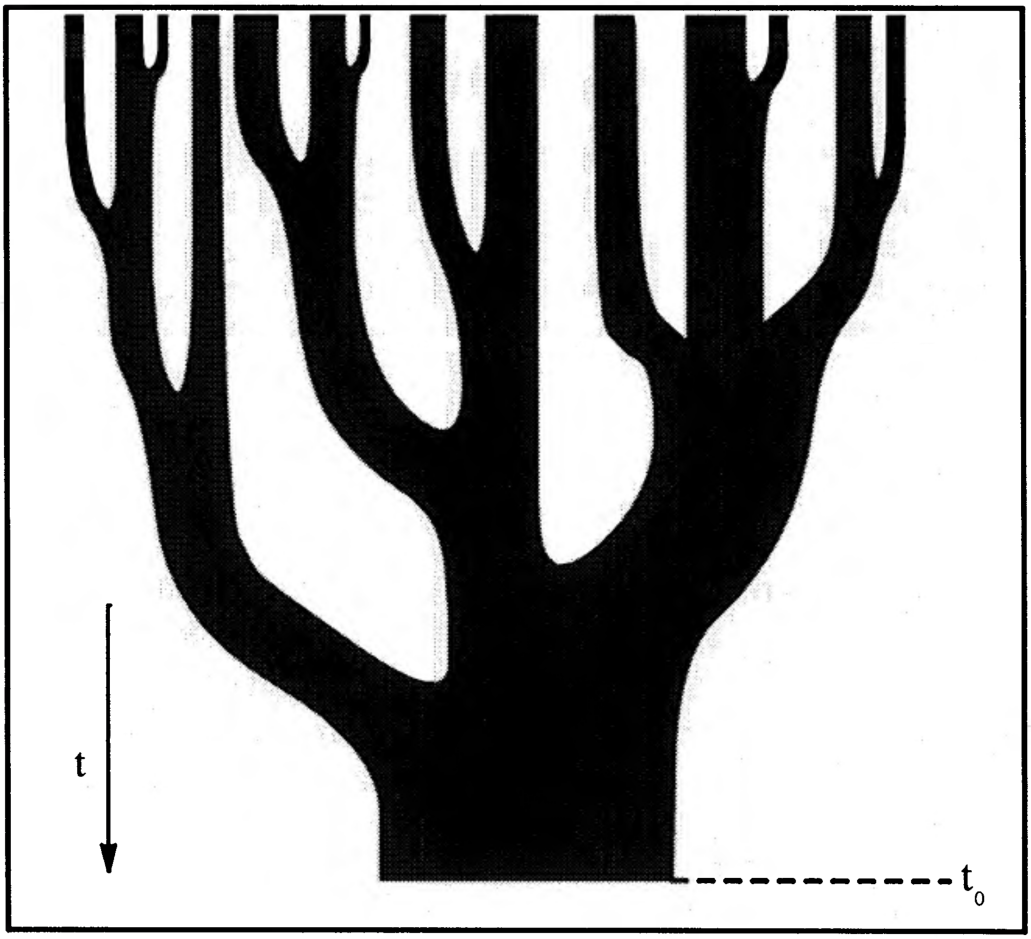


Figure 1.3: Illustration of the tree-like structure of halo mergers. A resultant halo is formed by the merging of smaller ones, which tend to form earlier. The width of the 'branches' represent the halo masses. Time increases downwards, and t_0 marks the present time. This figure is adapted from (Lacey & Cole, 1993).

1.3 The First Stars and Galaxies

Following the formation of nonlinear structure, gas density begins to increase to the point where baryonic processes begin to become relevant. Eventually gas can collapse and radiatively cool (via excitations/ionisations of primordial chemical species) to the point of forming the very first stars, bringing an end to the cosmic dark ages. From this point onward, the dynamics of the universe become a highly complex and nonlinear system, where dark matter, gas, dust, radiation, and cosmic rays are separate interacting fluids, and stars and black holes inject energy from the smallest scales to the largest in an *inverse energy cascade* (completely violating the self-similarity present in the dark matter-only evolution).

1.3.1 Primordial Chemistry

After recombination, gas collapses into haloes alongside dark matter, and begins to radiatively cool via excitations of the primordial atoms, ions and molecules within it. At first this gas consists almost entirely of neutral atomic hydrogen and helium, which are most efficiently excited by collisional ionisations, or recombinations (which require free electrons and ions to exist beforehand). The dominant atomic cooling processes here are collisional excitation of the Lyman-α transition, and collisional ionisation (leading to subsequent recombination)

The ionisation energy of hydrogen, and the energy of the Lyman-α line — 13.6 eV and 10.2 eV respectively — correspond to a temperature of $E/k_{\mathrm{B}} \approx 10^4$K, so the atomic cooling rate below that temperature has an exponential cut-off (from the Boltzmann distribution of gas particle energy) as excitations of 10.2 eV or more become highly unlikely (as shown in Figure 1.4). At this temperature, gas can support itself against gravity on large scales, so is unable to collapse and form stars; that is, the *Jeans length*, defined as the maximum radius of a gas cloud (for a given mass) such that it can still support itself against its own gravity

$$\lambda_J \equiv \sqrt{\frac{15 k_{\rm B} T}{4\pi G m_p \mu \rho}}, \tag{1.38}$$

(where μ is the average mass of gas particles in atomic units, and dark matter contributes to ρ) is comparable to the size of haloes. Therefore, atomic cooling alone is insufficient to allow the formation of stars and galaxies as we see them today.

However, the atomic hydrogen in the gas can undergo chemical processes, eventually forming H_2, which has a dense series of rotational and vibrational transitions at infrared rather than optical/UV energies. Thus, H_2 cooling is far more efficient than cooling from atomic H and He at temperatures below 10^4K (demonstrated in Figure 1.4). The most important H_2 formation processes (Abel et al., 1997) are the *H^- channel*

$$\mathrm{H} + \mathrm{e}^- \longrightarrow \mathrm{H}^-, \tag{1.39}$$

$$\mathrm{H}^- + \mathrm{H} \longrightarrow \mathrm{H}_2 + \mathrm{e}^-, \tag{1.40}$$

and the *H_2^+ channel*

$$\mathrm{H}^+ + \mathrm{H} \longrightarrow \mathrm{H}_2^+, \tag{1.41}$$

$$\mathrm{H}_2^+ + \mathrm{H} \longrightarrow \mathrm{H}_2 + \mathrm{H}^+, \tag{1.42}$$

although there is a rich set of other reactions (explored in the next chapter) that contribute to the formation and destruction of H_2, including the three-body reaction that becomes dominant at very high hydrogen densities ($n_H \sim 10^6$ cm^{-3}).

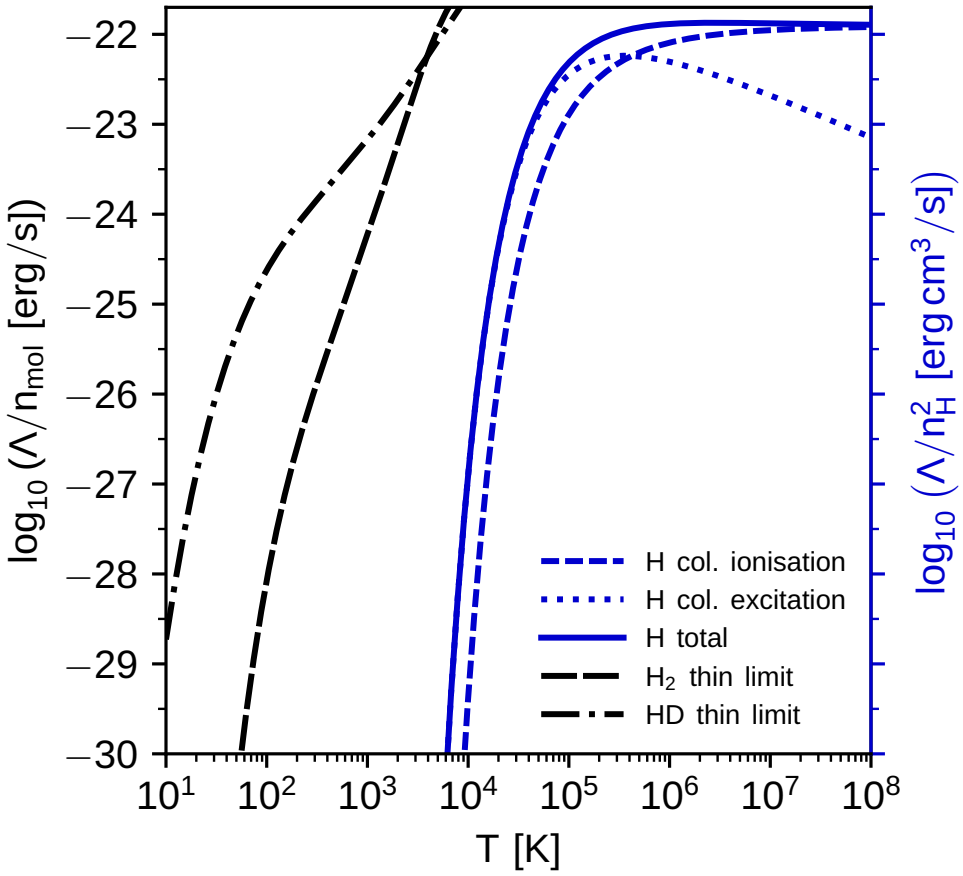


Figure 1.4: Radiative cooling rates per H nucleus vs. temperature for collisional ionisation and excitation (both from Cen 1992) in atomic hydrogen gas with an ionised fraction of 0.00031 (approximately the fraction after recombination) and proton density of $n_H = 1.0\,\mathrm{cm}^{-3}$, and per molecule for H_2 (Galli & Palla, 1998) and HD (Lipovka et al., 2005) in the optically thin limit. The steep cut-off of atomic cooling below $T \lesssim 10^4\,\mathrm{K}$ means that even small amounts of H_2 or HD can cool much more effectively than pure atomic gas at low temperature.

1.3.2 Population III Stars

The first stars to form in the universe are known as *Population III* stars, which have never been observed directly, but are the source of the first significant injection of metals into the initially pristine gas in the first galaxies, from which the next generation of stars form, and so are deeply important to galaxy evolution in the early universe. As they necessarily begin life in metal-free gas, their formation depends strongly on the formation of the primary low-temperature coolant H_2, so understanding the chemistry in primordial gas is key to predicting the properties of the first stars.

One-zone modelling by Omukai et al. (2005) predicts that gas fragments less strongly in lower-metallicity gas (as metals increase the efficiency of cooling, see Figure 1.5), allowing star-forming clumps of gas to remain intact at higher masses, potentially allowing

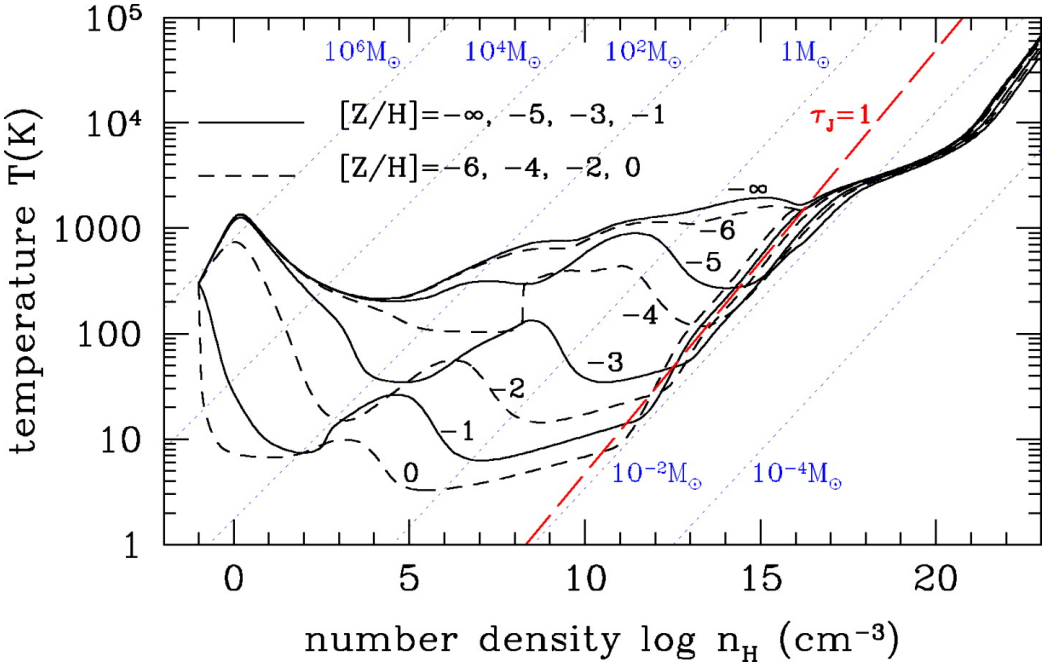


Figure 1.5: Evolution of temperature vs. number density for protostellar clouds at different metallicities, calculated with a one-zone model where chemicals are evolved according to a primordial network, and density evolves according to a free-fall law. Blue dashed lines mark the possible temperature and density of Jeans-stable clouds at the labelled masses. The red dashed line shows where the cloud becomes optically thick, at which point radiative cooling becomes inefficient. Lower metallicity clouds are able to remain hotter while they collapse, allowing clouds to remain Jeans-stable at higher masses, instead of fragmenting. This figure is from (Omukai et al., 2005).

Population III stars to form with greater masses than stars in e.g. the Milky Way. However, the initial mass function (IMF) for Population III stars is disputed, with other theoretical predictions spanning many orders of magnitude (Greif et al., 2011; Hirano et al., 2015; Stacy et al., 2016; Hosokawa et al., 2016; Hirano & Bromm, 2017; Sharda et al., 2019), and there is no conclusive observational evidence available.

Depending on their mass, Population III stars may end their lives in ways that other stars are not able to (Heger et al., 2003). Below $40\,M_\odot$, similarly to other stars, they should produce type II core-collapse supernovae after they burn through their nuclear fuel. If they are able to reach masses between $40\,M_\odot$ and $140\,M_\odot$ they are expected to collapse directly into a black hole, without causing a supernova or releasing any metals. Above $\sim 140\,M_\odot$, they may produce *pair-instability supernovae*, where very high energy gamma rays from the core of the star form electron-positron pairs (which then must annihilate to exert pressure, but this happens too slowly in less dense material), reducing the radiation pressure and allowing the star to collapse in on itself, increasing the core temperature, and thus the energy density of the gamma rays and the rate of the runaway reaction. These supernovae are extremely energetic, expected to completely destroy their progenitor star, leaving no remnant. At very high masses ($> 10^4\,M_\odot$), stars may succumb to the general relativistic instability, in which photons in the core begin to contribute gravitationally, acting against the pressure holding the core in equilibrium (Fowler, 1966; Chen et al., 2014).

Rest frame UV observations of star-forming galaxies in the epoch of reionisation (e.g. Bunker et al., 2010) suggest that a top-heavy stellar initial mass function — potentially a result of a low-metallicity environment — may be needed to explain UV luminosities at $z \sim 7$, to account for the apparent rate of reionisation. This is supported by submillimetre observations (rest frame infrared at this redshift), that are difficult to explain in simulations without a low Carbon-Oxygen abundance ratio (suggesting enrichment by core-collapse, rather than thermonuclear, supernovae), and an IMF with high power at the high mass end as a result of lower metallicity (Katz et al., 2022). Population III stars may be significantly responsible for the resultant metal enrichment of galaxies at this redshift, depending on the masses they are likely to form at, and how they affect the subsequent

evolution of the galaxies they form in.

Thus, the formation of Population III stars, and their (potential) subsequent violent supernovae, is critical in understanding the evolution of the first galaxies, and their role in reionisation. But the fact that the statistical distribution of their masses is largely unknown, and as-yet unobservable, makes studying them very difficult. Nonetheless, under certain conditions, galaxies containing population III stars may exhibit unique spectral signatures, depending on their properties (e.g. Woods et al., 2021; Katz et al., 2023), thus potentially becoming observable by JWST (Zackrisson et al., 2011) or the upcoming ELT (Grisdale et al., 2021). Lazar & Bromm (2022) also consider that the IMF of population III stars may be probed by observing the supernovae they produce, and the GAIA mission may even be able to detect individual low-mass zero-metallicity stars locally (Komiya et al., 2016), or higher mass zero-metallicity progenitors of gamma ray bursts (de Souza et al., 2012).

1.4 Reionisation

With the formation of the first stars, galaxies start to emit ionising radiation, ending the Dark Ages and beginning the Epoch of Reionisation (EoR). Now ionised bubbles form around galaxies in the neutral intergalactic medium, expanding until the entire medium is ionised. This proceeds at different rates for hydrogen and the two ionisation states of helium, due to the differing threshold energies, cross sections, and photon counts in each band, but 'reionisation' conventionally refers to that of only hydrogen, as this is the most important absorber of ionising far-UV light that escapes galaxies.

The first detection of reionisation in progress was in the spectrum of quasars (Becker et al., 2001), where Gunn & Peterson (1965) predicted that an absorption feature — the Gunn-Peterson Trough, see Figure 1.6 — would exist where light from a quasar has passed through neutral hydrogen, exciting Lyman-series transitions with very high efficiency. Quasars below $z \approx 6$ do not exhibit this trough, and so this was strong evidence for reionisation occurring earlier than redshift 6. This is further supported by later observations of many quasars (Fan et al., 2002, 2006; Hoag et al., 2019), and of the spectra of Lyman α

emitters (Ouchi et al., 2010).

Radio observations of the brightness temperature contrast (against the CMB) of the 21 cm hyperfine line in neutral hydrogen may enable a time-resolved measurement of the ionised fraction of the universe during the EoR. The neutral hydrogen in the IGM during the EoR absorbs 21 cm CMB photons, which reduces the observed brightness temperature of the global 21 cm signal, resulting in a dip across wavelengths associated with the rest frame 21 cm line at redshifts where neutral hydrogen exists. A measurement of this signal has been attempted by The Experiment to Detect the Global EoR Signature (EDGES; Monsalve et al., 2017), supporting an extended reionisation history with duration $\Delta z \gtrsim 0.5$ and ending at a redshift $z_{\rm re} \gtrsim 6.5$, but this has been found to be potentially erroneous, with the origin of the signal likely to be systematic, rather than cosmological (Singh et al., 2022). The upcoming Square Kilometre Array (SKA) is expected to be able to *spatially resolve* the 21 cm signal from reionisation down to $z > 20$ (Koopmans et al., 2015; Barry et al., 2022), measuring not just the global neutral fraction as a function of redshift, but the statistical fluctuations in the rate of reionisation across the sky.

Reionisation also has an effect on the CMB power spectrum. Thompson scattering of the CMB by electrons in the ionised IGM reduces the observed anisotropies, allowing a measurement of the 'optical depth to reionisation' by missions such as Planck (Planck Collaboration et al., 2020), which gives a model-independent redshift of reionisation of $z_{\rm re} = 7.64$ (midpoint, not end), and favours an extended, late-ending reionisation, compatible with quasar measurements.

What the dominant sources of ionising radiation are in this epoch, is an open question. The debate can be broken down into two key issues: low-mass vs. high-mass galaxies, and stars vs. quasars. It is not enough to simply count the stars or quasars as a function of galaxy mass, as ionising radiation is often obscured by gas in its host object (characterised by the *escape fraction* $f_{\rm esc}$). The escape fraction of an object depends on the configuration of gas and radiation within it, subject to the same stochasticity, complexities and non-linearities that affect other physical processes in galaxy formation, and so $f_{\rm esc}$ can vary greatly, even amongst objects of similar mass. This makes it very difficult to predict the average $f_{\rm esc}$ for a class of object without detailed 3D simulations of many such objects.

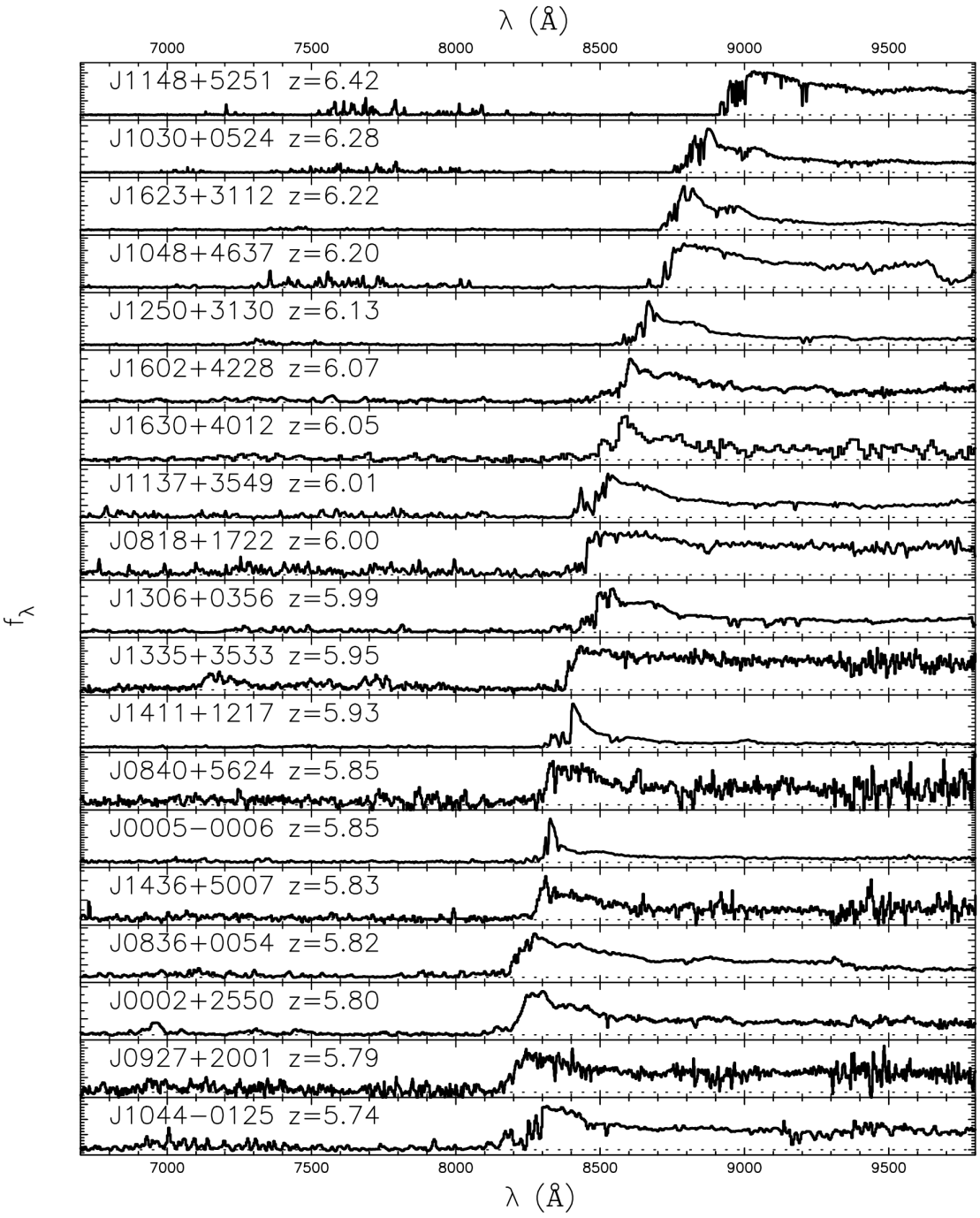


Figure 1.6: Observer-frame spectra of 19 quasars, sorted by redshift. Quasars with $z > 6$ show strong Gunn-Peterson troughs below the Lyman-alpha wavelength (left of the peak), but as redshift decreases past 6, features — the Lyman-alpha forest — begin to appear in some of the spectra in this region, indicating that there is no longer a dense column of neutral hydrogen along the line of sight. This figure is from (Fan et al., 2006).

This is complicated further by the fact that galaxies themselves are affected by reionisation, especially at lower masses, so a simulation must form galaxies across the mass spectrum and model their interactions self-consistently to be fully convincing. This is virtually impossible with current computers, as the resolution required to accurately capture the properties of the interstellar medium would be far too expensive to realise in a simulation of even a single high-mass galaxy. However, it is at least possible to simulate lower-mass galaxies at a high resolution (e.g. Rosdahl et al., 2018; Kimm et al., 2017; Wise et al., 2014), and analyses of these simulations and of observations (e.g. Haehnelt et al., 2001; Cowie et al., 2009; Bouwens et al., 2012; Fontanot et al., 2014; Jiang et al., 2022) favour stars in dwarf galaxies with masses $\sim 10^9 M_\odot$ as the primary sources of reionisation.

1.5 Thesis Outline

In the upcoming chapters of this thesis, I present a novel fully non-equilibrium, radiation-coupled model of molecule formation in primordial gas, self-consistently linked with the radiation transport and hydrodynamics of Ramses-rt. I aim to demonstrate the importance of using such a model (compared to simpler treatments that assume some degree of chemical equilibrium) in cosmological simulations to self-consistently capture the formation of Population III stars, and the subsequent effect of these on their host galaxies.

Chapter 2 is an introduction to the large collection of numerical methods involved in cosmological simulations of galaxy formation, containing a detailed description of the methods I use to simulate the chemistry of primordial gas. In Chapter 3, I present a suite of tests that I have performed to validate my treatment of the photochemistry. In Chapter 4, I show how cosmological simulations using my full non-equilibrium chemical model make different predictions to those that make more approximations concerning H_2 formation/cooling (either assuming an equilibrium concentration of H^- or H_2 itself), or omit radiation or non-equilibrium chemistry entirely. Finally, in Chapter 5, I summarise the key results.

Chapter 2

Numerical Methods

The cosmological simulations of galaxy formation in this thesis couple together numerical solvers for hydrodynamics, gravity and radiation with a collection of models for sub-grid scale astrophysical phenomena, including star formation, supernovae, and gas chemistry. In this chapter I describe the constituent models and methods of these cosmological simulations — including a rich photochemical system I have implemented — and the techniques used to post-process simulation data.

2.1 RAMSES-RT

The work in this thesis makes use of the Ramses-rt radiation-hydrodynamics code (Rosdahl et al., 2013), which is an extension to Ramses (Teyssier, 2002) that adds radiative transfer (RT) capability (from now on the code will be referred to as Ramses except in the context of RT). This massively parallel code features a state-of-the-art multigrid Poisson gravity solver (Guillet & Teyssier, 2011), a Particle Mesh treatment of stars and dark matter particles, *Eulerian* hydrodynamics (solved by a Finite Volume Method) with Adaptive Mesh Refinement, and Radiative Transfer with a finite speed of light. In this section I explain these elements in mathematical detail.

2.1.1 Gravity

Ignoring strong-field/high-velocity general-relativistic effects (but *not* the expansion of the universe according to the Friedmann equation), we can accept a universal, observer-independent gravitational field, which is produced by matter according to the Poisson equation

$$\nabla^2\phi = 4\pi G\rho, \tag{2.1}$$

where ρ is the density of matter, and ϕ is the *gravitational potential*, related to the *gravitational field* by $\vec{g} \equiv -\nabla\phi$. This field exerts a force on matter proportional to its mass, causing an acceleration independent of mass

$$\vec{a_g} = \vec{g} \equiv -\vec{\nabla}\phi. \tag{2.2}$$

This acceleration affects stars, dark matter and gas in the same way, but these are represented differently in Ramses. Stars are effectively non-collisional[1] (with each other and with gas), so they are modelled as particles in the code, although these do not necessarily represent individual stars — they can instead be statistical representations of populations, which is more useful in galaxies where it would be too expensive to compute dynamics for individual stars and store them in memory.

Similarly, dark matter is treated as a collection of particles that interact only gravitationally. Since in ΛCDM, dark matter is considered to be a pressure-free (hence collisionless) perfect fluid, these particles are a numerical discretisation of the fluid into packets: their mass is not the mass of the particles of whatever substance makes up dark matter, but rather a number chosen for numerical reasons — the lower the mass, the better resolved the fluid is.

[1] This is an approximation (though a very accurate one for most stars), as stars may occasionally collide with each other in reality, and have a kinematic interaction with the gas they move through and accrete. In the case of newly forming Population III stars, the accreted material may contribute more than the initial mass of the star, and so should dominate the kinematic properties, i.e. these stars can be *highly* collisional with their surrounding gas. This effect is not considered in the work in this thesis, as the initial formation and evolution of Population III stars is already poorly understood. Rather we seek to understand the effect a given population of Population III stars has on its host galaxy, rather than predict the properties of the population from first principles.

Exactly solving the gravitational dynamics of these particles and gas cells is a problem with N^2 complexity (for N cells and particles), so quickly becomes intractable for well-resolved simulations of large systems. However, there exist various high-accuracy approximate numerical techniques to solve this problem, in which masses can be grouped together so that the contribution of each individual pair does not have to be considered. I describe the most relevant and instructive of these here.

2.1.1.1 Exact Particle-Particle Force Calculation

In general, the gravitational force between two particles of masses m_1, m_2 at positions $\mathbf{x}_1, \mathbf{x}_2$ is

$$\frac{Gm_1m_2(\vec{x}_1 - \vec{x}_2)}{||\vec{x}_1 - \vec{x}_2||^3}. \tag{2.3}$$

As dark matter particles in our simulations are approximate representations of a pressure-free fluid, we wish to avoid artificial n-body scattering, so very short-range interactions should be smoothed over a scale ϵ. This should be chosen such that large-angle two-body scattering is impossible; that is, ϵ should be larger than the impact parameter for these interactions. Choosing a smaller particle mass allows a shorter smoothing length. Then, the inter-particle force (if either of 1 or 2 are a DM particle) is

$$\frac{Gm_1m_2(\vec{x}_1 - \vec{x}_2)}{(||\vec{x}_1 - \vec{x}_2||^2 + \epsilon^2)^{3/2}}. \tag{2.4}$$

Computing this for each of the N^2 pairs of masses in a simulation very quickly becomes impossible as the simulation size increases.

2.1.1.2 Barnes-Hut Algorithm

Barnes & Hut (1986) realised that contributions to the gravitational force on a particle from groups of distant masses could be grouped together, significantly reducing the complexity of the problem. They construct an octree (where each cell is either a 'leaf' node of the tree, or contains 8 equally-sized cells), such that every particle exclusively occupies a leaf cell (as shown in Figure 2.1). To calculate the force on a particle, the tree is then traversed,

looping through each node of the tree beginning at the root (the largest cell, containing all particles). If a node is sufficiently physically distant from the starting particle, the masses of the particles it contains can be summed, so that only a single force from the cell centre-of-mass is seen.

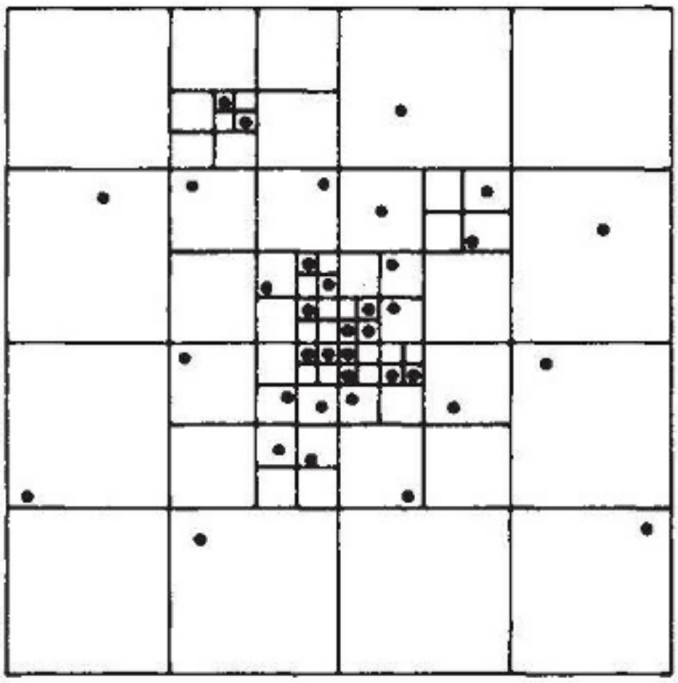


Figure 2.1: An example of a quadtree (2D equivalent to an octree) built around particles such that each exclusively occupies its own cell (Barnes & Hut, 1986).

This reduces the complexity to $N \log(N)$, allowing much more efficient simulation with little loss of accuracy. However, there are disadvantages compared to other methods. The technique requires that the tree is reconstructed completely every time step. Additionally, this algorithm does not make use of the fact that not only do distant masses in a group contribute similarly to the force on a particle, but particles that are close to each other should experience a similar acceleration from the same group of distant masses. This fact can be utilised by computing the gravitational field/potential across the space (as in the Particle Mesh method below), rather than immediately calculating the force on each particle.

2.1.1.3 Particle Mesh

Ramses employs a *Particle-Mesh* (PM) algorithm, in which the gravitational field is calculated on a discrete mesh by solving Poisson's equation, and then particles are evolved according to the gravitational acceleration computed locally from the field. Thus, information from distant masses is encoded in the field, and is efficiently reused by groups of nearby particles that see the same field. The code also uses an octree grid of gas cells for the Eulerian hydrodynamics (described in detail later in this section), which provides a convenient shared mesh structure for use in the PM calculation, that does not require total reconstruction every step.

To begin solving the Poisson equation, we must first compute the density field at the mesh cell centres. Here, there is a choice of discretisation in how point particles contribute to the density. The simplest choice is the *nearest grid point* (NGP) scheme, in which the density in a cell is the sum of the masses of particles inside divided by the volume; this can lead to numerical discontinuities in the density, as real fluid packets would overlap cell boundaries, rather than lying completely to one side or another. Ramses instead uses a *cloud-in-cell* (CIC) scheme, where each particle distributes its mass uniformly over a cube centred on the particle, with the same width as the host cell Δx. This prevents the density in a cell changing discontinuously as particles move through the boundary. Formally, the density in cell i with centre $\mathbf{x} = (x, y, z)$ is the sum over the masses of all N particles, subject to a window function W that describes how the mass in a particle is distributed

$$\rho_i = \frac{1}{(\Delta x)^3} \sum_p^N m_p W(x_p - x) W(y_p - y) W(z_p - z), \tag{2.5}$$

where the window function in the NGP scheme is

$$W^{\mathrm{NGP}}(x - x_p) = \begin{cases} 1 & |x - x_p| \leq \Delta x/2, \\ 0 & \text{otherwise}, \end{cases} \tag{2.6}$$

and the CIC window function is

$$W^{\mathrm{CIC}}(x - x_p) = \begin{cases} 1 - |x - x_p|/\Delta x & |x - x_p| \leq \Delta x, \\ 0 & \text{otherwise.} \end{cases} \tag{2.7}$$

With these window functions, it is clear that the sum over all particles in ρ_i reduces to a sum over only particles in the host and neighbouring cells (or indeed just the host cell with NGP), as the window function is zero if the separation exceeds the cell width Δx. Higher order window functions can be constructed via successive convolutions of the CIC window function, that move the discontinuity in the density contribution of a particle into even higher spatial derivatives, beginning with the Triangular-Shaped-Cloud (discontinuous in the 2$^{\mathrm{nd}}$ derivative) that covers up to 3 cells in length, allowing continuity of the first derivative (in time) of the density as particles move across the cell boundary.

To solve the Poisson equation numerically, the density and the potential must both be discretised. This has been achieved by representing the gravitational and density fields as a grid of cells, and using a CIC treatment of particles to compute the density. However, we must still choose a finite difference approximation for the Laplacian to turn the Poisson PDE into a purely discrete matrix equation. Ramses uses the 2$^{\mathrm{nd}}$-order stencil

$$\nabla^2 \phi \approx \frac{\phi_{i+1} - 2\phi_i + \phi_{i-1}}{(\Delta x)^2}, \tag{2.8}$$

where subscripts $i+1$ and $i-1$ label cells adjacent to cell i on the x axis, and the second and third spatial dimensions have been omitted. Thus, the Poisson equation is now entirely discrete, and can be expressed in the standard form

$$\mathbf{A}\vec{\phi} = \vec{\rho}, \tag{2.9}$$

where the potential and the density in each cell is represented in the elements of the vectors $\vec{\phi}$ and $\vec{\rho}$ respectively, and the differential operator has become the matrix $\mathbf{A}$. This can be solved via standard techniques, for example by direct matrix inversion (which would usually be far too expensive), or relaxation techniques, the simplest of which is *Jacobi iteration*, where the solution is obtained by successive iteration

$$\vec{\phi}^{n+1} = \mathbf{D}^{-1}(\vec{\rho} - (\mathbf{A} - \mathbf{D})\vec{\phi}^{n}), \tag{2.10}$$

(where $\mathbf{D}$ is the matrix containing only the diagonal elements of $\mathbf{A}$), which tends to outperform direct inversion in terms of computational expense. This is improved upon by the *Gauss-Seidel* method, which uses the decomposition of $\mathbf{A}$ into a lower triangular part $\mathbf{L}_* = \mathbf{L} + \mathbf{D}$ (including the diagonal) and strictly upper triangular (zero diagonal elements) $\mathbf{U}$ (such that $\mathbf{A} = \mathbf{L} + \mathbf{D} + \mathbf{U}$) in the iteration

$$\vec{\phi}^{n+1} = \mathbf{L}_*^{-1}(\vec{\rho} - \mathbf{U}\vec{\phi}^{n}). \tag{2.11}$$

This not only converges in fewer steps, but allows space-efficient in-place computation of the elements of $\vec{\phi}^{n+1}$ by forward substitution (possible due to the lower-triangular nature of $\mathbf{L}_*$), so only one instance of $\vec{\phi}$ needs to be stored in memory. The Gauss-Seidel method is a specific case of *successive over-relaxation* (SOR)

$$\vec{\phi}^{n+1} = (\mathbf{D} + \omega\mathbf{L})^{-1}(\omega\vec{\rho} - [\omega\mathbf{U} + (\omega - 1)\mathbf{D}]\vec{\phi}^{n}), \tag{2.12}$$

in which $1 < \omega < 2$ is the *relaxation factor* ($\omega = 1$ returns the Gauss-Seidel method). For some problems, SOR with $\omega \neq 1$ improves further on the convergence speed of Gauss-Seidel, where the choice of ω is problem dependent. Press et al. (1992) and Teyssier (2002) recommend a value of

$$\omega = \frac{2}{1 + \frac{\pi\alpha}{\langle L \rangle}}, \tag{2.13}$$

where $\alpha = 1, 2$ for Dirichlet and periodic boundary conditions respectively, and $\langle L \rangle$ is the typical number of cells in a patch[2] of the mesh (empirically found to be ∼20 in cosmological simulations by Teyssier (2002)). The tree-like mesh structure in Ramses has a very high dynamic range, i.e. the size of its cells span many orders of magnitude. This presents a problem for relaxation solvers, as the speed of convergence is slow for large scales on fine grids. This is the motivation for the *multigrid* method, in which the mesh is *restricted* into

[2] A 'patch' is a region of adjacent cells at the same refinement level.

coarser grids (by integrating over cells) at various scales, solutions are computed on these using SOR (or any similar relaxation method), and then *interpolated* back onto the original mesh. This allows multiple scales to converge simulataneously. Ramses uses a multigrid method (Guillet & Teyssier, 2011) in cosmological simulations with periodic boundaries, reaching sub-percent level accuracy in just a few cycles for typical problems.

Other techniques exist for solving Equation 2.9, for example, the *conjugate gradient* (CG) method (Hestenes et al., 1952), which Ramses also provides (though it is now obsoleted by the multigrid solver), making use of the fact that the solution is the unique minimum of the function

$$f(\vec{\phi}) = \frac{1}{2}\vec{\phi}^T \mathbf{A}\vec{\phi} - \vec{\phi}^T\vec{\rho}. \tag{2.14}$$

This can be seen by taking the gradient of $f(\vec{\phi})$:

$$\vec{\nabla}_\phi f(\vec{\phi}) = \mathbf{A}\vec{\phi} - \vec{\rho}, \tag{2.15}$$

which is equal to zero if $\vec{\phi}$ is a solution to Equation 2.9. Thus, the solution can be found by iteratively minimising $f(\vec{\phi})$. Rather than by an intuitive gradient descent, this minimisation is carried out by constructing the solution out of a successively growing orthogonal basis of vectors (of maximum size N – the number of columns/rows in $\mathbf{A}$, putting an upper bound on the number of iterations needed to find the solution). Though the CG method converges quickly for general linear problems, it is found to be significantly less performant than the multigrid method for the high dynamic range Poisson problems in cosmological simulations (Guillet & Teyssier, 2011), as it suffers from the same problem as SOR in the slow convergence of large scale components of the solution. However, unlike the multigrid method, it is possible to use the CG solver in non-periodic problems (though cosmological simulations are periodic, so this is not usually necessary).

Having solved the Poisson equation (which Ramses does using the multigrid method in cosmological simulations), all that remains is to update the position $\vec{x}$ and velocity $\vec{v}$ of particles. Ramses uses a *predictor-corrector* scheme to do this:

$$\vec{v}^{n+1/2} = \vec{v}^{n} + \vec{F}^{n}\frac{\Delta t}{2}, \tag{2.16}$$

$$\vec{x}^{n+1} = \vec{x}^{n} + \vec{v}^{n+1/2}\Delta t, \tag{2.17}$$

$$\vec{v}^{n+1} = \vec{v}^{n+1/2} + \vec{F}^{n+1}\frac{\Delta t}{2}, \tag{2.18}$$

in which the last ('corrector') step requires that the forward-in-time force $\vec{F}^{n+1}$ be calculated. Since the gravitational force depends only on the positions of particles (and gas), not the velocities, this corrector step can be delayed until the next time step, after the new positions have been calculated, thus efficiently avoiding a costly inversion of the system.

2.1.1.4 P^{3}M and Tree-PM

In the P^{3}M method (Hockney et al., 1973), forces on these particles are calculated differently depending on their distance from the source. Short range forces are calculated directly between particles, and long range forces are found by solving the Poisson equation and taking the gradient of the potential, as in the PM method above. As masses in the same cell do not contribute to the gradient of the potential in the cell, the resultant force on a particle i is a simple sum of the long-range component (captured in the field) and the local pairwise interactions

$$\vec{F}_i = -m_i\vec{\nabla}\phi + \sum_{j\neq i}^{\text{local parts.}} \frac{G m_i m_j (\vec{x}_i - \vec{x}_j)}{(||\vec{x}_i - \vec{x}_j||^2 + \epsilon^2)^{3/2}}. \tag{2.19}$$

This affords greater accuracy of short range interactions than in pure PM, and allows the use of a lower resolution mesh, at potentially greatly increased cost (if many particles occupy the same mesh cell). This cost can be mitigated by applying a Barnes-Hut-like tree algorithm inside each cell to compute the short-range term (tree-PM).

The cosmological hydrodynamics codes Gadget-2 (Springel, 2005) and Arepo (Weinberger et al., 2020) make use of combined tree-PM algorithms to calculate gravitational forces. These are both Lagrangian codes, so, unlike Ramses, they do not construct a high-resolution structured mesh for their hydrodynamics, that could otherwise be reused to

solve the Poisson equation with the PM method. Ramses already constructs a highly-refined mesh for its Eulerian hydrodynamics, that is re-used by the gravitational calculations, so the tree-PM technique is less applicable — short range forces are already better-resolved by a higher resolution mesh.

2.1.2 Hydrodynamics

To treat astrophysical gas, Ramses solves the Euler equations, a special case of the Navier-Stokes equations of fluid dynamics where there is no viscosity or conductivity. These are (Teyssier, 2002)

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot (\rho \vec{u}) = 0, \tag{2.20}$$

$$\frac{\partial (\rho \vec{u})}{\partial t} + \vec{\nabla} \cdot (\rho \vec{u} \otimes \vec{u}) = -\vec{\nabla} p - \rho \vec{\nabla} \phi, \tag{2.21}$$

$$\frac{\partial (\rho e)}{\partial t} + \vec{\nabla} \cdot ((\rho e + p) \vec{u}) = -\rho \vec{u} \cdot \vec{\nabla} \phi, \tag{2.22}$$

closed by the equation of state of an ideal gas (with adiabatic index γ)

$$p = (\gamma - 1)\rho \left(e - \frac{u^2}{2} \right), \tag{2.23}$$

where $\vec{u}$ is the gas velocity, ρ is the gas density (not total), p is the thermal pressure, e is the specific total energy, and ϕ is the total gravitational potential (due to dark matter, stars and gas). In one spatial dimension, ignoring gravity, these can be written in generalised form as

$$\begin{gathered} \frac{\partial Q}{\partial t} + \frac{\partial \mathcal{F}}{\partial x} = 0, \\ Q = \begin{bmatrix} \rho \\ \rho u \\ \rho e \end{bmatrix}, \quad \mathcal{F} = \begin{bmatrix} \rho u \\ \rho u^2 + p \\ u(\rho e + p) \end{bmatrix}. \end{gathered} \tag{2.24}$$

Then, by applying the chain rule $\frac{\partial \mathcal{F}_i}{\partial x} = \frac{\partial \mathcal{F}_i}{\partial Q_j} \frac{\partial Q_j}{\partial x}$ (with Einstein notation for summing

repeated indices) we have

$$\begin{aligned} \frac{\partial \boldsymbol{Q}}{\partial t} + \mathbf{J}\frac{\partial \boldsymbol{Q}}{\partial x} &= 0, \\ \mathbf{J}_{ij} &\equiv \frac{\partial \mathcal{F}_i}{\partial \boldsymbol{Q}_j}. \end{aligned} \tag{2.25}$$

The eigenvectors of the *Jacobian* matrix $\mathbf{J}$ then correspond to characteristic solutions of the Euler equations. Their eigenvalues, which give their velocities, are

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} u + \sqrt{\frac{\gamma p}{\rho}} \\ u \\ u - \sqrt{\frac{\gamma p}{\rho}} \end{bmatrix}, \tag{2.26}$$

which are the velocities of forward travelling sound waves, the bulk of the fluid itself, and backward sound waves respectively. This could be solved by a method of finite differences, similarly to the way particle states are updated, but this can have problems with mass/momentum/energy conservation, and capturing discontinuities. Ramses uses an alternative: the *finite volume method*, proposed by Godunov & Bohachevsky (1959). Consider the integral form of Equation 2.24 (over a single cell with size Δx and one timestep Δt)

$$\int_{\Delta x}\int_{\Delta t}\left(\frac{\partial \boldsymbol{Q}}{\partial t} + \frac{\partial \mathcal{F}}{\partial x}\right) dx dt = 0, \tag{2.27}$$

which we can apply the divergence theorem to, giving an expression in terms of integrals over time at the cell 'faces' (just two points in 1D) and integrals over the entire cell at the beginning and end of the step:

$$\int_{\Delta x} \boldsymbol{Q}\left(x, t^n\right) dx - \int_{\Delta x} \boldsymbol{Q}\left(x, t^{n+1}\right) dx + \int_{\Delta t} \mathcal{F}\left(x_{i-1/2}, t\right) dt - \int_{\Delta t} \mathcal{F}\left(x_{i+1/2}, t\right) dt = 0 \tag{2.28}$$

where n denotes the timestep number, and $x_{i-1/2}$ and $x_{i+1/2}$ are the positions of the left and right faces of the 1D cell respectively. Each cell of the mesh has an average gas state associated with it (rather than the state at the cell centre, as with the gravitational

potential), which is simply related to the spatial integrals above by dividing by the cell volume. This expression for the change in a cell's state in terms of time integrals of *fluxes* enables a simple, conservative numerical method for solving the system. Using a simple *forward Euler* time integration scheme, we have

$$\mathcal{Q}^{n+1} = \mathcal{Q}^n - \frac{\Delta t}{\Delta x}\left(\mathcal{F}_{i+1/2} - \mathcal{F}_{i-1/2}\right), \tag{2.29}$$

where $\mathcal{Q}^n = \int_{\Delta x} \mathcal{Q}\left(x, t^n\right) dx$, but the fluxes remain to be calculated. We cannot directly (without approximation) use the average state of the cell; we require $\mathcal{F}$ on the cell boundary: the *inter-cell flux* (averaged over time). Since $\mathcal{F}$ can be expressed as a function of $\mathcal{Q}$, we could ideally calculate $\mathcal{F}_{i\pm1/2}$ by knowing the time-averaged states at the cell faces $\mathcal{Q}_{L,R}$, but this is not explicitly stored or calculated. Godunov suggests a simple scheme in which the time-averaged inter-cell flux is the flux given by the instantaneous solution to the *Riemann Problem* between the two cells sharing the face at which the flux is defined

$$\mathcal{F}_{i+1/2} = \mathcal{F}\left(\mathcal{Q}_{i+1/2}^{n+1/2}\right), \tag{2.30}$$

where $\mathcal{Q}_{i+1/2}^{n+1/2} \approx \mathcal{R}\left[\mathcal{Q}_i, \mathcal{Q}_{i+1}\right]$ is the solution to the Riemann problem (a hydrodynamic system, initialised with two uniform states separated by a discontinuity) between the left and right cell states, at the interface between them. The flux is then computed directly from this solution to the Riemann problem, since $\mathcal{F}$ can be written as a function of $\mathcal{Q}$ at the same position and time. It must be stressed that, even with an exact Riemann solution (which is possible, but not cheap, to obtain), this is an approximation, and the resulting scheme is first-order at best. The time-averaged inter-cell flux is only given exactly by the solution to this Riemann problem when the conditions at the beginning of the step are truly piecewise uniform, and there are no other waves causally connected to the interface during the averaging period (originating earlier in time, from neighbouring cells).

2.1.2.1 The Riemann Problem

The *Riemann problem* (two regions of uniform state separated by a discontinuity; Toro 2009) is fundamental to solving hydrodynamics in Godunov's finite volume discretisation.

Formally, the Riemann problem is the initial value problem of Equation 2.24 (the Euler equations), with piecewise-constant initial conditions (around an interface at $x = 0$)

$$Q(x, t = 0) = \begin{cases} Q_{\mathrm{L}} & x < 0, \\ Q_{\mathrm{R}} & x > 0. \end{cases} \tag{2.31}$$

The solution to the problem can be found by considering the propagation of the characteristics of the Euler equations, and takes different forms depending on the initial conditions, featuring some, but not necessarily all, of the following key features (as seen in Figure 2.2):

- The **rarefaction wave** consists of a region where the wave speed increases from front to back (in the direction of travel of the wave), reducing the density within the region. This is a superposition of acoustic waves, each moving at the local speed of sound.

- The **contact discontinuity** separates regions of equal pressure, but different density and temperature (and specific entropy). This travels at the resultant flow velocity in the region between the outward travelling waves (conventionally called the "star" region in literature).

- The **shock** is a compressive wave. Here the characteristics converge into a single front, rather than spreading out as they do in the rarefaction wave. Notably, shocks are not well-described by the Euler equations alone; solutions are not single-valued across discontinuities without imposing an 'entropy condition' that forbids local characteristics ahead of the shock from travelling faster than the shock. Thus, shocks break the time-reversibility of the pure Euler equations. On the microscopic scale of the shock front (on the order of the kinetic mean free path), viscosity and heat conduction must be invoked to fully describe the dynamics, even in conditions where these are irrelevant on the macro-scale (Zeldovich & Raizer, 1967). Solutions to the Riemann problem need not consider the microscopic dynamics of shocks, but instead use thermodynamic arguments to infer the effect on the medium and the speed of propagation of the discontinuity. As the fluid behind the shock front cannot causally interact with the fluid ahead of it, the state of the shocked fluid is purely

a function of the unshocked fluid state and the shock speed. In the frame comoving with the shock, the dependence on the shock speed necessarily disappears, and the relationship between the unshocked state (subscripted by 1) and the shocked state (subscripted by 2) is a statement of mass, momentum and energy conservation (i.e. that the generalised flux $\mathcal{F}$ from Eq. 2.24 obeys $\Delta\mathcal{F} = 0$ across the shock)

$$\rho_1 u_1 = \rho_2 u_2, \tag{2.32}$$

$$\rho_1 u_1^2 + p_1 = \rho_2 u_2^2 + p_2, \tag{2.33}$$

$$u_1(\rho_1 e_1 + p_1) = u_2(\rho_2 e_2 + p_2). \tag{2.34}$$

These are the *Rankine-Hugoniot* jump conditions, which can be used to derive the shock speed (by transforming back to the grid frame) and state behind the shock.

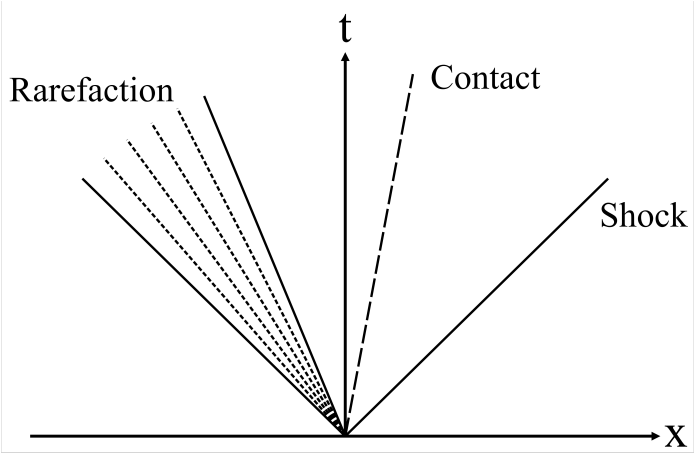


Figure 2.2: *The Riemann 'Fan':* an example of a possible solution to a Riemann problem, exhibiting the key features. Lines indicate the spacetime trajectory of waves in the gas. *Left:* the *rarefaction wave* spreads out as its characteristics have increasing speed from right to left. *Middle:* the *contact discontinuity* (or entropy wave) separates the post-rarefication and post-shock gas. *Right:* the *shock* (or compression wave) is a sharp front, as the characteristics have decreasing speed from left to right, so they converge.

General solutions for the Riemann problem must be found numerically. Various methods exist to do this approximately at the cell interface (the t-axis, in Figure 2.2), including:

- The Roe solver (Roe, 1997) linearises the Euler equations, solving the resulting problem analytically. The solutions then correspond exactly to the eigenvectors of the Jacobian (in Equation 2.25), which are simply forward/backward sound waves and the flow itself. The contact discontinuity, and the non-linear behaviour of shocks and rarefactions (where waves of different amplitudes travel at different speeds), are not captured.

- The Harten-Lax-van Leer (HLL) solver (Harten et al., 1983) assumes both the left and right waves always take the same symmetric form, travelling at a single speed in either direction, and that the state between the two waves is constant (so allows no contact discontinuity). Harten et al. (1983) compute these speeds by assuming that both waves are rarefactions, and that the gas dynamics are isentropic, but different choices for the speeds can be made (e.g. as in Toro et al. 1994, to better differentiate between shocks and rarefactions).

- The Harten-Lax-van Leer-Contact (HLLC) solver (Toro et al., 1994) improves on HLL by additionally tracking the contact discontinuity. This results in more accurate behaviour in some circumstances, for example around discontinuities that would not lead to shocks (where the densities in both regions may be discontinuous, but not the pressure). HLLC is an approximate three-wave model, and does not resolve the entire continuous 'fan' of characteristic waves, but still captures the key phenomena whilst enabling efficient computation. This is the default recommended Riemann solver in Ramses, and is used throughout this thesis.

Methods for numerically solving the exact Riemann problem can also be used (e.g. Godunov & Bohachevsky 1959 propose an iterative scheme for this, in addition to their finite volume method for hydrodynamics), which are less numerically diffusive than HLLC, but are conversely more prone to numerical oscillations and instabilities, and are more computationally expensive.

2.1.2.2 Numerical Stability

The choice of spatial discretisation and Riemann solver is a compromise between accuracy and numerical diffusion. Though numerical diffusion undesirably decreases solution accuracy, it is shown in this section that some amount of diffusion is necessary to ensure stability in schemes for hydrodynamics.

Consider a simplified advection equation

$$\frac{\partial Q}{\partial t} + a\frac{\partial Q}{\partial x} = 0, \tag{2.35}$$

where a is the dimensionless wave speed, and we choose $a > 0$. This could be solved by the following Forward Time Centred Space (FTCS) finite-difference scheme

$$\frac{Q_i^{n+1} - Q_i^n}{\Delta t} + a\frac{Q_{i+1}^n - Q_{i-1}^n}{2\Delta x} = 0. \tag{2.36}$$

The stability of this scheme can be determined using the standard method of *von Neumann analysis* (von Neumann & Richtmyer, 1947). On a uniform 1D grid of N points, the numerical solution can be represented by a truncated Fourier series (as a finite grid can only represent N distinct frequencies)

$$Q_i^n = \sum_{m=0}^{N} \hat{Q}_m^n e^{ik_m x}, \tag{2.37}$$

where $k_m = \frac{\pi\Delta x}{m}$, and $\hat{Q}_m^n$ are constant Fourier coefficients. Without loss of generality, we can consider the evolution of a solution composed of a single mode $Q_i^n = \hat{Q}_m^n e^{ik_m x}$

$$Q_i^{n+1} = \hat{Q}_m^n e^{ik_m x} - \frac{a\Delta t}{2\Delta x}\hat{Q}_m^n e^{ik_m x}\left(e^{ik_m\Delta x} - e^{-ik_m\Delta x}\right), \tag{2.38}$$

in which we see

$$Q_i^{n+1} = \left(1 - \frac{a\Delta t}{\Delta x} i\sin(k_m\Delta x)\right) Q_i^n, \tag{2.39}$$

where the prefactor of Q_i^n is called the *growth factor* $g(k_m)$. If $|g(k_m)|^2 > 1$, modes with wavenumber k_m grow indefinitely, and the scheme is unstable. We have

$$|g(k_m)|^2 = 1 + \left(\frac{a\Delta t}{\Delta x}\right)^2 \sin^2(k_m \Delta x) \; > 1, \tag{2.40}$$

so the FTCS scheme for our advection equation is unstable for every non-trivial mode $m > 0$, and so is always unstable for any spatially-varying initial conditions. Conversely, the *upwind* scheme

$$\frac{Q_i^{n+1} - Q_i^n}{\Delta t} + a\frac{Q_i^n - Q_{i-1}^n}{\Delta x} = 0, \tag{2.41}$$

that computes derivatives only using cells *opposite to the direction of the flow* (recalling that we have $a > 0$), has growth factor

$$g(k_m) = 1 - \frac{a\Delta t}{\Delta x}\left(1 - e^{ik_m \Delta x}\right), \tag{2.42}$$

which has modulus < 1 only if

$$\frac{a\Delta t}{\Delta x} < 1. \tag{2.43}$$

This is the *Courant-Friedrichs-Lewy* (CFL) condition, that says that the scheme is unstable if information is numerically propagated faster than the physical wave speed a. Though it has been derived only for this scheme, any explicit scheme for solving hyperbolic PDEs (like the scheme described above for solving the Euler equations) must obey this condition, since it is a fundamental statement of causality. The FTCS scheme described earlier in this section is not upwind — it uses information ahead of the flow — and its instability can be understood to come from its violation of causality.

Another way to write the upwind scheme is

$$\frac{Q_i^{n+1} - Q_i^n}{\Delta t} + a\frac{Q_{i+1}^n - Q_{i-1}^n}{2\Delta x} = \frac{a\Delta x}{2}\frac{Q_{i+1}^n - 2Q_i^n + Q_{i-1}^n}{\Delta x^2}, \tag{2.44}$$

where the term on the RHS is an approximation to the second derivative $\frac{a\Delta x}{2}\frac{\partial^2 Q}{\partial x^2}$. Thus, this is equivalent to the FTCS scheme with a *diffusive* source term, that grows proportionally to the wave speed.

Schemes for hydrodynamics must resolve multiple waves simultaneously (flow, acoustic waves, shocks, rarefactions), and so cannot avoid using information ahead of the flow of slower waves to accurately resolve the fastest (or equivalently, satisfying causality for all waves leaves the fastest sub-optimally resolved, and thus more diffusive), and so stability cannot be guaranteed without some amount of numerical diffusion. In general, schemes can be stabilised by adding numerical diffusion to cancel out non-causal terms ahead of the flow.

2.1.2.3 Higher-Order Methods

Godunov's first-order scheme for hydrodynamics (defined by Equations 2.29 and 2.30) seems undesirable at first, since first-order schemes are usually outperformed by higher-order schemes at a given computational cost. However, Godunov's theorem (Godunov & Bohachevsky, 1959) states that a linear (meaning that the new state is a linear function of the state at the previous step), *monotone* scheme (one that does not generate new extrema) cannot be more than first-order accurate. According to *Harten's Theorem* (Yee et al., 1985), only monotone schemes can be *total variation diminishing* (TVD), meaning that erroneous oscillations do not grow. Thus, a higher-order scheme for hydrodynamics must be non-linear in order to be useful.

Nonetheless, such schemes can be constructed, and achieve higher accuracy than Godunov's method in many situations, and so are considered useful (indeed, usually much more so than Godunov's original scheme). A classic example of these, serving as the inspiration for many that followed, is the *Monotonic Upstream-centered Scheme for Conservation Laws* (MUSCL; van Leer, 1979; Toro, 2009), which approximates $\boldsymbol{Q}$ as piecewise linear, so that the state in a cell is given by

$$\widetilde{\boldsymbol{Q}}_i(x) = \boldsymbol{Q}_i + (x - x_i)\frac{\Delta_i}{\Delta x}, \tag{2.45}$$

where Δ_i is the *slope* (remaining to be chosen), and x_i is the cell centre. Left and right boundary values can be extrapolated in each cell i

$$Q_{i,L} = Q_i - \frac{\Delta_i}{2} \quad ; \quad Q_{i,R} = Q_{i+1} + \frac{\Delta_i}{2}. \tag{2.46}$$

A discontinuity may exist at the cell face $i+1/2$, between the right extrapolated boundary value from the left cell and the left value from the right cell (i.e. $Q_{i,R} = Q_{i+1,L}$ is not guaranteed). One choice for the slope is a weighted average of the differences between the neighbouring states and Q_i

$$\Delta_i = \frac{1}{2}(1+\omega)(Q_i - Q_{i-1}) + \frac{1}{2}(1-\omega)(Q_{i+1} - Q_i), \tag{2.47}$$

where $\omega \in [-1, 1]$ is an arbitrary parameter that controls the weighting between the left and right states. Regardless of the chosen value of ω, this slope must be *limited* in order for the scheme to be TVD, necessarily introducing non-linearity (and a potentially controversial choice, since the presence of the limiter is not well-justified mathematically). A commonly used limiter (and the one used in this work) is the *MinMod* limiter (shown here with $\omega = 0$)

$$\Delta_i = \text{MinMod}\left(Q_i - Q_{i-1}, Q_{i+1} - Q_i\right), \tag{2.48}$$

$$\text{MinMod}(a, b) \equiv \begin{cases} a & |a| < |b| \;\&\; ab > 0, \\ b & |a| > |b| \;\&\; ab > 0, \\ 0 & ab \leq 0, \end{cases} \tag{2.49}$$

which selects the smallest of either slope at a cell boundary (independently for each component of the slopes), and limits the slope to zero where a positive slope neighbours a negative. Many other limiters can be used, that each make a different degree of compromise between numerical stability and accuracy (where the most aggressive slope limiters lead to more diffusive solutions, but less aggressive limiters may allow numerical oscillations). For example, *MinMod* is more diffusive than the *MonCen* limiter, but guarantees that the conservative variables Q are always non-negative.

The system may be evolved in time by, for example, a *predictor-corrector* method [3] (Toro, 2009), as follows:

1. **The Predictor Step:** The boundary value states at each interface are evolved in time by a half step $\frac{\Delta t}{2}$ with a Taylor expansion (making them *time-centered*)

$$Q_{i,L}^{n+1/2} = Q_{i,L}^{n} + \frac{\Delta t}{2}\frac{\partial \mathcal{F}}{\partial x}, \tag{2.50}$$

$$Q_{i,R}^{n+1/2} = Q_{i,R}^{n} + \frac{\Delta t}{2}\frac{\partial \mathcal{F}}{\partial x}, \tag{2.51}$$

   where $\frac{\partial \mathcal{F}}{\partial x}$ is approximated by the centred difference

$$\frac{\partial \mathcal{F}}{\partial x} \approx \frac{\mathcal{F}\left(Q_{i,L}^{n}\right) - \mathcal{F}\left(Q_{i,R}^{n}\right)}{\Delta x}. \tag{2.52}$$

2. **Solve the Riemann Problem:** The previous step returns two predictions for the state at each face — one from the left, one from the right. To compute $Q_{i+1/2}^{n+1/2}$ uniquely, the Riemann problem is solved at the face between the extrapolated boundary values found in step 1 (where the 'left' boundary value now comes from the cell to the right, as both values are taken at the same face)

$$Q_{i+1/2}^{n+1/2} \approx \mathcal{R}\left[Q_{i,R}^{n+1/2}, Q_{i+1,L}^{n+1/2}\right], \tag{2.53}$$

   which can be used to compute the time-centred inter-cell flux

$$\mathcal{F}_{i+1/2}^{n+1/2} = \mathcal{F}\left(Q_{i+1/2}^{n+1/2}\right). \tag{2.54}$$

3. **The Corrector Step:** Finally, the time-centred flux $\mathcal{F}_{i+1/2}^{n+1/2}$ is used to evolve the state

[3] Other methods can be used, such as Runge-Kutta schemes. The characterising property of the MUSCL scheme is that the inter-cell flux is the solution to a Riemann problem between piecewise-linear reconstructed cell states.

$$Q_i^{n+1} = Q_i^n - \frac{\Delta t}{\Delta x}\left(\mathcal{F}_{i+1/2}^{n+1/2} - \mathcal{F}_{i-1/2}^{n+1/2}\right). \tag{2.55}$$

This scheme is second order in space and time, and so is more efficient (in terms of computational cost) than Godunov's first-order scheme at a given level of accuracy.

Ramses uses the piecewise linear method (PLM) of Colella (1990) and Saltzman (1994), a variant of the method above in which the predictor step is now done by *tracing characteristics* from t^{n+1} back to the linear reconstructed data at t^n. This only retains information from waves that are travelling into the cell face, discarding the non-causal terms.

2.1.3 Adaptive Mesh Refinement

Ramses uses an octree mesh: a tree where each node may either contain a cell (making the node a *leaf*), or eight sub-nodes, each of which can either contain a leaf cell half the size of the parent or another sub-node (see Figure 2.3). The octree is *graded*, which guarantees that if adjacent cells differ in size, it is only by a factor of two. This mesh can be modified on the fly, adding or removing cells where the simulation requires it.

Ramses provides a choice of how to initialise the values of new cells created by the splitting of a refined cell (prolongation), or by joining of de-refined cells (restriction). The work in this thesis uses the simplest scheme (performing no interpolation, but conserving mass, energy and momentum), in which newly-refined cells directly inherit the density, momentum and internal energy density of their predecessor, and de-refined cells have the sum of the mass, momenta and internal energy of their children.

There exist a variety of strategies to choose which cells to refine; a commonly used one is *quasi-Lagrangian* refinement, in which a cell is refined into 8 sub-cells if it contains mass (including stars and dark matter as well as gas) in excess of $8M_{DM}$ (eight times the mass of a dark matter particle), ensuring cells always contain within a factor 2^3 the particle mass, regardless of the refinement level, thus the cells may be thought of similarly to particles in Lagrangian codes — in that they contain approximately constant mass. Other mass thresholds may be chosen, but it is always ensured that resolution is increased in proportion to matter density. Another strategy is to ensure cells always resolve a fraction

$\frac{1}{n}$ of the Jeans length (Equation 1.38), which means Jeans unstable cells are refined before they fully collapse, so the fragmentation of gas by radiative cooling is resolved properly.

The differing sizes of cells mean that time steps Δt must be set differently depending on the refinement level, to ensure that the CFL condition (where the wave speed in Eq. 2.43 is now the speed of sound waves propagating through the fastest flowing cells relative to the grid) always holds as cell size Δx decreases. Other conditions are imposed on the timestep, independently for each level of refinement: Δt must not exceed a user-specified fraction $C_1 < 1$ of the gravitational free-fall time (ensuring accuracy as gas collapses under its own gravity); particles must only move a fraction $C_2 < 1$ of the size of their host cell during Δt (ensuring validity of the operator splitting of the gravitational step); and the cosmological expansion factor a_{exp} must not change by more than a fraction $C_3 < 1$ over Δt (also ensuring validity of operator splitting for the gravitational update). These conditions, expressed mathematically for cells at the same refinement level, are

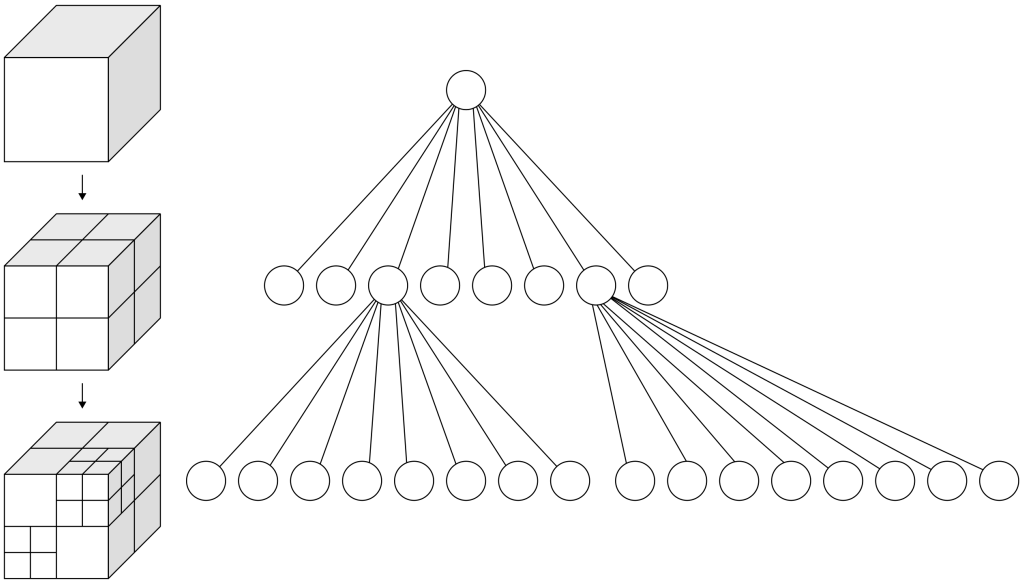


Figure 2.3: Illustration of a graded Octree mesh (left), and the graph structure of the associated nodes (right), with the root node at the top (Wikipedia contributors, 2023).

$$\Delta t_1 = C_1 \min(t_{\rm ff}) \; ; \quad t_{\rm ff} = \sqrt{\frac{3\pi}{32 G\rho}}, \tag{2.56}$$

$$\Delta t_2 = C_2 \frac{\Delta x}{\max(|\vec{v}_{DM}|)}, \tag{2.57}$$

$$\Delta t_3 = C_3 \frac{a_{\rm exp}}{\dot{a}_{\rm exp}}, \tag{2.58}$$

$$\Delta t_4 = \frac{a_{\rm CFL} \Delta x}{\max(|u_x| + c_s, |u_y| + c_s, |u_z| + c_s)}, \tag{2.59}$$

$$\Delta t \leq \min(\Delta t_1, \Delta t_2, \Delta t_3, \Delta t_4), \tag{2.60}$$

where $a_{\rm CFL} < 1$ is the *Courant factor*, which can be specified to reduce the timestep beyond the requirement of CFL condition if desired (this can be necessary for stability in practice). Finally, the time step between levels must be synchronised, so that the entire tree is eventually updated to arrive at the same time — at least before an output snapshot can be interpreted physically. One practical way to achieve this is to have cells in level $l + 1$ do twice as many steps (each with half the duration) as cells in the parent level l. This allows larger cells to avoid having to do many unnecessary steps, as their conditions are usually more easily satisfied by larger time steps than the smaller cells. Thus, the hierarchy of time steps is determined globally, such that the time step at level l is $\Delta t_l \propto 2^l$, chosen to be as large as possible whilst also satisfying the above conditions at every level.

2.1.4 Coordinates

When used for cosmological simulations, Ramses must account for the expansion of the universe. This is achieved through a change of variables (for a cubic domain of side length L at $a = 1$)

$$d\tilde{t} = H_0 \frac{dt}{a^2}, \tag{2.61}$$

$$\tilde{x} = \frac{x}{aL}, \tag{2.62}$$

$$\tilde{\rho} = \frac{a^3 \rho}{\Omega_m \rho_c}, \tag{2.63}$$

$$\tilde{p} = \frac{a^5 p}{\Omega_m \rho_c H_0^2 L^2}, \tag{2.64}$$

$$\tilde{\vec{u}} = \frac{a\vec{u}}{H_0 L}, \tag{2.65}$$

and a modified Poisson equation

$$\nabla^2 \tilde{\phi} = \frac{3a\Omega_m}{2}(\tilde{\rho} - 1). \tag{2.66}$$

These are called *super-comoving coordinates* (Martel & Shapiro, 1998), and leave the equations of hydrodynamics unchanged in form, as long as the adiabatic index is chosen to be $\gamma = 5/3$ (corresponding to that of atomic gas), which is the case for the work in this thesis. For smaller values of γ, other degrees of freedom than the three kinematic ones for atomic gas can contribute to the internal energy (for example, in molecular gas in LTE, some of the internal energy is found in the rotational and vibrational modes of molecules). Since these other degrees of freedom are unaffected by cosmological expansion (a molecule at $a = 1$ is no different to a molecule at $a = 0.1$), they transform differently under the change to super-comoving coordinates. This leads to a so-called 'drag' term in the energy equation that reduces the relative rate of change of the super-comoving internal energy under cosmological expansion as γ decreases below $5/3$.

2.1.5 Alternative Hydrodynamics Schemes

Ramses is an *Eulerian* hydrodynamics code: it represents the fluid as fields with respect to a fixed reference frame across the entire domain (cosmological expansion aside). The opposite description is *Lagrangian*, in which the fluid is represented by packets of constant mass that move exactly with the flow (and so their properties are described with respect

to the frame comoving with the flow). These packets may be referred to as 'particles', but do not behave as individual particles of the fluid would, but instead represent the motion of a large number of particles from the same region of space.

Described in this section are some alternative Lagrangian schemes to the Eulerian finite volume method of Ramses: *Smoothed-particle Hydrodynamics* (SPH) — as used in TREESPH (Hernquist & Katz, 1989) and Gadget (Springel et al., 2001; Springel, 2005) — and the 'hybrid' schemes of the Arepo (Springel, 2010; Weinberger et al., 2020) and Gizmo (Hopkins, 2015) codes.

2.1.5.1 Smoothed-particle Hydrodynamics

In SPH, 'smoothed' approximations to fields (subscripted by s), such as the density, are considered (Gingold & Monaghan, 1977)

$$\rho_s(\vec{x}) = \int W(\vec{x} - \vec{y})\, \rho(\vec{y})\, d^3\vec{y}, \tag{2.67}$$

where W is the a normalised *smoothing kernel*, satisfying

$$\int W(\vec{x})\, d^3\vec{x} = 1. \tag{2.68}$$

The smoothed field integral can be approximated by the Monte Carlo method, summing the contributions from N particles in the simulation (each with position $\vec{x}_i$). If each particle has the same mass $m = \frac{1}{N}\int \rho(\vec{y})\, d^3\vec{y}$, and they are distributed consistently with the density ρ, then the smoothed density field can be approximated by

$$\rho_s(\vec{x}) \approx \rho_N(\vec{x}) \equiv \frac{1}{N}\sum_{i=1}^{N} W(\vec{x} - \vec{x}_i) \int \rho(\vec{y})\, d^3\vec{y}. \tag{2.69}$$

That is, the approximate smoothed density at $\vec{x}$ is the particle mass, multiplied by the sum of the values of their smoothing kernels at $\vec{x}$–$\vec{x}_i$. Choosing a kernel so that $\frac{1}{N}\sum_{i=1}^{N} W(\vec{x} - \vec{x}_i) \rightarrow \int \delta^3(\vec{x} - \vec{z}) d^3\vec{z}$ as $N \rightarrow \infty$ ensures that ρ_N converges to the unsmoothed field as the number of particles increases

$$\lim_{N\to\infty} \rho_N(\vec{x}) = \int\int \rho(\vec{y})\,\delta^3(\vec{x}-\vec{z})\,d^3\vec{y}d^3\vec{z} = \rho(\vec{x})\,. \tag{2.70}$$

It is useful to choose W of the form

$$\frac{1}{h^3} f\left(\frac{||\vec{x}||}{h}\right), \tag{2.71}$$

where h is called the *smoothing length*. If $f(r)$ satisfies

$$|r^2 f(r)| \to 0 \text{ as } r \to \infty, \tag{2.72}$$

$$\int_0^\infty f(r) 4\pi r^2 dr = 1, \tag{2.73}$$

$$\int_0^\infty |f(r)| 4\pi r^2 dr < \infty, \tag{2.74}$$

$$\tag{2.75}$$

then W behaves like the Dirac delta in the limit $N \to \infty$ (as $h \to 0$) and Equation 2.70 is upheld. A simple example of a kernel that satisfies these conditions is the step function

$$W(\vec{x}) = \frac{3H\left(1 - \frac{||\vec{x}||}{h}\right)}{4\pi h^3}, \tag{2.76}$$

which is normalised, and only nonzero within the unit sphere centred on the origin.

In more sophisticated SPH formulations particles may each have different masses m_i and smoothing lengths h_i. An intuitive choice of smoothing length is one that ensures each particle has, on average, the same mass within its kernel volume $N_{\text{sph}}\bar{m}$ (the product of the number of neighbours N_{sph} with the average particle mass $\bar{m}$), so that random noise affects all particle interactions uniformly. That is, the smoothing length h_i must satisfy

$$\frac{4\pi}{3} h_i^3 \rho_i = N_{\text{sph}}\bar{m}, \tag{2.77}$$

(Springel, 2005) which must be solved on a per-particle basis (choosing a universal value for N_{sph}) by summing the mass around it to determine ρ_i

$$\rho_i \equiv \rho_N\left(\vec{x}_i\right) = \sum_{j=1}^{N} m_j W\left(\vec{x}_i - \vec{x}_j\right). \tag{2.78}$$

This is consistent with Equation 2.69 when each particle has the same mass.

Derivatives of fields can be calculated similarly to Equation 2.78, for example, the divergence of the velocity at particle i is a mass-weighted average of the gradients between particles

$$\left(\vec{\nabla}\cdot\vec{v}\right)_i = \frac{\sum_{j=1}^{N} m_j\left(\vec{v}_i - \vec{v}_j\right)\cdot\vec{\nabla}_i W\left(\vec{x}_i - \vec{x}_j\right)}{\sum_{j=1}^{N} m_j W\left(\vec{x}_i - \vec{x}_j\right)} = \frac{1}{\rho_i}\sum_{j=1}^{N} m_j\left(\vec{v}_i - \vec{v}_j\right)\cdot\vec{\nabla}_i W\left(\vec{x}_i - \vec{x}_j\right), \tag{2.79}$$

where $\vec{\nabla}_i$ is the gradient w.r.t. $\vec{x}_i$, and $\vec{v}_i$ is the particle velocity.

Unlike in the Eulerian formulation, the dynamics are specified in terms of *convective derivatives*, which give the rate of change of a quantity in the frame of reference of a particle moving with the flow

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \vec{u}\cdot\vec{\nabla}. \tag{2.80}$$

Then, the Euler equations are (Zeldovich & Raizer, 1967)

$$\frac{D\rho}{Dt} = -\rho\nabla\cdot\vec{v}, \tag{2.81}$$

$$\frac{D\vec{v}}{Dt} = \frac{\vec{\nabla}p}{\rho} - \vec{\nabla}\phi, \tag{2.82}$$

$$\frac{De}{Dt} = -\frac{p}{\rho}\vec{\nabla}\cdot\vec{v}, \tag{2.83}$$

in which the density equation becomes trivial, consistent with the idea that particles have constant mass. Continuing naïvely, the numerical acceleration of the particles could be taken directly from the second equation above

$$\vec{a}_i = \frac{\vec{\nabla}p}{\rho} - \vec{\nabla}\phi, \tag{2.84}$$

however, this fails to capture shocks. As discussed previously, shocks either require resolution of dynamics on an untenably-small scale (with realistic viscosity and conduction), or explicit tracking and special treatment of discontinuities (as done by Riemann solvers), neither of which are present so far. However, Von Neumann & Richtmyer (1950) originated a method to smooth out the shock front across a few particles (or vertices, or cells, depending on the discretisation scheme used), called *artificial viscosity*, thus allowing the use of a scheme that does not treat discontinuities. Many possible forms for this artificial visocisty exist, but an accurate formulation will always ensure that the viscosity only affects the flow within shocks, and that the Rankine-Hugoniot conditions hold on scales above the artificial shock thickness. Von Neumann & Richtmyer (1950) achieve this through a term

$$q = \begin{cases} c^2 \rho \Delta x^2 \left(\frac{\partial u}{\partial x}\right)^2 & \frac{\partial u}{\partial x} < 0, \\ 0 & \text{otherwise}, \end{cases} \tag{2.85}$$

(shown in one dimensional form here) that is added to the pressure, where c^2 is a dimensionless constant close to unity, and the term is non-zero only for compressive flow (introducing the time irreversibility inherent in physical shocks). GADGET employs the extended form of Monaghan & Gingold (1983), that additionally includes a bulk viscosity, proportional to $\nabla \cdot u$, that damps oscillation on the sub smoothing-length scale, particularly necessary in SPH due to the random nature of N-particle motion that means particles are often very close to one another (as opposed to e.g. Lagrangian mesh codes, as considered by von Neumann).

There are advantages and disadvantages to SPH compared to the Eulerian finite volume method.

- Due to its use of the Lagrangian representation, SPH advects steady flows very well. Particles will only deviate from their straight line trajectory if a force acts on them; unlike with the FVM, there is little numerical dissipation of density or momentum in shock-free flows, and both linear and angular momentum are well-conserved. Angular momentum is particularly challenging to conserve in Eulerian

codes with cubic grids, in which a numerical torque acts to align systems with the grid directions (Hahn et al., 2010).

- As SPH is fundamentally a Monte Carlo method, it is subject to random noise, that can only be overcome by increasing the particle density, thus increasing the number of neighbours in force calculations and samples in reproduction of the fields. The errors introduced by this noise can seed instabilities, whereas Eulerian schemes could otherwise damp erroneous oscillations (by their TVD property). This places a greater burden on the artificial viscosity, to not only capture shocks, but also to damp numerical instabilities (achieved by the term $\propto \nabla \cdot u$ introduced above), leading to a more severely dissipative form that can diminish the advantage in angular/linear momentum conservation.

- SPH does not properly resolve Kelvin-Helmholtz (vortices arising from shear flow) and Rayleigh-Taylor (vortices occurring when a lighter fluid moves into a dense one) instabilities (Agertz et al., 2007). This is due to the way particles behave in steep density gradients, that leads to a poor treatment of contact discontinuities. Forces are estimated as a weighted sum of gradients of the smoothing kernel evaluated at the neighbours of a particle. So, in the case of sharp changes in the field over a shorter length scale than the smoothing length, gradients are either overestimated if many neighbours lie at the outer edge of the kernel, or underestimated if many neighbours are very close. A particle approaching a boundary between regions of different densities will tend to either be strongly repelled (if approaching from the rarefied side), or strongly pulled through (if approaching from the dense side), leading to a restorative resultant force (by momentum conservation) on the boundary that damps instabilities.

Despite fundamental differences between SPH and Eulerian methods, the properties of simulated galaxies are not necessarily very sensitive to the choice of hydrodynamics scheme (Vazza et al., 2011), and the choice of feedback models are of at least comparable importance (Scannapieco et al., 2012). This highlights the control feedback has over

the energy flow in galaxies, rather than diminishing the importance of accurate numerical hydrodynamics in general.

2.1.5.2 Moving Mesh Methods — Arepo

Purely Eulerian schemes lack *Galilean-invariance*, in that the error in the numerical solution is sensitive to bulk flow velocities (recalling from Section 2.1.2.2 that the numerical diffusion required to enable stability grows with the wave speed relative to the grid), even though the physical equations could be constructed and solved in the moving frame insensitive to a large-scale background flow (as in SPH). Previously the finite volume method has been considered on a static (though adaptively-refined) mesh, but it is possible to formulate the FVM on an arbitrarily-moving mesh in an *Arbitrary-Lagrangian-Eulerian* (ALE) method Hirt et al. (1974). In ALE schemes a *rezoning* step is performed, in which cells may be continuously deformed and/or moved (such that cell surfaces have a well-defined velocity), and hydrodynamic fields are updated taking into account their fluxes through the moving faces of the cell. This Galilean-invariance can be utilised to reduce numerical diffusion in high Mach number flows by allowing the mesh to move with the flow.

However, allowing deformations in the mesh introduces complications. If a mesh follows a rotating flow, it can become tangled, as in Figure 2.4, leading to complete failure of a simulation unless dealt with (cells must not overlap). This can be mitigated by careful set up of initial mesh structure for some problems, but in general the mesh must be occasionally discontinuously remapped, by interpolating the flow fields onto a new, more regular grid. This introduces numerical diffusion, and can reduce the accuracy advantage over pure Eulerian methods if performed too often or too severely. Schemes are also more computationally intense when translated from a structured mesh to an unstructured one. On a structured grid, the position of neighbouring cells can be found by integer arithmetic, but with an arbitrary mesh a spatial search into the grid data structure may be required.

In Arepo (Springel, 2010; Weinberger et al., 2020), the mesh is a *Voronoi tesselation* of the space (Figure 2.5 shows an example), in which the volume of a given cell is the set of points closest to its 'generating point' (which could have an arbitrary position). These

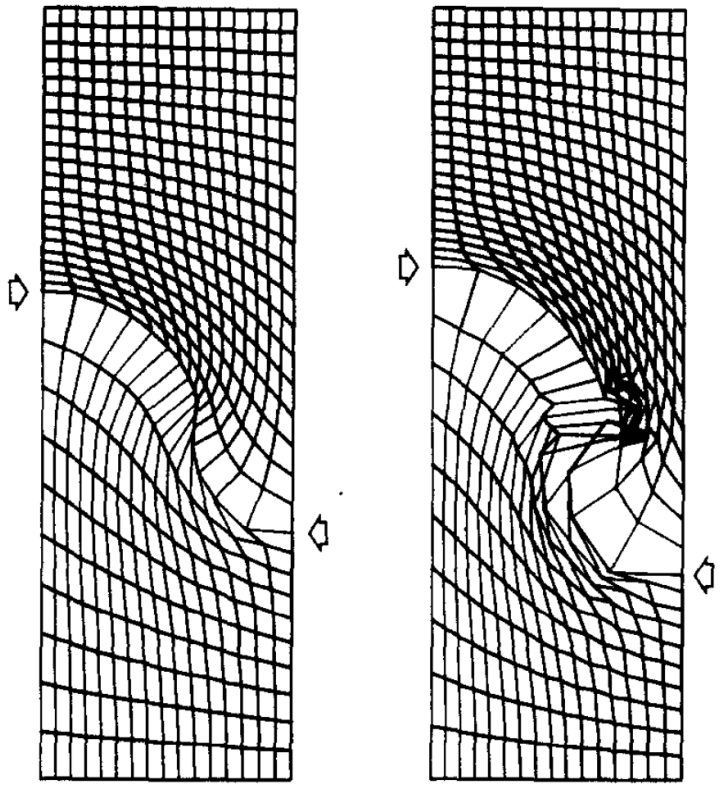


Figure 2.4: A Rayleigh-Taylor instability between two fluids simulated on an ALE mesh (Hirt et al., 1974) (evolving from **left** to **right**). The rotational flow has tangled the mesh to the point of failure.

generating points can be moved with the flow, like particles in a Lagrangian code. Unlike conventional ALE meshes, Arepo's Voronoi mesh is independently rebuilt each step around the generating points, leading to occasionally discontinuous changes in the *surface* (vertices, edges and faces may appear and disappear), but always ensuring a continuously deformed *volume*. This continuous deformation means that the flux of fields through the surface of a cell under the rezoning step can always be calculated, and so discontinuous remapping of the grid is never necessary.

Mathematically, Arepo's scheme closely resembles that of Ramses. If the mesh generating points are held static, instead of following the flow, the only difference between the schemes of the two codes is in the shape of the mesh. The Voronoi mesh's shape is more complicated to evaluate gradients on than a cubic grid, but still simpler than the fully-general case in unstructured ALE methods, since the Voronoi cell faces are guaranteed to be perpendicular to the line connecting cell generating points. Arepo, like Ramses, uses

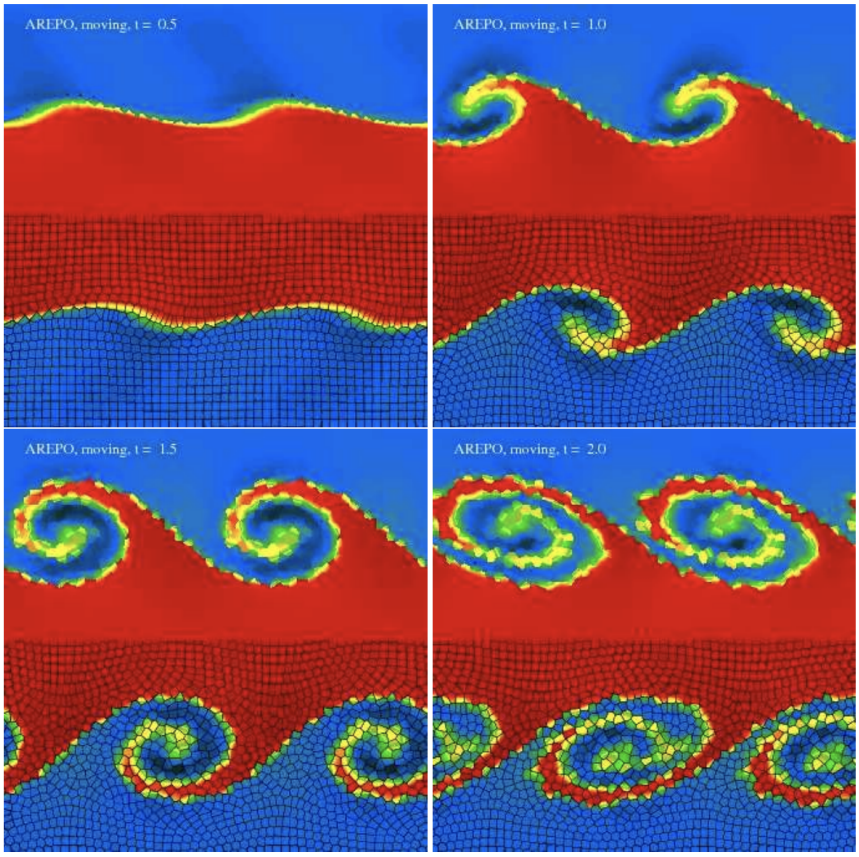


Figure 2.5: AREPO simulation of a Kelvin-Helmholtz instability (with the Voronoi mesh edges overlaid on the bottom half of each panel). Unlike with SPH, the instability is well-resolved here.

a linear reconstruction of fluid fields in a MUSCL scheme. The reconstruction is exact if the underlying field is linear, and the mesh generating point of each cell lies on its centre of mass.

The mesh generating point $\vec{r}_i$ of cell i obeys the equation of motion

$$\frac{d\vec{r}_i}{dt} = \vec{u}_i + \begin{cases} 0 & \frac{d_i}{\eta R_i} < 0.9, \\ c_i \frac{\vec{s}_i - \vec{r}_i}{d_i} \cdot \frac{d_i - 0.9\eta R_i}{0.2\eta R_i} & 0.9 \leq \frac{d_i}{\eta R_i} < 1.1, \\ c_i \frac{\vec{s}_i - \vec{r}_i}{d_i} & 1.1 \leq \frac{d_i}{\eta R_i}, \end{cases} \tag{2.86}$$

where $\vec{u}_i$ is the flow velocity in the cell, $\vec{s}_i$ is the centre of mass of the cell, $d_i = ||\vec{s}_i - \vec{r}_i||$, c_i is the local speed of sound, $R_i = \left(\frac{3V_i}{4\pi}\right)^{\frac{1}{3}}$ (the radius of an equivalent-volume sphere), and $\eta < 1$ is a dimensionless user-specified parameter. Without the second term on the RHS,

the equation is simply a specification of Lagrangian motion of the generating points. The purpose of the additional term is to *regularise* the mesh, bringing the generating points closer to the centres of mass and making cells more spherical, in order to reduce reconstruction error, and avoid bias towards particular regions or directions. This takes inspiration from Lloyd's Algorithm (Lloyd, 1982), which is designed to find an approximate *centroidal* Voronoi tesselation (in which the cell centres are also the generating points) by iteratively relaxing the generating points toward the cell centroids, regenerating the mesh each step. The smoothing term here accomplishes this over the course of a few time steps (as a new mesh is already generated each step) by moving the generating point toward the centre of mass if $d_i \lesssim \eta R_i$, in which η is the threshold for application of the smoothing (typically taken to be 0.25).

Similarly to Ramses, Arepo's mesh may be refined on the fly. This could be done by arbitrarily inserting new mesh generating points, regenerating the mesh, and then interpolating the fields from the old mesh onto the new one, but this would significantly increase numerical error around the newly introduced point. However, if a new mesh point is added infinitesimally close to another (or practically, as closely as they can be numerically represented without being identical), then the new pair of cells splits the old one exactly in half and together occupy the same volume with the same boundary as the old cell, leaving the rest of the mesh unchanged. De-refinement of cells is less elegant: a generating point is simply removed, and the mass, momentum and energy content of the removed cell is subsumed by its neighbours in proportion to the covering fraction of the old cell; in doing this, some error is introduced to the neighbours of the de-refined cell. This is a key difference with the AMR scheme of Ramses: though adding new cells is simple in Arepo, and refinement criteria can be applied unambiguously, de-refinement cannot be done without affecting neighbouring cells (also introducing error in the process), so the order of de-refinement may change the set of cells that are eventually removed. Though, if refinement is only applied to keep mass per-cell approximately constant as gas accretes (as in the quasi-Lagrangian refinement of Ramses) mesh de-refinement is rarely required, as the mass in any cell rarely decreases if the mesh moves in a Lagrangian way (unlike in

Eulerian AMR codes, where cells are continuously refined/de-refined to follow the movement of mass). However, if other refinement criteria are used (e.g. to resolve shocks, radiation fronts, chemical gradients etc.), frequent de-refinement could be required, incurring an accuracy penalty.

Arepo's scheme combines the most important advantages of Lagrangian schemes (stemming from their Galilean-invariance) with those of Eulerian schemes: predictable and locally-controllable solution accuracy. However, this comes at a greatly increased cost compared to codes with a structured mesh, in both memory and compute time, for a number of reasons. First, the mesh vertices must be stored in memory (whereas a structured mesh does not need to be explicitly stored at all), taking up hundreds of bytes per vertex. Second, the nearest-neighbour search performed in SPH to find the nearby particles that a given particle should interact with is replaced here by the generation of the mesh itself — the information is encoded in the mesh data structure ahead of flux calculation, but the compute cost is not avoided.

The unstructured mesh also makes it far more difficult to construct a magnetic monopole-free MHD scheme without including a high level of numerical dissipation, or using a divergence cleaning scheme, which also introduces error (Pakmor et al., 2011). On a cubic grid it is possible to use *constrained transport* instead Evans & Hawley 1988, implemented in Ramses by Teyssier et al. 2006; Fromang et al. 2006), which generates no monopoles by construction (using a staggered mesh for the magnetic field), so does not require cleaning or added dissipation.

2.1.5.3 Mesh-Free ALE — Gizmo

Gizmo (Hopkins, 2015) takes a philosophically different approach to Arepo. Its scheme splits up the domain around numerical 'particles' (as shown in Figure 2.6), similar to those of SPH, or Arepo's mesh generating points, each of which has a volume-partition $\psi_i(\vec{x})$ assigned to it, such that the volume associated with particle i is $V_i = \int \psi_i(\vec{x})\, d\vec{x}$. Unlike with the smoothing kernel in SPH, it is ensured that every point in space is fully 'owned' by the particles in the vicinity, by use of a normalised partition function

$$\psi_i(\vec{x}) = \frac{1}{\omega(\vec{x})} W\left(|\vec{x} - \vec{x}_i|; h(\vec{x})\right), \tag{2.87}$$

where

$$\omega(\vec{x}) \equiv \sum_j W\left(|\vec{x} - \vec{x}_j|; h(\vec{x})\right) \tag{2.88}$$

ensures the normalisation (i.e. that every point in space is accounted for in the partition with weight 1, though a given particle may only own a fraction of it), and W is an arbitrary continuous 1D function, with scale length $h(\vec{x})$ that sets the sharpness of the partition between particles. In the limit that the scale length becomes infinitesimally small, the volume around a particle will contain, with exclusive ownership, all the points that are closer to this particle than any others — this is a Voronoi tessellation by definition. In general, Gizmo's volume partition resembles a 'fuzzy' Voronoi mesh, for any choice of ψ (so long as it is the same for each particle).

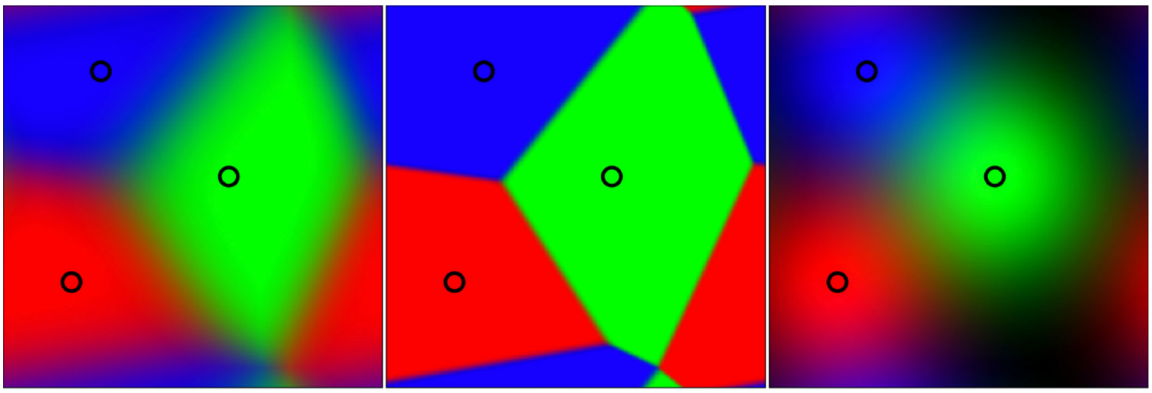


Figure 2.6: Partitioning of volume around 'particles' in Gizmo with a cubic spline partitioning function (*left*), Voronoi moving mesh methods (*middle*), and SPH (*right*). It is clear here that the kernel functions of SPH are not a true partitioning of volume, and the covering of the space is subject to numerical Monte Carlo noise.

A scheme similar to Godunov's finite volume method can be constructed by considering the volume integral form of the Euler equations, and approximating fields as piecewise polynomial (indexing vectors and tensors by superscripts α, β, and adopting Einstein summation convention), relating the vector of fields $\boldsymbol{Q}_i$ at particle i to the fluxes between i and each other particle j

$$\frac{d(V_i Q_i^\alpha)}{dt} + \sum_j \mathcal{F}_{ij}^{\alpha\beta} \left(V_i \Psi_j^\beta (x_i) - V_j \Psi_i^\beta (x_j) \right) = 0, \tag{2.89}$$

where $\mathcal{F}_{ij}^{\alpha\beta} = -\mathcal{F}_{ji}^{\alpha\beta}$ is a tensor that gives the fluxes (with components β) of each element Q_i^α from i to j, and Ψ_i^β is a vector of quadratic weightings of ψ_i, chosen to reconstruct gradients of piecewise-linear fields exactly, similarly to the aforementioned MUSCL schemes of Ramses and Arepo. Defining an effective area $A_{ij}^\beta \equiv V_i \Psi_j^\beta (\vec{x}_i) - V_j \Psi_i^\beta (\vec{x}_j)$ between i and j, the above can be written in standard finite volume form

$$\frac{d(V_i Q_i^\alpha)}{dt} + \sum_j \mathcal{F}_{ij}^{\alpha\beta} A_{ij}^\beta = 0. \tag{2.90}$$

Unlike in traditional FVM codes, this equation has been arrived at by an explicit volume integral, rather than transforming this into a surface integral around a cell using the divergence theorem, but the final result is mathematically identical to a Godunov scheme, and has very similar properties to the scheme of Arepo. Indeed, Gizmo demonstrates near identical accuracy and convergence to Arepo in a variety of test problems.

Like Arepo, Gizmo's scheme sets up Riemann problems parallel to the lines connecting particles, and updates the state of a well-defined (though now fuzzy) volume. These volumes do not necessarily need to follow the flow, so this is also an ALE scheme. Gizmo defines the position of the effective 'face' in each Riemann problem between neighbours as

$$\vec{x}_{ij} \equiv \vec{x}_i + \frac{h_i}{h_i + h_j} \left(\vec{x}_j - \vec{x}_i \right), \tag{2.91}$$

(which says that the face is located an equal proportion of the scale length from each particle), whereas Arepo makes this choice ahead of time during mesh construction (where faces separate regions containing points closest to their respective generators).

There are some important practical differences in the implementations of the two codes' schemes, however. First, Gizmo does not explicitly store a mesh, but rather performs a nearest neighbour search around each particle (as in SPH). This may save some computation compared to the cost of mesh construction, but more importantly, it vastly

reduces the memory footprint, since only particle centres and states need to be stored (identically to SPH codes), and greatly simplifies the software implementation and parallelisation. Second, Gizmo does not have an on the fly mesh refinement scheme, though it may be possible to implement an Arepo-like cell-splitting method. Computational differences aside, Gizmo shares the key numerical advantages and disadvantages of Arepo: it is an ALE scheme, enabling greater flexibility and uniformity of numerical accuracy than static meshes, without mesh-tangling/remapping issues, at the price of increased mathematical complexity and, for some problems, increased cost[4].

Ultimately, Ramses-rt was chosen as the base for the work in this thesis, for a few reasons. First, The radiation transport model, rather than the hydrodynamics, dominates the computational cost in the simulations performed here, and Ramses-rt's treatment is well-tested and well-used, with both source code and data from other cosmological simulations available for comparison to my model (Arepo was not publically available at the time of starting this work, and Gizmo's radiation transport scheme was not as mature). Using a Lagrangian mesh moving with the gas flow (as in Arepo) offers no advantage to the radiation transport scheme, as the mesh is effectively static compared to light; indeed, the use of an unstructured moving mesh only complicates the RT, increasing cost, and adding numerical diffusion of the radiation field during the mesh-remapping step. Secondly, Ramses-rt itself is a relatively simple, well-organised Fortran program that is easy to extend, and is able to natively interact with the Fortran source generated by krome, without further wrapping (as would be needed in a C or C++ code). Finally, SPH codes were ruled out as a host for my non-equilibrium chemical model, as reaction rates would depend strongly on the local density around a particle, which could be excessively noisy (being a Monte Carlo estimator of the actual density), causing a difficult-to-quantify reduction in accuracy of chemical predictions.

[4]However, Hopkins 2015 argues that in many galaxy-formation relevant cases, a given accuracy can be achieved more cheaply with Gizmo than with AMR codes. This is a complex issue; there may be fewer computations to perform in a single time step with structured grid codes than with particle codes, but Δt may be chosen differently to enforce stability, and, for high-velocity flows, Eulerian methods will need greater resolution to achieve the same accuracy as Lagrangian schemes.

2.1.6 Radiative Transfer

The RT in RAMSES-RT stands for *radiative transfer*. It is a deeply important part of cosmological simulations, as radiation from stars is responsible for reionising the universe, and affects the state of the gas in the interstellar medium (ISM) where successive generations of stars form.

2.1.6.1 Equations of Radiative Transfer

In classical geometric optics (ignoring wave and quantum effects), light can be described by a single field called the *spectral radiance* $I_\nu(\vec{x}, \hat{n}, t)$. This is defined such that

$$I_\nu \, d\nu \, d\Omega \, dA \, dt \tag{2.92}$$

is the energy in the radiation field in the spectral range $d\nu$, over a solid angle $d\Omega$, through an area dA, during a time dt. The equation of radiative transfer describes how the spectral radiance of radiation travelling in direction $\hat{n}$ evolves in a medium that interacts with light (Mihalas & Weibel-Mihalas, 1984)

$$\frac{1}{c}\frac{\partial I_\nu}{\partial t} + \hat{n} \cdot \vec{\nabla} I_\nu = -\kappa_\nu I_\nu + \eta_\nu, \tag{2.93}$$

where $\kappa_\nu(\vec{x}, \hat{n}, t)$ is the *absorption coefficient* and $\eta_\nu(\vec{x}, \hat{n}, t)$ is a source term. Various methods exist for numerically solving this equation; for example, the radiation field can be discretised into packets of photons travelling in the same direction, or *rays*, that can be cast from sources to gas cells in the simulation, integrating along the path to calculate the resultant field at the destination. Ray-based schemes have some disadvantages; they often assume an infinite speed of light (or incur increased cost if not), which can lead to incorrect propagation speed of ionisation fronts (I-fronts) around radiation sources, especially at low densities; their computational expense is also proportional to the number of sources (and scattering events), so in simulations that form many stars (or light up a large volume of optically thick medium) they can become intractable. Instead, RAMSES-RT uses a moment-based scheme, in which the radiation field is treated like a fluid, that can be advected across a grid similarly to gas in the numerical hydrodynamics scheme. This is

achieved by taking *angular moments* of Equation 2.93. The zeroth angular moment of I_ν is the integral over all solid angles (proportional to the photon number density N_ν)

$$h\nu c N_\nu(\vec{x}, t) \equiv \int_\Omega I_\nu \left(\vec{x}, \hat{\Omega}, t\right) d\Omega, \tag{2.94}$$

the first angular moment is the spectral flux density

$$F_\nu(\vec{x}, t) \equiv \int_\Omega I_\nu \left(\vec{x}, \hat{\Omega}, t\right) \hat{\Omega}\, d\Omega, \tag{2.95}$$

and the second moment is the radiative pressure tensor

$$\mathbb{P}_\nu(\vec{x}, t) \equiv \frac{1}{c} \int_\Omega I_\nu \left(\vec{x}, \hat{\Omega}, t\right) \hat{\Omega} \otimes \hat{\Omega}\, d\Omega. \tag{2.96}$$

Taking the zeroth and first moments (as in Rosdahl et al., 2013) of the equation of radiative transfer gives

$$\frac{\partial N_\nu}{\partial t} + \vec{\nabla} \cdot \vec{F}_\nu = \dot{N}_\nu^\star + \dot{N}_\nu^{\text{rec}} - c N_\nu \sum_j^{\text{species}} n_j \sigma_j(\nu), \tag{2.97}$$

$$\frac{\partial \vec{F}_\nu}{\partial t} + c^2 \vec{\nabla} \cdot \mathbb{P}_\nu = -c \vec{F}_\nu \sum_j^{\text{species}} n_j \sigma_j(\nu), \tag{2.98}$$

where $\int_\mathcal{V} \int_{\nu_0}^{\nu_1} \dot{N}_\nu^\star(\vec{x}, t) d\vec{x} d\nu$ is the rate of photons emitted by stars in a volume $\mathcal{V}$ and frequency range $[\nu_0, \nu_1]$ (see Section 2.1.7.2), similarly $\dot{N}_\nu^{\text{rec}}$ quantifies the emission of light due to recombinations in ionised gas (together these replace the source term η_ν), and the absorption term κ_ν has been split into a sum of the photoabsorption by all the chemical species in the simulation, each with a density n_j and cross section $\sigma_j(\nu)$. These absorption and recombination terms are associated with the photochemistry of the gas (covered in detail in Section 2.2), and are the sole mechanism by which radiation and gas interact.

These equations must be discretised in frequency to be numerically solvable, so we integrate over ν to treat the radiation in photon energy bins, rather than as a continuous

spectrum. Each energy bin i has a minimum photon energy ϵ_i, average photon energy $\bar{\epsilon}_i$, corresponding minimum and average frequencies ν_i and $\bar{\nu}_i$, and average cross sections and energy-weighted cross sections $\bar{\sigma}_{ij}$ and $\bar{\sigma}^E_{ij}$ for each species j. To calculate these, an underlying spectral energy distribution (SED) $J(\nu)$ must be assumed, which, in this scheme, must be considered a property of the radiation field itself, not of each individual source, that is, it must be the same for all sources in the simulation[5], although it is allowed to vary with time (or even space, though we do not use such a scheme). Thus, if the photon groups cover the spectrum without gaps, and mutually exclusively so that ν_{i+1} is both the minimum frequency in bin $i+1$ and the maximum frequency in bin i (or simply just the right edge if bin $i+1$ does not exist, usually taken to be ∞), we have

$$\bar{\sigma}_{ij} = \frac{\int_{\nu_i}^{\nu_{i+1}} \sigma_j(\nu) J(\nu)/h\nu \; d\nu}{\int_{\nu_i}^{\nu_{i+1}} J(\nu)/h\nu \; d\nu}, \tag{2.99}$$

$$\bar{\sigma}^E_{ij} = \frac{\int_{\nu_i}^{\nu_{i+1}} \sigma_j(\nu) J(\nu) \; d\nu}{\int_{\nu_i}^{\nu_{i+1}} J(\nu) \; d\nu}, \tag{2.100}$$

and

$$\bar{\epsilon}_i = \frac{\int_{\nu_i}^{\nu_{i+1}} J(\nu) \; d\nu}{\int_{\nu_i}^{\nu_{i+1}} J(\nu)/h\nu \; d\nu}. \tag{2.101}$$

Now the equations of radiative transfer can be transformed into a finite set of PDEs by integrating over each bin's frequency range, giving us

$$\frac{\partial N_i}{\partial t} + \vec{\nabla} \cdot \vec{F}_i = \dot{N}^\star_i + \dot{N}^{\text{rec}}_i - cN_i \sum_j^{\text{species}} n_j \bar{\sigma}_{ij}, \tag{2.102}$$

$$\frac{\partial \vec{F}_i}{\partial t} + c^2 \vec{\nabla} \cdot \mathbb{P}_i = -c\vec{F}_i \sum_j^{\text{species}} n_j \bar{\sigma}_{ij}, \tag{2.103}$$

where terms with subscript i represent the integrated quantity over the energy bin i.

[5] A multi-SED treatment could be achieved by using separate grids for each source, but this is not computationally viable for a non-trivial number of stars.

Finally, the equations must be *closed* by determining the pressure tensor $\mathbb{P}_i$; this can be done in theory by taking further angular moments of the radiative transfer equation, but eventually an approximation must be made for the highest order moment, leaving a finite system of equations to solve. RAMSES-RT employs the *M1 closure* (Levermore, 1984), in which the *Eddington tensor* $\mathbb{D}$, defined such that

$$\mathbb{P}_i = \mathbb{D}_i N_i, \tag{2.104}$$

is approximated to be

$$\mathbb{D}_i = \frac{1-\chi_i}{2}\mathbb{I} + \frac{3\chi_i - 1}{2}\hat{n}_i \otimes \hat{n}_i, \tag{2.105}$$

where $\mathbb{I}$ is the identity, and

$$\hat{n}_i = \frac{\vec{F}_i}{||\vec{F}_i||}, \quad \chi_i = \frac{3+4f_i^2}{5+2\sqrt{4-3f_i^2}}, \quad f_i = \frac{||\vec{F}_i||}{cN_i}, \tag{2.106}$$

which are, respectively, the unit vector in the direction of the photon flux, the *Eddington factor*, and the *reduced flux* which characterises the directionality of photons, taking values from 0 to 1, where 0 indicates an isotropic field, and 1 corresponds to a collimated beam. $\mathbb{D}$ is thus composed of an isotropic diffusive term (proportional to $\mathbb{1}$), and a free-streaming term (proportional to $\hat{n}_i \otimes \hat{n}_i$), each of which is exact in the corresponding optically thin ($f = 1$) and optically thick ($f = 0$) limits, but in between the limits $\mathbb{D}$ is approximate.

Other closures can be chosen to make a different compromise between computational cost and resolution of anisotropies. For example, the *diffusive approximation* is commonly applied to radiation-hydrodynamics problems, in which I_ν is assumed to be locally isotropic (as it is in the optically thick limit), so the flux density is instantaneously related to the number density gradient

$$F_\nu(\vec{x}, t) = -\frac{c}{3\sum_i^{\text{species}} n_i \sigma_i(\nu)}\vec{\nabla} N_\nu, \tag{2.107}$$

which can also be reproduced by neglecting the time derivative in Eq. 2.103 and taking the isotropic limit ($\mathbb{P}_\nu = \frac{N_\nu}{3}\mathbb{1}$) of the M1 closure. The radiative transfer equation then takes the form of the diffusion/heat equation (hence the name '*diffusive*')

$$\frac{\partial N_\nu}{\partial t} - \frac{c}{3\sum_i^{\text{species}} n_i\sigma_i(\nu)}\nabla^2 N_\nu = \dot{N}_\nu^\star + \dot{N}_\nu^{\text{rec}} - cN_\nu \sum_i^{\text{species}} n_i\sigma_i(\nu). \tag{2.108}$$

This is only a good approximation in optically thick gas, but leaves only one PDE to solve, so the computational cost is significantly less than with the M1 closure[6].

2.1.6.2 Numerics

Once closed, the equations of radiative transfer can be solved numerically. Notice that Equations 2.102 and 2.103 have the form of conservation laws (like the Euler equations 2.20, 2.21 and 2.22), more clearly written as

$$\frac{\partial \mathcal{U}}{\partial t} + \vec{\nabla}\cdot\mathcal{F} = 0, \tag{2.109}$$

$$\mathcal{U} = \begin{bmatrix} N_i \\ \vec{F}_i \end{bmatrix}, \quad \mathcal{F} = \mathcal{F}(\mathcal{U}) = \begin{bmatrix} \vec{F}_i \\ c^2\mathbb{P}_i \end{bmatrix} \tag{2.110}$$

where the source and sink terms have been neglected. These equations can therefore be solved by a finite volume method, very similarly to the hydrodynamics described in Section 2.1.2. However, as we are dealing with light, rather than a fluid, there are some details in which the schemes must differ, and some drawbacks to be aware of. For example, with the M1 closure, beams travelling into each other collide and are diverted perpendicularly at the point of collision (as shown in Figure 2.7), rather than simply passing through each other independently as they would in reality (although for astrophysical problems this tends not to be a major concern). One choice of inter-cell flux scheme is the Global Lax Friedrich (GLF), which sets the eigenvalues of the Jacobian $\frac{\partial \mathcal{F}_i}{\partial \mathcal{U}_j}$ to be $\pm c$, in its treatment

[6]Note that the parabolic form imposes an additional stability condition that $\Delta t \propto (\Delta x)^2$ (which would drastically increase the computational expense), unless a *flux limiter* is used — forcing the flux to not exceed the maximum physically-allowed value of cN. This can also be achieved through use of a slope limiter on the reconstructed photon field, as is done for the hydrodynamics; RAMSES-RT uses this method.

of the characteristics, returning (in one spatial dimension)

$$\mathcal{F}_{k+1/2} \approx \frac{\mathcal{F}_k + \mathcal{F}_{k+1}}{2} - c\frac{\mathcal{U}_{k+1} - \mathcal{U}_k}{2} \tag{2.111}$$

where k is the spatial index of the cell. This is a very diffusive scheme, so beams have a tendency to de-collimate, but this helps to overcome the numerical bias towards transport along grid directions for isotropic sources, so that the isotropy is better captured here than with less diffusive schemes (for example HLL, which is also implemented in Ramses-rt, that takes the reduced flux into account in its calculation of the wave speed eigenvalues, better capturing directionality); this is ideal for many astrophysical sources such as stars within galaxies, which are perfectly-isotropic emitters. The work in this thesis makes exclusive use of the GLF scheme.

Another numerical consideration is that the CFL condition

$$\Delta t < \frac{\Delta x}{3c}, \tag{2.112}$$

when applied to light, requires a very small timestep in comparison with the Δt chosen for the hydrodynamics (in proportion to the maximum sound speed in gas divided by the speed of light, which can be typically $< \frac{1}{300}$). In order to match the RT and hydrodynamics time steps, Ramses-rt employs a *reduced speed of light approximation* (RSLA; originated by Gnedin & Abel 2001), in which a new speed of light $\bar{c} << c$ replaces c in the equations of radiative transfer. This can bring the RT time step back into line with the hydrodynamics, but the drawbacks must be carefully considered. The effect of a reduced c on I-fronts can be demonstrated in an idealised scenario: the Strömgren Sphere (Strömgren, 1939, explored in detail in Chapter 3). Here a star emits ionising photons at a rate $\dot{N}$ into a homogeneous static medium of initially neutral hydrogen of density n. The I-front expands out to the *Strömgren radius*

$$r_S = \left(\frac{3\dot{N}}{4\pi\alpha n^2}\right)^{\frac{1}{3}}, \tag{2.113}$$

at which the recombination rate α equals the flux of ionising photons, so the front stops

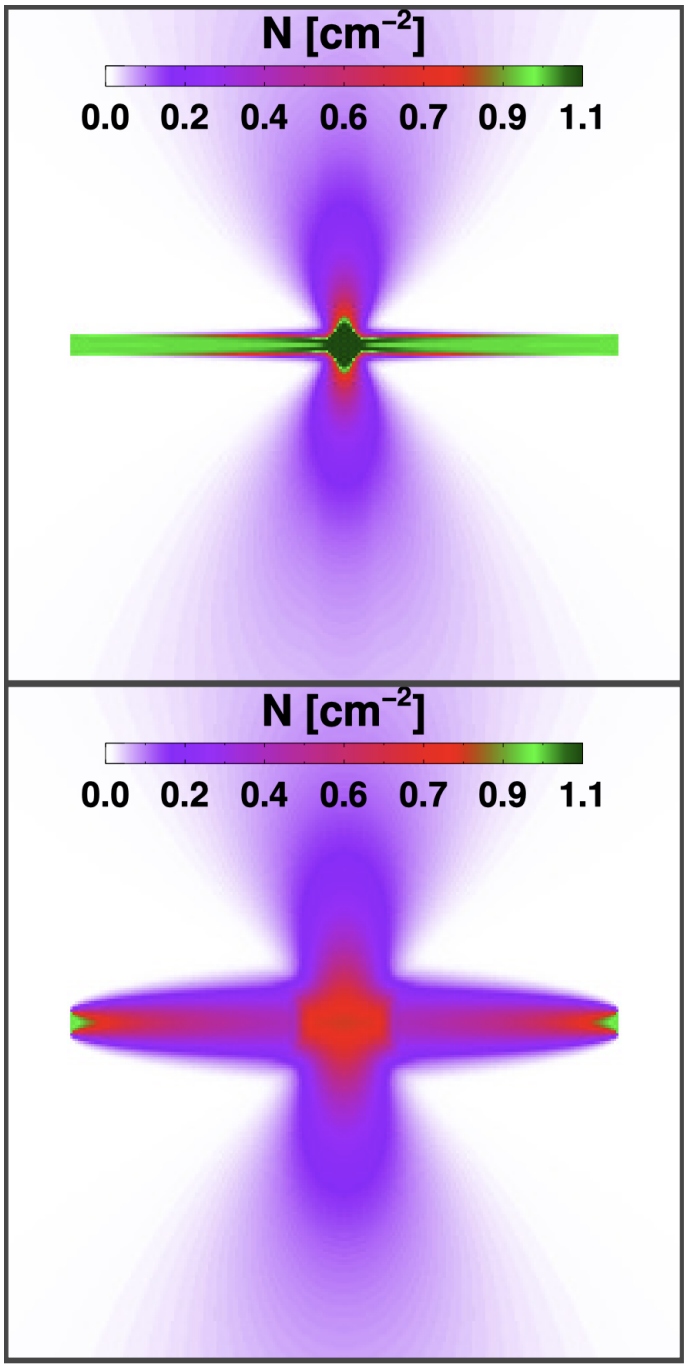


Figure 2.7: Photon number density for two directional sources facing each other, simulated by RAMSES-RT (from Rosdahl et al. (2013)). The beams meet in the middle and unphysically collide, spreading out, with little directionality, from the centre. *Top:* using HLL to calculate inter-cell fluxes. *Bottom:* using GLF to calculate inter-cell fluxes; the beam diffuses much more strongly here than with HLL.

expanding. Shapiro et al. (2006) derived the evolution of the I-front up to this point: the dimensionless time $w = t/t_{\rm rec}$ (where $t_{\rm rec} = 1/(n\alpha)$ is the recombination time) and radius $y = r/r_S$ obey

$$w = qy - \ln\left(1 - y^3\right), \tag{2.114}$$

where $q = r_s/ct_{\rm rec}$ is the ratio of the Strömgren Sphere light crossing time to the recombination time. For $q = 0$ (equivalent to the $c \to \infty$ limit), where recombination is relatively slow, we have

$$y = (1 - e^{-w})^{\frac{1}{3}}, \tag{2.115}$$

$$= \left(1 - \left(1 - w + \frac{w^2}{2!} + \cdots\right)\right)^{\frac{1}{3}}, \tag{2.116}$$

$$\approx w^{\frac{1}{3}} \quad \text{for } w << 1, \tag{2.117}$$

$$\implies r = r_S \left(\frac{t}{t_{\rm rec}}\right)^{\frac{1}{3}}, \tag{2.118}$$

whereas for $q >> 1$, where the crossing time is much larger than the recombination time,

$$y \approx \frac{wct_{\rm rec}}{r_S}, \quad \text{for } w << 1 \tag{2.119}$$

$$\implies r = ct, \tag{2.120}$$

so, when recombination is relatively fast (or put another way, for high photon emission rate), I-fronts expand at the speed of light. Indeed, for any positive value of q, the speed of expansion is initially c, until the $\ln\left(1 - y^3\right)$ term becomes more significant (after $\sim 1t_{\rm cross}$), so the front slows down to follow $r \propto t^{\frac{1}{3}}$, before reaching r_S and stopping. Thus, a reduced speed of light scheme fails to capture the correct expansion speed in the high q regime, and I-fronts are not well-resolved until the simulation time exceeds one Strömgren Sphere light crossing time, when the front would have sufficiently slowed down. Rosdahl et al.

(2013) demonstrate that a reduced speed of light is inappropriate for capturing I-fronts in the intergalactic medium (IGM) that originate from the light from entire galaxies (where the low density leads to a high q factor), but is more valid for stars in the denser interstellar medium (ISM), where the high-q phase of Strömgren Sphere expansion is relatively short-lived.

2.1.6.3 Update Steps

The combined radiation-hydrodynamic system (Equations 2.20, 2.21, 2.22, 2.102, and 2.103) is solved by *operator splitting* the radiation transport (described by Eqs. 2.102 and 2.103) from the hydrodynamics (described by Eqs. 2.20, 2.21, and 2.22); that is, in each global time step, the hydrodynamic state is calculated independently, before the radiative (and thermochemical) state is updated — discretising the photochemical coupling between the hydrodynamics and radiative transfer. The radiative update itself is split further into an injection step, transport step, and thermochemistry step, carried out in this order.

In the injection step, the $\dot{N}_i^\star$ term is added to the total number of photons in bin i

$$N_i^{n+1} = N_i^n + \dot{N}_i^\star \Delta t, \tag{2.121}$$

where the source term is a function of the mass, age and metallicity of the star particle, and the SED $J(\nu)$ (discussed further in Section 2.1.7.2). By splitting updates of the source and absorption terms into the injection and thermochemistry steps, the equations of radiative transfer are in the pure conservation form of Equation 2.109, and can be solved by the finite volume GLF scheme described earlier. Finally the thermochemistry step deals with the absorption and emission of radiation by chemical processes, and the associated change in thermal and chemical state of the gas (described comprehensively in Section 2.2). These steps share the time step Δt of the hydrodynamics, and require that the CFL condition for the radiation (Eq. 2.112) be satisfied in addition to the hydrodynamic conditions (Eq. 2.60).

2.1.7 Sub-grid Physics

After stars form in a galaxy, energy is injected into gas at small scales by stellar radiation, winds and supernovae, eventually being transferred to large scales via hydrodynamic and radiative processes, in an *inverse energy cascade*. The inverse energy cascade places a very high dynamic range requirement on simulations: in order to capture the transfer of energy from small to large scales, the simulation must resolve these scales, and those in between, simultaneously. However, it is computationally impossible for any simulation of even just one dwarf galaxy to resolve the scale of a star; indeed modern simulations of such objects are limited to $\sim$ 1pc resolution at best. It is then also impossible to consistently resolve the formation of stars.

So, cosmological simulations can neither self-consistently simulate the formation of stars within galaxies, nor afford to ignore their energy output via radiation and supernovae. An alternative approach to fully simulating gas on the scale of stars is necessary. Simulations instead employ *sub-grid* models, that approximate the behaviour of stars (or other compact objects) as a direct interaction with the integrated state of gas and radiation over their host grid cell (typically at least 0.7 pc in size, for the work in this thesis). There are three key sub-grid models that are used in the simulations in this thesis that collectively capture the interaction between gas, radiation and stars in the early universe: star formation, radiation and supernovae. I ignore stellar winds for the purposes of this thesis, as they are much less significant than radiation and supernovae for massive stars (Geen et al., 2015).

2.1.7.1 Star Formation

Since gas collapse cannot be followed all the way down to protostellar scales, a sub-grid model must take over. RAMSES-RT implements various recipes for star formation, but the one used in this thesis is that of Rosdahl et al. (2018) (similar to that of Kimm et al. 2017, inspired by Federrath & Klessen 2012). Here, stars are formed according to a Schmidt law (Schmidt, 1959), in which the density of stars $\rho_\star$ changes in proportion to the gas density divided by the gas free-fall time

$$\frac{d\rho_\star}{dt} = \epsilon_{\rm ff}\frac{\rho_{\rm gas}}{t_{\rm ff}}\ ; \quad t_{\rm ff} = \sqrt{\frac{3\pi}{32G\rho_{\rm gas}}}, \tag{2.122}$$

where $\epsilon_{\rm ff}$ is the *efficiency*, which is determined on a per-cell basis (as in Kimm et al. 2017), using a model (the 'The Multi-freefall PN Model' from Federrath & Klessen 2012) that considers the free-fall times of a lognormal distribution of sub-grid clouds, formed by an interacting system of shocks (Padoan et al., 1997). In this model,

$$\epsilon_{\rm ff} = \frac{\epsilon}{2\phi_t}\exp\left(\frac{3\sigma_s^2}{8}\right)\left[1+\mathrm{erf}\left(\frac{\sigma_s^2 - s_{\rm crit}}{\sqrt{2\sigma_s^2}}\right)\right] \tag{2.123}$$

where $\epsilon \approx 0.5$ is the fraction of gas that accretes onto protostars (the rest being pushed away by the protostar itself), $\phi_t \approx 0.57$ is a factor used to correct for the uncertainty in the calculation of free-fall time of clouds within the cell , $\sigma_s^2 = \ln(1 + b^2\mathcal{M}^2)$ is the variance of the log density contrast $\ln(\rho/\rho_0)$ (where b is the *turbulent forcing parameter* — a value of 0.4 is used, corresponding to *mixed* forcing; see Federrath et al. (2008) — and $\mathcal{M} = ||\vec{u}||/c_s$ is the *Mach number*), and $s_{\rm crit} = \ln(0.067\theta^{-2}\alpha_{\rm vir}\mathcal{M}^2)$ is the minimum log density contrast for sub-grid clouds to collapse (Padoan & Nordlund, 2011), $\theta \approx 0.33$ corrects for uncertainty in post-shock thickness of clouds, and $\alpha_{\rm vir} = 2E_{\rm kin}/E_{\rm grav}$ is the *virial parameter*. ϵ is calibrated against observed galactic clouds, and the values of ϕ_t, θ are from fits to high-resolution simulations of star-forming clouds by Federrath & Klessen (2012).

Stars may only form in a cell if it is refined to the maximum level, if the flow is convergent ($\vec{\nabla}\cdot\vec{u} < 0$), and if the *turbulent Jeans length* (Federrath & Klessen, 2012) is unresolved at cell size Δx:

$$\lambda_{J,\rm turb} = \frac{\pi\sigma_v^2 + \sqrt{36\pi c_s^2 G\Delta x^2\rho_{\rm gas} + \pi^2\sigma_v^4}}{6G\rho_{\rm gas}\Delta x} < \Delta x, \tag{2.124}$$

(where σ_v^2 is the gas velocity dispersion in the cell), i.e. if the sum of the thermal and turbulent pressure is insufficient to support the gas inside the cell against its own gravity.

Once these conditions are satisfied, a cell forms a star particle with mass $m_\star = N_\star m_{\star,\rm min}$, sampling the random variable $N_\star$ from a Poisson distribution with mean

$$\lambda = \epsilon_{\rm ff} \frac{\rho_{\rm gas} \Delta x^3}{m_{\star,\rm min}} \frac{\Delta t}{t_{\rm ff}}, \tag{2.125}$$

such that $N_\star$ is statistically consistent with Equation 2.122. For Population II stars, these particles each represent a distribution of stars formed with a Chabrier IMF (Chabrier, 2005), with masses between $0.1\,\mathrm{M}_\odot$ and $150\,\mathrm{M}_\odot$, where the luminosity is normalised to the particle mass. $m_{\star,\rm min}$ is a numerical parameter that controls how many stars particles are formed in a simulation; choosing a lower mass leads to more particles being formed (and thus more realistic results, up to a point) but also incurs greater computational cost. The maximum mass of individual stars *of interest* ($\sim 100 M_\odot$ for the work here) sets a scale below which it is no longer sensible to discretise their distribution with particles representing a population (the mass of a single star should not be split into more than one particle) — instead, particles representing individual stars should be used.

Population III star particles are treated differently (as in Kimm et al. 2017). These are only allowed to form in gas where the metallicity is below $10^{-6} Z_\odot$ (where $Z_\odot$ is the Solar metallicity). Following Wise et al. (2012), their mass is randomly sampled from a Salpeter-like IMF:

$$\frac{dN_{\star,\rm III}}{d\log M} \propto M^{-1.3} \exp\left[-\left(\frac{M_{\rm char}}{M}\right)^{1.6}\right], \tag{2.126}$$

where $M_{\rm char} = 100\,\mathrm{M}_\odot$ is the characteristic mass, above which there is an exponential cutoff in the formation rate. Unlike Pop. II star particles, these particles each represent individual stars rather than a population.

2.1.7.2 Radiative Feedback

Population II star particles emit light according to the spectral evolution model of Bruzual & Charlot (2003). Given that each particle represents a population formed with a Chabrier IMF, with metallicity Z (inherited from the gas at formation) and a certain age, the resultant SED can be calculated, and thus the contribution from a star into the photon energy bins N_i. Population III stars are treated independently of the others (using the $Z = 0$ case of the model of Schaerer 2002), as these stars can potentially be much more massive and

thus emit significantly more ionising radiation.

The moment-based radiation treatment must assume a single SED $J(\nu)$ for all of the photons in each bin, so although the photon contribution from each star to bin N_i can be calculated consistently, the underlying SED must be approximated as the average SED of all the stars in the simulation in the computation of average bin energies and cross-sections. This average global SED $J\left(\nu; t^{n+1}\right)$ is updated on-the-fly, every time step t^{n+1}, as stars form and age in a simulation. This is only a good approximation when the spectra of stars are of similar shapes to each other in a given bin (also improving as the width of bins decreases). Even for a mixture of Pop. III and II stars, this is likely to be true for the UV bins, as these will be mostly occupied by the tail of a black body distribution (whose shape is not very sensitive to the temperature, so long as it is well beyond the peak).

2.1.7.3 Supernovae

Since the early stages of the expansion of supernovae can happen on scales well below 1pc, it not possible to capture their behaviour in simulations self-consistently without much higher grid resolution, as depositing the equivalent energy into a cell that is too large causes the shock to be numerically over-cooled — the temperature is too low, and radiative cooling takes away much more relative energy (e.g. Katz, 1992; Kim & Ostriker, 2015) — so the momentum is not efficiently transferred into neighbouring cells.

There exist various numerical approaches to solving this issue, for example, a delayed cooling scheme can be used where radiative cooling is switched off until the shock has propagated sufficiently (e.g. Mori et al., 1997; Stinson et al., 2006), or momentum can be injected directly instead of thermal energy (e.g. Navarro & White, 1993) (although the resulting shocks can then still be over-cooled), but these have to be tuned carefully to the specific resolution of the simulation they are used in. Kimm & Cen (2014) have suggested a more robust alternative (updated in Kimm et al. 2017), where momentum, mass and thermal energy are injected as a function of the state of the surrounding cells.

Supernovae, in their adiabatic stage before radiative cooling becomes relevant, behave as Sedov-Taylor blast waves (Sedov, 1946; Bethe et al., 1958; Taylor, 1950), where the radius r (and, by taking the time derivative, the speed u) of the blast at time t are given by the

famous solution (obtained by dimensional analysis)

$$r \propto \left(\frac{E_0 t^2}{\rho_0}\right)^{\frac{1}{5}}, \tag{2.127}$$

$$u \propto \left(\frac{E_0}{\rho_0 t^3}\right)^{\frac{1}{5}}, \tag{2.128}$$

with an initial energy injection of E_0 in a medium with density ρ_0. This stage of the blast is energy conserving, as radiative losses are relatively low. The blast wave then transitions into a regime where the shock is driven by the pressure of the low-density bubble left behind by the Sedov-Taylor blast (Chevalier, 1974; Blondin et al., 1998), called the *snowplow* phase. In a fit to the results of simulations by Geen et al. (2015), which account for the effect of photoionisation of the medium by the host star (reducing the density and thus radiative cooling rate, increasing the momentum transferred), Kimm et al. (2017) find the total momentum deposited up to this point to be

$$p_{SN,PH} = 4.2 \times 10^5 \,\mathrm{kms}^{-1} M_\odot E_{51}^{16/17} (Z/Z_\odot)^{-0.14}, \tag{2.129}$$

although in simulations with radiative transfer, the Strömgren Sphere may actually be resolved, in which case the reduction in density by photoionisation is already accounted for, and instead a fit to to simulations (without photoionisation) by Thornton et al. (1998) applies instead

$$p_{SN,snow} = 3 \times 10^5 \,\mathrm{kms}^{-1} M_\odot n_H^{-2/17} E_{51}^{16/17} (Z/Z_\odot)^{-0.14}, \tag{2.130}$$

where n_H is the hydrogen number density in units of cm^{-3}, E_{51} is the energy output of the supernova in units of 10^{51} erg and $Z/Z_\odot$ is the metallicity of the medium in solar units. A successful sub-grid model should impart this momentum into cells where the Sedov-Taylor stage of the blast is unresolved, but in lower density gas, the blast radius may extend beyond a cell length before it transitions into the pressure-driven stage, in which case less momentum should be injected. Thus, the simulations in this thesis use

the scheme of Kimm & Cen (2014) and Kimm et al. (2017), where momentum is added to cells neighbouring a supernova (subtending solid angle $\Delta\Omega$ to the host cell, at angular position Ω) according to the stage of the blast

$$\Delta p = \frac{\Delta\Omega}{4\pi} \cdot \begin{cases} \sqrt{2\chi M_{\rm ej} f_e E_{\rm SN}} & \chi < \chi_{\rm tr} \quad \text{(Sedov-Taylor)}, \\ p_{\rm SN} & \chi \geq \chi_{\rm tr} \quad \text{(Snowplow)}, \end{cases} \tag{2.131}$$

where $E_{\rm SN}$ is the supernova energy, $M_{\rm ej}$ is the mass ejected by the supernova, $p_{\rm SN}$ is the total momentum imparted up to the Sedov-Taylor stage, interpolated between the regimes of Thornton et al. (1998) and Geen et al. (2015) depending on whether the Strömgren radius r_S is resolved at cell length Δx

$$p_{\rm SN} = p_{SN,snow} \exp\left[\frac{-\Delta x}{r_S}\right] + p_{SN,PH}\left(1 - \exp\left[\frac{-\Delta x}{r_S}\right]\right), \tag{2.132}$$

where a minimum metallicity of $(Z/Z_\odot)_{\rm min} = 0.01$ is taken, and $\chi(\Omega) \equiv \Delta M_{\rm shell}/\Delta M_{\rm ej}$ is the mass ratio of the blast shell (containing ejecta and swept up mass) to the mass of the ejecta in cells subtending solid angle $\Delta\Omega$. χ is used to determine which stage the blast would be in as it passes through a given cell, with transition between regimes at

$$\chi_{\rm tr} = \frac{900 n_H^{-4/17} E_{51}^{-2/17} (Z/Z_\odot)^{-0.28}}{f_e^{\rm tr}(M_{\rm ej}/M_\odot)}, \tag{2.133}$$

obtained by equating $p_{\rm SN,snow}$ with the Sedov-Taylor momentum at transition $(2\chi_{\rm tr} M_{\rm ej}$ $f_e^{\rm tr} E_{\rm SN})^{1/2}$, where $f_e^{tr} = 0.676$ is the fraction of energy left in the bubble (not yet lost radiatively) at transition (Blondin et al., 1998). $f_e(\chi) \equiv 1 - \frac{\chi-1}{3(\chi_{\rm tr}-1)}$ is the energy fraction as a function of χ, smoothly interpolating between the two regimes.

The mass ejected into a given cell is

$$\Delta M_{\rm ej} = \frac{(1-\beta) M_{\rm ej}}{N_{\rm nbor}}, \tag{2.134}$$

where $\beta = 4/52$ is chosen such that mass is distributed evenly between the host and neighbouring cells, and $N_{\rm nbor}$ is the number of neighbouring cells — 48 in the case where the neighbours are refined one level more than the host (if the neighbours are not refined,

their properties are rescaled so they can be treated as such). The mass swept up into a cell with density $\rho_{\rm nbor}$ (including the initial ejecta), around a host cell with density $\rho_{\rm host}$ and size Δx, is

$$\Delta M_{\rm shell} = \rho_{\rm nbor}\left(\frac{\Delta x}{2}\right)^3 + \frac{(1-\beta)\rho_{\rm host}\Delta x^3}{N_{\rm nbor}} + \Delta M_{\rm ej}. \tag{2.135}$$

In order to conserve energy, the total kinetic energy added as a result of injecting momentum into neighbouring cells is subtracted from the supernova energy $E_{\rm SN}$, and this leftover energy is shared equally between the neighbouring cells. The ejected mass is also added to the cells involved, and subtracted from the host star particle.

The value of $M_{\rm ej}$ depends on whether a star particle represents Population II or Population III stars; if Population II, the mass ejected is a fixed fraction of the particle mass, which for a Chabrier IMF is $\alpha \approx 0.317$; if Population III, it is assumed that stars with masses $140M_\odot < M < 230M_\odot$ explode in pair-instability supernovae (Heger et al., 2003), in which the entire mass of the star is ejected, else if stars with masses $M < 40M_\odot$ are treated the same as Population II stars (except for the energy output and metal enrichment, see below). Population III stars with masses $40M_\odot < M < 140M_\odot$ are assumed to not explode, and those with masses $M > 230M_\odot$ directly collapse into black holes, ejecting no mass (although these are highly unlikely to form in our prescription anyway). For Population II supernovae, the energy output is a a constant $E_{\rm SN} = 10^{51}$ erg; for Population III supernovae, the energy is (following the prescriptions compiled by Wise et al. 2012)

$$E_{\rm SN,III} = 10^{51}\ {\rm erg}\cdot\begin{cases} 1 & 11 \le M/M_\odot < 20, \\ -13.714 + 1.086M & 20 \le M/M_\odot < 40, \\ 5.0 + 1.304\cdot(M_{\rm He} - 64) & 140 \le M/M_\odot < 230, \\ 0 & \text{otherwise}, \end{cases} \tag{2.136}$$

where $M_{\rm He} = \frac{13}{24}(M-20)$ is the helium core mass. For Population II supernovae, 5% of the ejecta is assumed to be metallic (Arnett, 1996), and for Population III supernovae, the

mass of metals ejected is (again following Wise et al. 2012)

$$M_{\mathrm{Z,III}} = M_{\odot} \cdot \begin{cases} 0.1077 + 0.3383(M - 11) & 11 \leq M/M_{\odot} < 20, \\ -2.7650 + 0.2794M & 20 \leq M/M_{\odot} < 40, \\ M_{\mathrm{He}} & 140 \leq M/M_{\odot} < 230, \\ 0 & \text{otherwise}, \end{cases} \tag{2.137}$$

Finally, the lifetime of Population II progenitor stars is randomly sampled from the integrated supernova occurrence rate of the STARBURST99 model (Leitherer et al., 1999), such that the supernova rate for a star particle is statistically consistent with the model predictions given the particle mass and Chabrier IMF. If a Population III star is in a mass range that allows it to explode, its lifetime is calculated from a fit to the results of Schaerer (2002); these are very short (~2 Myr) for massive Population III stars, so pair instability supernovae occur very quickly in simulations after forming massive Population III stars.

2.2 Thermochemistry

In addition to hydrodynamical heating (by compression), chemical and radiative processes can transfer energy in and out of gas. Indeed, without radiative cooling, gas would never be able to reach high enough densities to form stars, which come to dominate the energy flow in galaxies. Thus, astrophysical simulations must make their best effort to capture the processes that lead to gas cooling to form stars (or being prevented from doing so). In this section I describe a number of thermochemical models for use in simulations, including a scheme I have developed myself that solves a rich chemical network to accurately capture H_2 formation and cooling in primordial gas.

2.2.1 Thermal Evolution

The temperature T of gas is related to the specific thermal energy $\varepsilon = \epsilon - \frac{u^2}{2}$ (where ϵ is the specific total energy) by

$$T = \varepsilon \frac{(\gamma - 1) m_p \mu}{\rho k_\mathrm{B}}, \tag{2.138}$$

where m_p is the proton mass, μ is the mean particle mass in units of m_p (e.g. $\mu = 1.0$ for atomic hydrogen gas), ρ is the gas density, and γ is the adiabatic index (which may be allowed to vary for the thermochemical update, but is fixed to $\frac{5}{3}$ in the hydrodynamics solver for the work in this thesis). Due to the fast evolution of μ in some situations (by strong ionisation, for example), it is easier to numerically consider the quantity

$$T_\mu \equiv \frac{T}{\mu}, \tag{2.139}$$

which obeys

$$\frac{\partial T_\mu}{\partial t} = \frac{(\gamma - 1) m_p}{\rho k_\mathrm{B}} \dot{\varepsilon}, \tag{2.140}$$

where the rate of change of the specific energy (not including the effect of hydrodynamic compression/expansion, which are treated separately in the code) is the sum of the heating ($\mathcal{H}$) and cooling ($\mathcal{L}$) rates $\dot{\varepsilon} = \mathcal{H} + \mathcal{L}$. In general, these rates depend on the incident radiation field and chemical makeup of the gas.

2.2.2 Cooling Processes

In astrophysical gas, the cooling term $\mathcal{L}$ can be very complex, as any process within the gas that emits radiation is a potential source of cooling (and these tend to interact with each other as gas gets optically thick); practical models must only deal with the major processes, as it is impossible to account for the radiation of every possible chemical species (indeed even in primordial gas, there is a rich spectrum that is difficult to capture). The major sources of cooling (all of which are included in my model) are:

- **Collisional ionisation** — electrons thermally collide with another species A, with ionisation state n, to excite another electron to the point of being unbound

$$\mathrm{e^- + A^n \longrightarrow A^{n+1} + e^- + e^-}, \tag{2.141}$$

which removes thermal energy from the gas equivalent to the ionisation energy of species A. Even today, the collisional ionisation rates used in simulations for hydrogen and helium (Cen, 1992; Black, 1981) are a fit to results from lab-based experiments (Lotz, 1967).

- **Collisional excitation** — collisions with electrons can also excite bound electrons to higher energy levels (such that they remain bound, i.e. without ionising the host species), which then decay, releasing radiation that escapes the gas (if it is optically thin to that frequency on the scale of a simulation cell, which is usually assumed). This cooling rate can be tabulated from numerical models or, for some levels, approximated analytically (see the compilation of Black 1981). In general, this rate can also depend on the incident radiation field in a non-trivial way (especially for metal ions and molecules which have a more dense spectrum), but rates used in astrophysical simulations do not usually capture this.

- **Bremsstrahlung** — "Braking radiation" in English. Thermal electrons in plasma are slowed down by their electromagnetic interactions with positively charged nuclei, emitting radiation in the process. Assuming low optical depth, this can be calculated analytically to a good approximation (Osterbrock & Ferland, 2006).

- **Recombination** — A free electron can combine with an ion, being bound into a free energy level in the ion, then decaying into a more stable state, releasing energy in the form of radiation which can escape the gas

$$e^- + A^{n+1} \longrightarrow A^n + \gamma\,. \tag{2.142}$$

  In the case of *free-to-ground* recombination, the free electron is captured directly into the ground state, releasing a photon with enough energy to ionise another atom of the same type. Many simulations (including those described later in this thesis) take the approximation that gas is optically thick to this radiation, such that these photons immediately ionise a local atom — the *on-the-spot approximation* (OTSA). Under the OTSA, only case B recombination (all transitions except free-to-ground,

as opposed to case A which covers all possible transitions) needs to be considered. Cooling rates due to hydrogen and helium recombinations can be calculated using analytic/numerical theoretical predictions; e.g. Hui & Gnedin (1997) provide fits to the results of Ferland et al. (1992) and Burgess & Seaton (1960).

- **Compton cooling(/heating)** — CMB photons Compton scatter off free electrons in plasma, either gaining or losing energy depending on the difference in temperature between the radiation field and the electrons (Cen, 1992). In the early universe, there exists some gas which may be cooler than the CMB itself for some time (and is thus heated by it), but eventually the CMB temperature drops such that this is a cooling term for most gas — the CMB is inverse Compton scattered, with photons gaining energy at the expense of electrons in the gas.

- **Molecular cooling** — Molecules (such as H_2 or HD) have much more rich absorption/emission spectra than atoms, as their structure allows them to rotate or vibrate in addition to having electrons excited around their host atoms. These *rovibrational* transitions can be excited in collisions by much smaller energies than atomic energy levels, and emit infrared photons when they decay, rather than UV or optical. Thus, molecules can radiatively cool gas much more efficiently than atoms at low temperatures. As molecules tend to occur in relatively high density gas, the optical depth should be considered in this cooling term. There exist various ways of modelling the internal opacity of a single cell, discussed later in this thesis. Another important molecular process is *collision-induced emission* (CIE) — separate from collisional excitation above — in which molecules collide with other atoms or molecules, temporarily forming a 'supermolecule' with an electric dipole which than decays, emitting a photon (Ripamonti & Abel, 2004). CIE from H_2–H_2 interactions dominates the cooling in highly-molecular, dense primordial gas, and thus is an important contributor to the formation of the first stars.

- **Metal line emission** — metal ions can have emission lines at lower energies than hydrogen and helium, due to their more complex electronic structure, and so become much more effective at cooling gas (per mass of coolant) at intermediate

temperatures ($\lesssim 10^4$K). This is very difficult to model line-by-line, as a system of many metal ions can interact in a non-trivial way. Codes can employ a lookup table created by running 1D radiative transfer simulations with a comprehensive photochemical network (e.g. CLOUDY Ferland et al. 2013, 2017), to approximate the cooling rate as a function of the gas metallicity, density and temperature. 'Metal line emission' is nothing but a combination of the excitation and recombination processes mentioned earlier in this list, but is listed separately here as it requires a drastically different treatment to these same processes in hydrogen and helium, due to the complexity of the system of energy levels in metallic gas.

- **Chemical reactions** — endothermic reactions take thermal energy from the gas. These include the aforementioned collisional ionisations, and also collisional dissociation of H_2 by electrons, atomic hydrogen, and H_2 itself. Omukai (2000) provide cooling rates associated with the dissociation of H_2, and heating rates for its formation.

A number of other cooling processes exist, but are considered too minor/difficult to model in the simulations in this thesis; for example, gas can also cool or heat by collisions with dust (which then very efficiently radiates in the infrared).

2.2.3 Heating Processes

Gas can be heated either by internal exothermic chemical reactions (for example the formation of H_2), by Compton scattering of hot CMB photons in the early universe (as mentioned previously), or by absorbing photons in ionisations or dissociations. Of these, photon absorption the most important in galaxies after stars have formed. For each absorption process k with cross section $\sigma_k(\nu)$, threshold energy ϵ_k and associated absorber species j with density $n_{j(k)}$ (where the species j is a non-injective function of the process k; for example, absorption by both H_2 dissociation and ionisation are associated with n_{H_2}), the contribution to the heating term is

$$\mathcal{H}_k^{\text{abs}} = c n_{j(k)} \int_0^\infty N_\nu \sigma_k(\nu) \, (h\nu - \epsilon_k) \, d\nu. \tag{2.143}$$

Then, discretising into photon bin i and taking the RSLA, with average cross section for process k over the bin $\bar{\sigma}_{ik}$, energy weighted-cross section $\bar{\sigma}^E_{ik}$, the contribution to heating by bin i (remembering that N_i and $\bar{\epsilon}_i$ are the photon concentration and the average photon energy in bin i, and $\bar{c}$ is the reduced speed of light) is

$$\mathcal{H}^{\text{abs}}_{ik} = \bar{c} N_i n_{j(k)} \left(\bar{\epsilon}_i \bar{\sigma}^E_{ik} - \epsilon_k \bar{\sigma}_{ik} \right), \tag{2.144}$$

thus, the total heating by all photoabsorption processes is obtained by summing over all bins i and absorption processes k (where the species $j = j(k)$ — labelling the number density — is a function of the absorption process k)

$$\mathcal{H}^{\text{abs}} = \sum_i \sum_k c N_i n_{j(k)} \left(\bar{\epsilon}_i \bar{\sigma}^E_{ik} - \epsilon_k \bar{\sigma}_{ik} \right). \tag{2.145}$$

In molecular hydrogen, more complex photon absorption processes are possible than in atoms, and not all the energy absorbed should be added to the medium as suggested by Equation 2.143, as excited states can fluoresce in the infrared, allowing energy to escape the gas without being efficiently reabsorbed. In photodissociation, a molecule is excited by an incoming UV photon, which then decays into the vibrational continuum in which the atoms become unbound, causing it to dissociate. The dissociated hydrogen atoms are themselves excited, and lose further energy as they relax. Additionally, not all absorptions of Lyman-Werner photons lead to dissociations; $\sim 85\%$ of the time, the excited state decays into bound states, which are the collisionally or radiatively de-excited. Collisional de-excitations may transfer the bound energy into the gas, whereas in radiative de-excitations (which dominate in optically thin gas) it is lost as infrared radiation escapes.

Molecular hydrogen photoheating thus requires special treatment. I use the prescription of Baczynski et al. (2015) for dissociation heating and UV pumping, in which 0.4 eV is deposited into the gas for every dissociation (Black & Dalgarno, 1977), and the ratio of fluorescent absorptions (that heat without dissociating) to dissociations is taken to be 6.94 (found by Draine & Bertoldi 1996 to be slightly higher in typical ISM conditions than in optically thin gas). For each non-dissociating absorption, the energy deposited is given by Hollenbach (1990)

$$2\,\mathrm{eV} \cdot \frac{C_d}{C_d + C_r}; \tag{2.146}$$

where $C_d = 10^{-12}(1.4 \cdot e^{\frac{-18100}{T+1200}} \frac{n_{\mathrm{chH2}}}{n} + e^{\frac{-1000}{T}} \frac{n_H}{n}) n\sqrt{T}\, s^{-1}$ is the collisional de-excitation rate (with the total number density given by n in cm^{-3}, and the temperature in Kelvin by T), and $C_r = 2 \times 10^{-7}\, s^{-1}$ is the radiative de-excitation rate.

2.2.4 Cosmological UV Background

In simulations of individual galaxies, it is not possible to self-consistently model the illumination from a background of other galaxies, that may be close enough to contribute significant UV radiation, but are not captured in the simulation box[7]. Instead, a universal cosmological background of UV radiation can be assumed, contributing to photoionisation and heating terms in the thermochemistry of every cell, without tracking the radiation explicitly. This can be done separately from the normal RT calculations, in that the photoionisation and heating can be calculated without the same assumed $J(\nu)$; the simulations in this thesis that make use of a UV background use the empirically-calibrated redshift-dependent model of Faucher-Giguère et al. (2009). Of course, in reality some regions of gas would be better shielded from this background radiation than others, but this can be difficult to capture if the radiation is not tracked directly. A reasonable approximation can be made that the UV radiation is attenuated as a function of density in a cell (although this fails to capture shielding from adjacent dense cells); following Rosdahl & Blaizot (2012), in simulations using a UV background, the incident background flux is attenuated exponentially above a density threshold by a factor

$$f_{\mathrm{shd}}^{\mathrm{UV}} = \exp\left[\frac{-\rho X / m_p}{0.01\,\mathrm{cm}^{-3}}\right], \tag{2.147}$$

where X is the hydrogen fraction of the universe (usually taken to be ≈ 0.76 immediately after the Big Bang), such that the numerator of the exponent is the number of hydrogen atoms in cosmological gas of equivalent density.

[7] Though this is not so important for more isolated galaxies, such as the ones studied in a later chapter of this thesis.

2.2.5 Equilibrium Thermochemistry

The simplest possible treatment of heating and cooling assumes that chemicals in the gas are always in chemical equilibrium, so that the cooling rate is purely a function of density and temperature. For pristine ($Z = 0$) gas, RAMSES (the parent code of RAMSES-RT[8]) employs the prescription of Katz et al. (1996) (assuming no UV background), in which H, H^+, He, He^+ and He^{2+} are in ionisation equilibrium with each other at the gas temperature, assuming the gas is optically thin (so photons emitted by recombination are not re-absorbed locally). By assuming this equilibrium, cooling via collisional excitations/ionisations, recombinations, free-free bremsstrahlung emission and Compton scattering of CMB photons can be calculated. For metal-enriched gas, an interpolation is made between this and the cooling functions of Sutherland & Dopita (1993), who model gas at solar metallicity with a code that simulates a complex system of ions, in order to additionally include cooling by metal line emission. Cooling by H_2 (which is very important in the early universe) is not considered.

The problem with this treatment is that the cooling rate is not accurately estimated in gas that would not actually be in chemical equilibrium. Indeed, this is the case for very much of the gas in galaxies, as photoionisation can be very strong, and timescales of gravitational collapse are not necessarily greater than timescales of chemical processes in low-density gas, where collisions are rare, so the equilibrium assumption effectively leads to overestimation of chemical reaction rates, as the density increases too quickly.

2.2.6 Non-equilibrium H and He Photochemistry (RAMSES-RT)

RAMSES-RT simulates the photoionisation of hydrogen and helium as part of its radiative transfer treatment, so non-equilibrium information can be used in calculating the photoionisation heating rate, and various cooling rates that depend on ionisation fraction, rather than having to assume concentrations of ions based on ρ and T. This involves solving a system of ODEs that account for collisional ionisations, photoionisations, and recombinations of the ions of H and He, which are (expressed in terms of the ionisation

[8]RAMSES-RT also uses this prescription if RT and non-equilibrium chemistry are switched off.

fractions $x_{\mathrm{HII}}, x_{\mathrm{HeI}}, x_{\mathrm{HeII}}, x_{\mathrm{HeIII}}$, with M photon energy groups)

$$\frac{\partial x_{\mathrm{HII}}}{\partial t} = x_{\mathrm{HI}} \left[\beta_{\mathrm{HI}} n_e + \sum_{i=1}^{M} \bar{c} N_i \bar{\sigma}_{i\mathrm{HI}} \right] - x_{\mathrm{HII}} \alpha_{\mathrm{HII}} n_e, \tag{2.148}$$

$$\begin{aligned} \frac{\partial x_{\mathrm{HeII}}}{\partial t} = {} & x_{\mathrm{HeI}} \left[\beta_{\mathrm{HeI}} n_e + \sum_{i=1}^{M} \bar{c} N_i \bar{\sigma}_{i\mathrm{HeI}} \right] \\ & - x_{\mathrm{HeII}} \left[\beta_{\mathrm{HeII}} n_e + \alpha_{\mathrm{HeII}} n_e + \sum_{i=1}^{M} \bar{c} N_i \bar{\sigma}_{i\mathrm{HeII}} \right] + x_{\mathrm{HeIII}} \alpha_{\mathrm{HeIII}} n_e, \end{aligned} \tag{2.149}$$

$$\frac{\partial x_{\mathrm{HeIII}}}{\partial t} = x_{\mathrm{HeII}} \left[\beta_{\mathrm{HeII}} n_e + \sum_{i=1}^{M} \bar{c} N_i \bar{\sigma}_{i\mathrm{HeII}} \right] - x_{\mathrm{HeIII}} \alpha_{\mathrm{HeIII}} n_e, \tag{2.150}$$

where n_e is the electron density, $\beta_{\ldots}$ and $\alpha_{\ldots}$ are collisional ionisation and recombination coefficients for each species, and $\bar{\sigma}_{i\ldots}$ are average photoionisation cross sections over group i. Nucleus number conservation ($x_{\mathrm{HI}} + x_{\mathrm{HII}} = 1$ and $x_{\mathrm{HeI}} + x_{\mathrm{HeII}} + x_{\mathrm{HeIII}} = 1$) is used to eliminate the need to solve the ODEs for x_{HI} and x_{HeI}, and charge conservation allows the ODE for n_e to be additionally eliminated.

Photon groups are conventionally chosen such that the left edge of the energy bin $h\nu_i$ is the ionisation energy of a corresponding species, so there is a sharp distinction between, for example, radiation that can ionise only hydrogen and radiation that can also ionise neutral helium, thus capturing the steepest changes in the absorption spectrum. In this case that means using 3 energy bins: $[13.6\,\mathrm{eV} - 24.59\,\mathrm{eV})$, $[24.59\,\mathrm{eV} - 54.42\,\mathrm{eV})$, $[54.42\,\mathrm{eV} - \infty)$, where the left edges are hydrogen's ionisation energy, helium's first ionisation energy, and helium's second ionisation energy respectively.

Solving for the photoionisation of gas on-the-fly allows Ramses-rt to calculate non-equilibrium heating and cooling rates, including photoionisation heating, collisional ionisation cooling (Cen, 1992), collisional excitation cooling (Black, 1981), and recombination cooling (Hui & Gnedin, 1997). Indeed, all processes mentioned in Section 2.2.2 except for molecular, metal and chemical cooling are calculated here as a function of ionisation state.

Associated with the equations above, the photon densities and fluxes must also evolve as a result of photoabsorptions and recombinations. These evolve according to Equations

2.102 and 2.103, with the advection and stellar source terms *operator-split* into a different update step (as discussed previously), leaving

$$\frac{\partial N_i}{\partial t} = \dot{N}_i^{\mathrm{rec}} - \bar{c} N_i \sum_j^{\{\mathrm{HII},\, \mathrm{HeII},\, \mathrm{HeIII}\}} n_j \bar{\sigma}_{ij}, \tag{2.151}$$

$$\frac{\partial F_i}{\partial t} = -\bar{c} F_i \sum_j^{\{\mathrm{HII},\, \mathrm{HeII},\, \mathrm{HeIII}\}} n_j \bar{\sigma}_{ij}. \tag{2.152}$$

Now we have a system of ODEs describing the evolution of the thermal, chemical and radiative state of a cell (separate from the flow of gas or radiation) that must be solved numerically. This can be done in numerous ways. For example, the simplest possible scheme is *forward Euler*

$$\vec{y}^{n+1} = \vec{y}^n + \Delta t \mathbf{A} \vec{y}^n, \tag{2.153}$$

where we have written the system in matrix form $\frac{\partial}{\partial t}\vec{y} = \mathbf{A}\vec{y}$, and $\mathbf{A} = \mathbf{A}(\vec{y}, t)$ can vary in general (i.e. the equation may actually be non-linear). This scheme is first-order, so its accuracy scales as $O(\Delta t)$, meaning that many time steps must be done to achieve high-accuracy. The scheme is also explicit, meaning that the updated value $\vec{y}^{n+1}$ depends explicitly on the backwards-in-time value $\vec{y}^n$, which means it takes at least two time steps to capture a two-way coupling between ODEs (appearing as off-diagonal terms in $\mathbf{A}$), meaning an even smaller choice of Δt may be required, in the case of strong coupling. *Runge-Kutta methods* (Runge, 1895; Kutta, 1901) improve on the accuracy of explicit Euler at the cost of having to evaluate the derivative matrix $\mathbf{A}$ more than once, but are still explicit, so suffer from the problem with strong coupling. The simplest example of an implicit method is the first-order *implicit Euler* scheme

$$\vec{y}^{n+1} = \vec{y}^n + \Delta t \mathbf{A} \vec{y}^{n+1}, \tag{2.154}$$

$$\implies \vec{y}^{n+1} = (\mathbb{I} - \Delta t \mathbf{A})^{-1} \vec{y}^n \tag{2.155}$$

which cannot be solved immediately, but instead requires a costly matrix inversion (or solution of the implicit equation by iterative methods). However, in inverting the matrix, coupling is resolved immediately (though still to only first-order accuracy), as each element of the forward-in-time vector has dependence on the entire matrix of derivatives, not just one row. Thus, implicit schemes have the potential to outperform explicit schemes if they enable the use of a much larger timestep Δt by better resolving two-way coupling.

RAMSES-RT employs a *semi-implicit scheme* with a *10% rule*, in which each variable is updated one at a time (with the variables expected a-priori to evolve the fastest updated last), but forward-in-time values are used for variables that have already been updated, making the scheme more 'implicit' for variables updated later (which better captures coupling between variables from cross-terms in the ODEs). For example, a dummy system of variables x, y, z that obey

$$\frac{\partial}{\partial t}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} f(x,y,z) \\ g(x,y,z) \\ h(x,y,z) \end{bmatrix}, \tag{2.156}$$

where z is expected to change the fastest, then y, then x, would be updated (with a forward Euler scheme here for simplicity) according to

$$x^{n+1} = x^n + \Delta t \cdot f(x^n, y^n, z^n), \tag{2.157}$$

$$y^{n+1} = y^n + \Delta t \cdot g(x^{n+1}, y^n, z^n), \tag{2.158}$$

$$z^{n+1} = z^n + \Delta t \cdot h(x^{n+1}, y^{n+1}, z^n), \tag{2.159}$$

which allows the update of each variable to use incrementally more information. RAMSES-RT uses this semi-implicit technique, but with a second-order update for each variable, instead of the forward Euler shown above; if a variable x obeys $\frac{\partial x}{\partial t} = f(\vec{x})$, where $\vec{x}$ contains the most recently updated system (a mix of backwards and forwards-in-time values), then it is updated with

$$x^{n+1} = x^n + \frac{f(\vec{x})\Delta t}{1 - \left.\frac{\partial f}{\partial x}\right|_{\vec{x}} \Delta t}. \tag{2.160}$$

This may still have quite poor accuracy used in isolation for ODEs with strong two-way coupling, so a 10% rule is applied, in which, if a single variable changes by more than 10%, the update step is restarted with $\Delta t \rightarrow \Delta/2$, and twice as many steps are done to make up the total Δt required. This rule is applied recursively, ensuring that coupling is always captured at least at the 10% level. The semi-implicit scheme tends to perform very well in cells where the photoionisation is not too strong, only taking a few cycles to satisfy the 10% rule, but where there is a strong ionising radiation field, cells may need hundreds of cycles after a few recursive applications of the 10% rule, in which case a fully implicit scheme may perform better. However, integrated over all the cells in a simulation over all time, the scheme is a good compromise between speed and accuracy for this particular system of ODEs, as photoionisation events may be short lived, and a large number of cells in the simulation may not be subject to fast changes in thermochemical state for very long.

2.2.7 Semi-equilibrium H_2 Photochemistry

The treatment of Ramses-rt still does not account for H_2 formation, which is particularly important in the low-metallicity early universe. Kimm et al. (2017) and Katz et al. (2017) suggest a simple addition to the system above, that assumes chemical equilibrium for the intermediate reactant H^- in H_2 formation, and immediate dissociation of ionised H_2^+, thus avoiding having to track these species separately[9].

The main H_2 formation channel in primordial gas is

[9]It should be noted that Nickerson et al. (2018) have also developed a very similar H_2 photochemsitry system in Ramses, making similar equilibrium assumptions about intermediate reactants and products, but I choose to use the scheme of Kimm et al. (2017) and Katz et al. (2017) for simulations in later chapters of this thesis, due to their code being more well-integrated with the other sub-grid methods I use at the time of writing.

$$\mathrm{H} + \mathrm{e}^- \longrightarrow \mathrm{H}^-, \tag{2.161}$$

$$\mathrm{H}^- + \mathrm{H} \longrightarrow \mathrm{H}_2 + \mathrm{e}^-, \tag{2.162}$$

where the concentration of H^- is also determined by

$$\mathrm{H}^- + \mathrm{H}^+ \longrightarrow \mathrm{H} + \mathrm{H}, \tag{2.163}$$

$$\mathrm{H}^- + \mathrm{e}^- \longrightarrow \mathrm{H} + \mathrm{e}^- + \mathrm{e}^-. \tag{2.164}$$

These are reactions 1, 2, 5 and 13 of Glover et al. (2010), with reaction rate coefficients k_m that depend on gas temperature (where m is the reaction number); so in equilibrium we have

$$k_1 n_{\mathrm{HI}} n_e = k_2 n_{\mathrm{HI}} n_{\mathrm{H}^-} + k_5 n_{\mathrm{HII}} n_{\mathrm{H}^-} + k_{13} n_e n_{\mathrm{H}^-}, \tag{2.165}$$

so the rate of H_2 formation via the primordial channel, assuming an equilibrium concentration of H^- is

$$R_p(T) = \frac{k_1 k_2}{k_2 + k_5 x_{\mathrm{HII}} + k_{13} x_e}, \tag{2.166}$$

where $n_{...}$ are number densities and $x_{...} = \frac{n_{...}}{n_{\mathrm{H}}}$ are number fractions of the subscripted species relative to the total number of hydrogen nuclei (in any ionisation state). H_2 can also form on dust, which can still be important compared to the primordial channel, even at low gas metallicities. For this, the approximation of Gnedin et al. (2009) is used, which estimates the dust concentration as a function of the metallicity, then the H_2 formation rate on this approximate dust population as

$$R_d = 3.5 \times 10^{-17} Z C_f \ \mathrm{cm}^3 \mathrm{s}^{-1}, \tag{2.167}$$

where $C_f \approx 10$ is the 'clumping factor', representing the effect of sub-grid scale density

variance, which increases the formation rate (as, microscopically, the formation rate is approximately quadratic in gas density, being at least a two-body interaction).

H_2 can be dissociated or ionised by radiation, in one of the following reactions

$$H_2 + \gamma_{LW} \longrightarrow H + H, \tag{2.168}$$

$$H_2 + \gamma_{UV} \longrightarrow H_2^+ + e^-, \tag{2.169}$$

where γ_{LW} is a photon in the *Lyman-Werner band* that can dissociate H_2, with energy between 11.2 eV and 13.6 eV, and γ_{UV} is a photon with energy of at least 15.2 eV, that can ionise H_2. In this prescription, it is assumed that H_2^+ ions immediately dissociate via

$$H_2^+ + e^- \longrightarrow H + H, \tag{2.170}$$

so photoionisations effectively act as another dissociation channel. In this prescription, it is sensible to add two photon energy bins with left edges at 11.2 eV and 15.2 eV, splitting the default bins of RAMSES-RT for a total of five. H_2 has many absorption lines in the Lyman-Werner band, which are broadened by the Doppler effect, in addition to their natural width (a result of the short lifetime of the excited states). The line width is enhanced in columns of dense gas, as even the 'wings' of lines reach significant opacities, absorbing light continuously across the band, which allows the outside of gas clouds to very effectively *shield* the inside from dissociating radiation, as a steep nonlinear function of density. The spectrum is discretised very coarsely, so it is impossible to resolve the correct behaviour self-consistently within the RT solver when lines overlap, so an explicit model for self-shielding model must be used. Kimm et al. (2017); Katz et al. (2017) use the prescription of Gnedin et al. (2009) (modified from Draine & Bertoldi 1996), that multiplies the dissociation cross section in the Lyman-Werner band by a factor

$$f_{shd} = \frac{0.8}{(1+x)^2} + \frac{0.2}{\sqrt{1+x}} e^{-0.00085\sqrt{1+x}}, \tag{2.171}$$

where $x \equiv N_{H_2}/5{\times}10^{14}\ \mathrm{cm}^{-2}$ and N_{H_2} is the column density of H_2. Since radiation is treated

in bulk here (rather than tracking individual rays), the column density experienced by radiation from each source is not known. Instead, the H_2 column density in the expression above is approximated as $N_{H_2} = n_{H_2}\Delta x$, and f_{shd} is calculated for each cell (rather than per-ray or per-source). Although this scheme realistically represents the reduced *global* dissociation rate in dense columns of molecular gas, it unphysically *increases* the transmission of Lyman-Werner radiation through denser cells, which may lead to dissociation rates being increased *locally* in regions illuminated by this transmitted radiation, and also increase the escape fraction of Lyman-Werner radiation in galaxies.

Finally, the ODE describing the H_2 fraction is

$$\frac{\partial x_{H_2}}{\partial t} = -C x_{H_2} - \sum_{i=1}^{M} \bar{c} N_i \left[f_{\rm shd} \bar{\sigma}^{\rm dis}_{iH_2} + \bar{\sigma}^{\rm ion}_{iH_2} \right] + \left(R_p + R_d \right) n_{\rm HI}, \tag{2.172}$$

where C is the sum of collisional dissociation coefficients by e^-, H, He and H_2 (given by Glover & Abel 2008), and $\bar{\sigma}^{\rm dis}_{iH_2}$, $\bar{\sigma}^{\rm ion}_{iH_2}$ are the H_2 photodissociation and photoionisation cross sections in photon group i. The H_2 formation and dissociation terms must also be added to the $x_{\rm HI}$ ODE for consistency (and the photoabsorption terms to the RT equations), but otherwise the system is unchanged from the previous section, and is solved in the same way. An H_2 cooling term is added to $\mathcal{L}$ from Hollenbach & McKee (1979), and heating due to photoionisations, photodissociations, and non-dissociative absorption then subsequent fluorescence (*UV pumping*) is also accounted for.

2.2.8 Fully Non-Equilibrium H_2 and HD Photochemistry

Assuming an equilibrium concentration of H^- is not necessarily valid. Indeed, the gravitational collapse of gas into the first stars can happen too quickly for an equilibrium concentration of H^- (and thus H_2) to be found in the cores of clouds, so equilibrium models may overestimate cooling by molecular hydrogen in the early universe. Once stars form, neglecting the effect of photoionisation (which creates the free electrons needed to form H_2) may have the opposite effect. There is also a potentially significant abundance of Deuterium formed during the Big Bang ($[2.55 \pm 0.02] \times 10^{-5}$ of total hydrogen; Riemer-Sørensen & Jenssen 2017), which can form the molecule HD, allowing cooling via another

channel in primordial gas, which is even more efficient than H_2 at low temperatures (Figure 1.4).

I have modified Ramses-rt to solve a large system of the most important chemical and photo reactions involved in the formation and destruction of the primordial molecules H_2 and HD, self-consistently coupled to the radiative transfer. The existing implementation is almost entirely replaced by a solution based on krome (Grassi et al., 2014), which generates Fortran code to solve the ODEs associated with a user-specified chemical network (and the heating and cooling of gas by the species involved), with an implicit, adaptive, sparse-matrix scheme suitable for large, stiff systems that exhibit dynamics over a range of timescales. I have also modified this version of krome to ensure the network is properly coupled to the moment-based RT of Ramses-rt, so both the photon fluxes and densities are properly evolved.

For a general network of chemical reactions, the density of each species n_j obeys

$$\frac{dn_j}{dt} = \sum_{a \in F_j} \left(k_a \prod_{r \in R_a} n_r \right) - \sum_{b \in D_j} \left(k_\mathrm{B} \prod_{r \in R_b} n_r \right), \tag{2.173}$$

which states that the rate of change of the concentration n_j of species j is equal to the sum of the reaction rates k of the set of reactions $\{a \in F_j\}$ that form the species minus the same for reactions $\{b \in D_j\}$ that destroy it. Each reaction a has a rate coefficient k_a and a set of reactants $\{r \in R_a\}$. I use a modified version of a network (shown in Table 2.2) from Grassi et al. (2014), originally based on Omukai (2000), with some added reactions concerning deuterium from Galli & Palla (2002). Frequency dependent photoabsorption cross section tables are sourced from Heays, A. N. et al. (2017) and the Phidrates database (Huebner & Mukherjee, 2015), which are integrated on-the-fly during simulations to calculate average cross sections. The network requires a fixed photon group configuration, where each photoabsorption reaction has a left bin edge associated with it (Table 2.1). It includes:

- Collisional ionisations, photoionisations and case B recombinations of H and He ions as in Ramses-rt.
- The most important reactions in deciding the concentration of the hydrogen anion

Group	Photon energy range / eV	Left edge significance
1	[0.75 – 2.65)	H^- detachment
2	[2.65 – 11.2)	H_2^+ dissociation
3	[11.2 – 13.59)	H_2 dissociation
4	[13.59 – 15.42)	H ionisation
5	[15.42 – 24.58)	H_2 ionisation
6	[24.58 – 54.41)	He ionisation
7	[54.41 – ∞)	He^+ ionisation

Table 2.1: The fixed photon energy groups used in my modified code. The left edges are chosen to align with minimum ionisation/dissociation energies. Group 3 is the *Lyman-Werner Band*, containing H_2 dissociating photons.

H^- (which participates in the main H_2 formation channel, including the photodetachment of H^- by near-infrared/optical photons).

- Explicit tracking of the H_2^+ ion, and the processes which create and destroy it (including photodissociation), which allows separate treatment of H_2 dissociations and ionisations, and together with the H^- prescription, captures the interaction of radiation with H_2 formation via H_2^+ and H^-.

- Three channels of H_2 formation: via H^-, and two kinds of 3-body hydrogen interactions (which become important at high densities); and various dissociation and ionisation processes of H_2, both collisionally and by Lyman-Werner and more energetic UV photons.

- The main reactions in determining the concentration of HD: charge transfer between hydrogen and Deuterium, and Deuteron exchanges with H_2 (and the inverse reactions).

A term from Gnedin et al. (2009) is also added to the n_{H_2} ODE to account for the H_2 formation on dust, as in the semi-equilibrium treatment of Kimm et al. (2017); Katz et al. (2017),

$$\frac{dn_{H_2}}{dt} = \ldots + R_d n_H^2, \tag{2.174}$$

where R_d given by Equation 2.167.

This fully non-equilibrium, radiation-coupled treatment of H_2 and HD formation and destruction allows self-consistent calculation of the cooling rates associated with these molecules, which together with the cooling from ions and free electrons, account for all major cooling processes in pristine primordial gas. Heating and cooling rates are provided by KROME (Grassi et al., 2014), but the recombination cooling is modified to be consistent with the OTSA (only case B recombinations contribute) using the rates provided by Hui & Gnedin (1997). I use the treatment of Baczynski et al. (2015) (as described previously in Section 2.2.3) for dissociation heating and UV pumping.

Table 2.2: My reaction network. γ_i represents a photon from energy bin i. *Reactions 32-35 are free-to-ground recombinations that would emit photons in reality, but we use the On The Spot Approximation (OTSA), in which we assume ionising photons emitted by these recombinations are immediately re-absorbed, thus the photons shown in these reactions are ignored in the model, and case B recombination rates are used in both the reaction rates and the cooling function. Rate coefficients for each reaction are given in Table 2.3.

№	Reaction	Reference
1	$H + e^- \longrightarrow H^+ + e^- + e^-$	Cen (1992)
2	$He + e^- \longrightarrow He^+ + e^- + e^-$	Cen (1992)
3	$He^+ + e^- \longrightarrow He^{2+} + e^- + e^-$	Cen (1992)
4	$H^- + H \longrightarrow H_2 + e^-$	Kreckel et al. (2010)
5	$H + H^+ \longrightarrow H_2^+$	Coppola et al. (2011)
6	$H_2^+ + H \longrightarrow H_2 + H^+$	Karpas et al. (1979)
7	$H_2 + H^+ \longrightarrow H_2^+ + H$	Savin et al. (2004)
8	$H_2 + e^- \longrightarrow H + H^-$	Capitelli et al. (2007)
9	$H_2 + e^- \longrightarrow H + H + e^-$	Mitchell & Deveau (1983)
10	$H_2 + H \longrightarrow H + H + H$	Martin et al. (1996)
11	$H^- + e^- \longrightarrow H + e^- + e^-$	Janev et al. (1987)
12	$H^- + H \longrightarrow H + H + e^-$	Abel et al. (1997)
13	$H^- + H^+ \longrightarrow H + H$	Stenrup et al. (2009)
14	$H^- + H^+ \longrightarrow H_2^+ + e^-$	Poulaert et al. (1978)
15	$H_2^+ + e^- \longrightarrow H + H$	Coppola et al. (2011)
16	$H_2^+ + H^- \longrightarrow H + H_2$	Dalgarno & Lepp (1987)
17	$H + H + H \longrightarrow H_2 + H$	Forrey (2013)
18	$H_2 + H + H \longrightarrow H_2 + H_2$	Glover & Abel (2008)
19	$H_2 + H_2 \longrightarrow H + H + H_2$	Omukai (2000)
20	$D + H_2 \longrightarrow HD + H$	Galli & Palla (2002)
21	$D^+ + H_2 \longrightarrow HD + H^+$	Galli & Palla (2002)
22	$HD + H \longrightarrow D + H_2$	Galli & Palla (2002)
23	$HD + H^+ \longrightarrow D^+ + H_2$	Galli & Palla (2002)
24	$H^+ + D \longrightarrow H + D^+$	Galli & Palla (2002)

№	Reaction	Reference
25	$H + D^+ \longrightarrow H^+ + D$	Galli & Palla (2002)
26	$H_2^+ + H_2 \longrightarrow H_3^+ + H$	Galli & Palla (1998)
27	$H_2 + H^+ \longrightarrow H_3^+$	Galli & Palla (1998)
28	$H_3^+ + e^- \longrightarrow H_2 + H$	Galli & Palla (1998)
29	$H_3^+ + H \longrightarrow H_2^+ + H_2$	Galli & Palla (1998)
30	$He^+ + H \longrightarrow He + H^+$	Yoshida et al. (2006)
31	$He + H^+ \longrightarrow He^+ + H$	Yoshida et al. (2006)
32*	$H^+ + e^- \longrightarrow H + \gamma_4$	Hui & Gnedin (1997)
33*	$He^+ + e^- \longrightarrow He + \gamma_6$	Hui & Gnedin (1997)
34*	$He^{2+} + e^- \longrightarrow He^+ + \gamma_7$	Hui & Gnedin (1997)
35*	$H + e^- \longrightarrow H^- + \gamma_1$	Stancil et al. (1998)
36	$H^- + \gamma_1 \longrightarrow H + e^-$	Heays, A. N. et al. (2017)
37	$H^- + \gamma_2 \longrightarrow H + e^-$	Heays, A. N. et al. (2017)
38	$H^- + \gamma_3 \longrightarrow H + e^-$	Heays, A. N. et al. (2017)
39	$H^- + \gamma_4 \longrightarrow H + e^-$	Heays, A. N. et al. (2017)
40	$H_2^+ + \gamma_2 \longrightarrow H^+ + H$	Heays, A. N. et al. (2017)
41	$H_2^+ + \gamma_3 \longrightarrow H^+ + H$	Heays, A. N. et al. (2017)
42	$H_2^+ + \gamma_4 \longrightarrow H^+ + H$	Heays, A. N. et al. (2017)
43	$H_2 + \gamma_3 \longrightarrow H + H$	Heays, A. N. et al. (2017)
44	$H_2 + \gamma_4 \longrightarrow H + H$	Heays, A. N. et al. (2017)
45	$H_2 + \gamma_5 \longrightarrow H + H$	Heays, A. N. et al. (2017)
46	$H + \gamma_4 \longrightarrow H^+ + e^-$	Heays, A. N. et al. (2017)
47	$H + \gamma_5 \longrightarrow H^+ + e^-$	Heays, A. N. et al. (2017)
48	$H + \gamma_6 \longrightarrow H^+ + e^-$	Heays, A. N. et al. (2017)
49	$H + \gamma_7 \longrightarrow H^+ + e^-$	Heays, A. N. et al. (2017)
50	$H_2 + \gamma_5 \longrightarrow H_2^+ + e^-$	Heays, A. N. et al. (2017)
51	$H_2 + \gamma_6 \longrightarrow H_2^+ + e^-$	Heays, A. N. et al. (2017)
52	$H_2 + \gamma_7 \longrightarrow H_2^+ + e^-$	Heays, A. N. et al. (2017)
53	$He + \gamma_6 \longrightarrow He^+ + e^-$	Huebner & Mukherjee (2015)
54	$He + \gamma_7 \longrightarrow He^+ + e^-$	Huebner & Mukherjee (2015)
55	$He^+ + \gamma_7 \longrightarrow He^{2+} + e^-$	Huebner & Mukherjee (2015)

Table 2.3: Reaction rate coefficients for each reaction, from the references in Table 2.2. In reactions 11 and 12, $T_e = k_B T/(1\,\mathrm{eV})$, i.e. the thermal energy in electronvolts.

№	Reaction Rate Coefficient k_i
1	$\beta_{\mathrm{H}} = 5.85 \times 10^{-11}\sqrt{T}\left(1+\sqrt{T/10^5}\right)^{-1}\exp(-157809.1/T)$
2	$\beta_{\mathrm{He}} = 2.38 \times 10^{-11}\sqrt{T}\left(1+\sqrt{T/10^5}\right)^{-1}\exp(-285335.4/T)$
3	$\beta_{\mathrm{He^+}} = 5.68 \times 10^{-12}\sqrt{T}\left(1+\sqrt{T/10^5}\right)^{-1}\exp(-631515/T)$

№	Reaction Rate Coefficient k_i
4	$1.35 \times 10^{-9} \left(T^{9.8493\times10^{-2}} + 3.2852 \times 10^{-1} T^{5.5610\times10^{-1}} + 2.7710 \times 10^{-7} T^{2.1826}\right)$ $/\left(1 + 6.1910 \times 10^{-3} T^{1.0461} + 8.9712 \times 10^{-11} T^{3.0424} + 3.2576 \times 10^{-14} T^{3.7741}\right)$
5	$\begin{cases} 2.1 \times 10^{-20} (T/30)^{-0.15} & T < 30\,\mathrm{K} \\ 10^{(-18.20 - 3.194 \log_{10}(T) + 1.786 (\log_{10}(T))^2 - 0.2072 (\log_{10}(T))^3)} & T \geq 30\,\mathrm{K} \end{cases}$
6	6×10^{-10}
7	$(-3.323\,218\,3 \times 10^{-7} + 3.373\,538\,2 \times 10^{-7} \ln T - 1.449\,136\,8 \times 10^{-7} (\ln T)^2 + 3.417\,280\,5 \times 10^{-8} (\ln T)^3 - 4.781\,372\,8 \times 10^{-9} (\ln T)^4 + 3.973\,154\,2 \times 10^{-10} (\ln T)^5 - 1.817\,141\,1 \times 10^{-11} (\ln T)^6 + 3.531\,193\,2 \times 10^{-13} (\ln T)^7) \exp(2.123\,715\,0 \times 10^4 / T)$
8	$35.5 T^{-2.28} \exp[-46707/T]$
9	$4.38 \times 10^{-10} T^{0.35} \exp[-102000/T]$
10	Long numerical function; see Martin et al. (1996) and implementation in source code of KROME (Grassi et al., 2014)
11	$\exp(-18.01849334273 + 2.360852208681 \ln T_e - 0.2827443061704 (\ln T_e)^2 + 0.01623316639567 (\ln T_e)^3 - 0.03365012031362999 (\ln T_e)^4 + 0.01178329782711 (\ln T_e)^5 - 0.001656194699504 (\ln T_e)^6 + 0.0001068275202678 (\ln T_e)^7 - 2.631285809207 \times 10^{-6} (\ln T_e)^8)$
12	$\begin{cases} 2.56 \times 10^{-9} T_e^{1.78186} & T \leq 1160\,\mathrm{K} \\ \exp(-20.37260896533324 + 1.139449335841631 \ln T_e \\ \quad - 0.1421013521554148 (\ln T_e)^2 + 0.00846445538663 (\ln T_e)^3 \\ \quad - 0.0014327641212992 (\ln T_e)^4 + 0.0002012250284791 (\ln T_e)^5 \\ \quad + 0.0000866396324309 (\ln T_e)^6 - 0.0000258500968026 4 (\ln T_e)^7 \\ \quad + 2.4555011970392 \times 10^{-6} (\ln T_e)^8 \\ \quad - 8.06838246118 \times 10^{-8} (\ln T_e)^9) \times 10^{-9} T_e^{1.78186} & T > 1160\,\mathrm{K} \end{cases}$
13	$2.96 \times 10^{-6} / \sqrt{T} - 1.73 \times 10^{-9} + 2.50 \times 10^{-10} \sqrt{T} - 7.77 \times 10^{-13} T$
14	$1.0 \times 10^{-8} T^{-0.4}$
15	$10^6 (4.2278 \times 10^{-14} - 2.3088 \times 10^{-17} T + 7.3428 \times 10^{-21} T^2 - 7.5474 \times 10^{-25} T^3 + 3.3468 \times 10^{-29} T^4 - 5.528 \times 10^{-34} T^5)$
16	$5 \times 10^{-7} \sqrt{10^2 / T}$
17	$6 \times 10^{-32} T^{-0.25} + 2 \times 10^{-31} T^{-0.5}$
18	$(6 \times 10^{-32} T^{-0.25} + 2 \times 10^{-31} T^{-0.5})/8$

№	Reaction Rate Coefficient k_i
19	$k_H^{1-a} k_L^a$, where $k_L = 1.18 \times 10^{-10} \exp(-6.95 \times 10^4 / T)$; $k_H = 8.125 \times 10^{-8} T^{-0.5} \exp(-5.2 \times 10^4 / T)[1 - \exp(-6.0 \times 10^3 / T)]$; $a = (1 + n_H / n_{\rm cr})^{-1}$; $\log_{10} n_{\rm cr} = 4.845 - 1.3 \log_{10}(T/10^4) + 1.62 \left[\log_{10}(T/10^4)\right]^2$
20	$1.69 \times 10^{-10} \exp(-4680/T + 198800/T^2)$
21	$10^{-9} \exp\left(-0.417 + 0.846 \log T - 0.137 (\log T)^2\right)$
22	$5.25 \times 10^{-11} \exp(-4430/T + 173900/T^2)$
23	$1.1 \times 10^{-9} \exp(-448/T)$
24	$2.0 \times 10^{-10} (T^{0.402}) \exp(-37.1/T) - 3.31 \times 10^{-17} T^{1.48}$
25	$2.0 \times 10^{-10} (T^{0.396}) \exp(-33/T) - 2.03 \times 10^{-9} T^{-0.332}$
26	2.08×10^{-9}
27	1.0×10^{-16}
28	$4.6 \times 10^{-6} T^{-0.65}$
29	$7.7 \times 10^{-9} \exp(-17560/T)$
30	$1.20 \times 10^{-15} (T/300)^{0.25}$
31	$\begin{cases} 1.26 \times 10^{-9} T^{-0.75} \exp(-1.275 \times 10^5 / T) & T \leq 1000\,\mathrm{K} \\ 4 \times 10^{-37} T^{4.74} & T > 1000\,\mathrm{K} \end{cases}$
32	$\alpha^B_{\rm HII} = 2.753 \times 10^{-14} (315614/T)^{1.5} / (1 + (315614/2.74T)^{0.407})^{2.242}$
33	$\alpha^B_{\rm HeII} = 1.26 \times 10^{-14} (570670/T)^{0.75}$
34	$\alpha^B_{\rm HeIII} = 5.506 \times 10^{-14} (1263030/T)^{1.5} / (1 + (1263030/2.74T)^{0.407})^{2.242}$
35	$3.0 \times 10^{-16} (T/300)^{0.95} \exp(-T/9320)$
...	All further reactions are photoabsorptions, which have their cross sections integrated on-the-fly from the tables of Heays, A. N. et al. (2017) and Huebner & Mukherjee (2015).

To capture the self-shielding of H_2, instead of the scheme of Gnedin et al. (2009), I use that of Nickerson et al. (2018), which tries to capture the physics on the scale of cells, rather than H_2 columns. Here, the rate of photon destruction by H_2 dissociation in the Lyman-Werner band is multiplied by a constant factor $S > 1$ (I use $S = 400$ as recommended by the authors), but not the corresponding dissociation rate of H_2 itself. That is, more photons

are destroyed than there are dissociations. Indeed, this is true in reality, with only $\sim 10\%$ of absorptions leading to dissociations (Stecher & Williams, 1967). In this prescription, columns or clouds of gas must be resolved by multiple cells in order to self-shield: the illuminated outer cells experience the normal dissociation rate, but light in the LW band is destroyed much more strongly, shielding the inner cells. A quantitative comparison of the schemes of Gnedin et al. (2009) and Nickerson et al. (2018) follows in the next chapter.

In this thermochemical scheme, the chemistry and photon concentration update steps are operator split from the thermal and photon flux updates. That is, in a single thermochemical update (which is itself split from the photon injection and transport steps), first the concentrations of chemical species (n_i) and photons (N_i) are updated using DLSODES, an implicit, adaptive, sparse-matrix ODE solver from ODEPACK (Hindmarsh, 1983), that selects its time step duration automatically in order to achieve a user-specified error tolerance. Thus, the Δt provided to DLSODES can simply be the RAMSES-RT global time step (that of the hydrodynamics, gravity and RT transport), and DLSODES will automatically perform the correct number of internal steps to evolve the system by the global Δt. Being implicit, the strong coupling between the concentrations of chemical species is captured without the need for excessively small time steps, and the use of a sparse-matrix representation improves speed where the derivative matrix contains many zeros (as is the case here) by only operating on the nonzero values.

After the chemical and photon density update, the temperature and photon fluxes are updated using the semi-implicit scheme of RAMSES-RT. This is done for two reasons, firstly because keeping the temperature constant during the chemistry step allows rates to be calculated once for the global step, rather than every internal step, which greatly improves performance, and secondly because there are 21 flux variables (1 for each spatial dimension, for each photon group), which would significantly increase the size of the system of ODEs to solve (and thus the computational cost), so these are updated separately (which is not a great concern, as their ODEs do not contain source terms from stars or recombination, and their values do not appear in the ODEs of any chemical species, so coupling is weak and one-way when split from the transport step). To ensure that coupling is properly captured when the cooling/heating time is short, I use the same 10% rule

as Ramses-rt, restarting the step with half the duration if the temperature changes too quickly (which was found in testing to outperform the fully-implicit update where T was allowed to vary).

2.3 Halo Finding

It can be useful to identify haloes in simulation outputs, in order to analyse galaxies separately from their backgrounds, and understand the formation of their structure resolved in time and space. In this thesis, an implementation (Tweed et al., 2009) of the AdaptaHOP algorithm (Aubert et al., 2004) is used to find dark matter haloes, which can also identify hierarchies of substructure within host haloes. TreeMaker (Tweed et al., 2009) is then used over many simulation snapshots to construct *merger trees*, which, for a given resultant halo, contain the haloes of each snapshot that contributed to its assembly. Thus, with a merger tree, one can study the history of all the matter that eventually comes together to form a galaxy.

AdaptaHOP extends the HOP algorithm (Eisenstein & Hut, 1998), which identifies local maxima in the dark matter density field, and the set of particles around each maximum that contribute to a cosmological overdensity. This is done by assigning a 'density' to each particle, calculated by integrating the mass in the neighbourhood of the particle through a smoothing kernel function. Beginning at a particle, the density of it and its neighbours are calculated this way, then a path is traced to the densest of the neighbours, repeatedly until a particle is found that is its own densest neighbour. All particles that can trace paths in this manner to the densest particle are considered members of the same halo, where the densest particle is considered the centre. If many such paths can be drawn, particles are assigned to belong to the halo whose centre is closest to them. A density threshold is applied such that only particles above the threshold may be considered members of a halo.

AdaptaHOP then considers the saddle points of the density field, that connect local maxima to each other; if two local maxima are connected along a saddle point, one may be considered a substructure of the other. The higher the density of the local maximum,

the deeper in the tree of structure the halo is considered to be (i.e. substructures have higher densities than their host haloes). The densities of the connecting saddle points are used as thresholds to separate particles between substructure and host halo, such that the particles within a given substructure have densities bounded by the minimum and maximum densities of saddle points connected to it. In the tree constructed by this method, only the leaves may be considered physical haloes, as non-leaf nodes of the tree (host haloes) have their most dense particles artificially removed and assigned to deeper nodes. Thus, further processing is required to fully identify both host haloes and their intermediate substructure (that is then host to the deepest substructure — leaves of the tree). Multiple methods exist to do this, but the work in this thesis uses the *most massive sub-node method* (MSM). Here, the host halo is chosen to be the union of the particles contained in the branch of the tree containing both the most massive leaf and the root node, thus grouping particles across the entire range of densities, and assigning a section of the tree to a physical object. This is applied recursively, so the branch (of the tree that remains after the first step, so this branch will not extend back to the root) that contains the next most massive leaf node is a physical sub-halo of the host, and so on until the entire tree is split up into branches (or solitary leaf nodes) corresponding to physical haloes.

Once structures have been identified in many simulation snapshots, a merger tree can be created. TreeMaker does this by identifying haloes that share particles in common between snapshots, so that, for a given original halo in a snapshot and a set of haloes in a later snapshot, the halo in the later snapshot with the most mass of particles in common with the original is considered to be the successor. Haloes may only have one successor, but many progenitors, thus forming a tree of haloes connected between snapshots. It is often useful to consider the properties of the *main branch* of a root halo, containing only the most massive progenitor at each level of the merger tree in prior snapshots.

In high resolution simulations of the early universe, dark matter haloes are not always fully virialised before baryonic physics of interest begins to happen (e.g. star formation), which presents a problem for halo finders, as dark matter structure is much less smooth and more fragmented. This can lead to haloes being identified as separate when they are in fact already gravitationally bound together, leading to what may look like a very

steep, or even non-monotonic, mass history of the main branch of progenitors. This can be overcome by choosing different criteria for selecting the main progenitor at each level of the tree; for example, one could insist that the main progenitor is the halo with the most star particles in common with the successor, rather than dark matter particles, or that the main progenitor is simply the most massive of all the possible progenitors (of the haloes whose dark matter particle mass shared with the successor is above a certain threshold). Gas is also not necessarily perfectly local to not-yet-virialised dark matter haloes, so adjustment of the halo centre can be made by also considering either the density or centre of mass of gas (or stars) within the bounds of the dark matter halo. The virial radius of the halo can also be adjusted in a similar manner, increasing it to be consistent with the total contained mass including gas and stars, rather than just the dark matter.

Chapter 3

Testing the Non-equilibrium Chemistry

My custom thermochemistry implementation (described in Section 2.2.8) uses a fundamentally different numerical scheme to the well-tested implementation of thermochemistry in RAMSES-RT; thus it has been subject to testing to show that it produces similar results in physically identical situations, and agrees with analytic calculations (where such predictions can be made). In this chapter I present a collection of tests of my modified version of RAMSES-RT that demonstrate the accurate two-way coupling of the chemical network to the radiation field and the hydrodynamic state of the gas. I also compare numerical H_2 self-shielding prescriptions and the effect of varying metallicity on gas collapse with my chemical network.

3.1 Strömgren Spheres

The *Strömgren Sphere* (Strömgren, 1939) is a classic study of astrophysical radiative transfer, in which a star emits a constant rate of ionising photons $\dot{N}$ into a homogeneous, static medium of pure atomic hydrogen with number density n. Assuming a constant final temperature, an equilibrium solution can be found analytically by recognising that a sphere of ionised gas will initially expand from around the star and eventually reach a radius r_S (the *Strömgren radius*), at which the rate of photoionisation in the sphere is equal to the rate of recombination, and thus stop expanding. Recombinations to the ground state of H

emit a photon that can go on to ionise another hydrogen atom, and so these do not contribute to the total recombination rate; that is, only case B recombinations (in which the free electron is bound into any state but the ground) should be considered. If n_e and n_p are the electron and proton number densities, $\alpha_{\rm B}$ is the case B recombination rate (such that $\alpha_{\rm B} n_e n_p$ is the number of recombinations per unit volume), and it is assumed that ionising photons are instantly absorbed by neutral hydrogen atoms (so that the I-front is infinitely sharp), we can write

$$\dot{N} = \alpha_{\rm B} n_e n_p \left(\frac{4}{3}\pi r_S^3\right), \tag{3.1}$$

which says that the rate of emission of ionising photons (LHS) is equal to the recombination rate (RHS). This equation defines the Strömgren radius r_S (as seen previously in Section 2.1.6)

$$r_S = \left(\frac{3\dot{N}}{4\pi\alpha_{\rm B} n^2}\right)^{\frac{1}{3}}, \tag{3.2}$$

where, in fully ionised gas, n_e and n_p are both equal to the number density of hydrogen atoms n in the initially neutral medium.

However, in reality the ionisation front is not infinitely sharp, as the photons are not instantly absorbed in neutral gas, but instead stream some distance and are eventually absorbed due to random collisions with hydrogen atoms. This random interaction is characterised by the *cross section* σ, defined such that probability of a photon causing an ionisation in a slab of neutral hydrogen of number density n is the same as that of a ray intersecting a sphere of geometric cross section $\sigma = \pi r^2$ in a slab of these spheres uniformly distributed with number density n. In a slab of thickness δl, the fraction of area of an incident beam obscured by target spheres, and thus the probability of intersecting an incident ray, is

$$n\sigma\delta l, \tag{3.3}$$

assuming negligible overlap of spheres (i.e. $n\sigma\delta l << 1$). If N such photons illuminate the

slab, the change in photon number due to ionisations is

$$\delta N = -Nn\sigma\delta l. \tag{3.4}$$

Taking the infinitesimal limit of δl, we have an ODE for the photon count as the beam passes through a slab, at any distance l into the slab

$$\frac{dN}{dl} = -Nn\sigma. \tag{3.5}$$

If there are N_0 incident photons, this ODE has solution

$$N(l) = N_0 e^{-n\sigma l}. \tag{3.6}$$

Thus, ionising photons passing a distance l through a medium of neutral hydrogen of density n have their density reduced by a factor $e^{-n\sigma l}$, in which the exponent $n\sigma l$ is generally called the *optical depth*, denoted τ. More generally, if n varies with l, the solution is

$$N(l) = N_0 e^{-\int_0^l n(l')\sigma dl'}, \tag{3.7}$$

now with $\tau = \int_0^l n(l')\sigma dl'$.

Consider a spherical shell of gas at radius r around the star with infinitesimal thickness δr, in which a fraction $\eta(r)$ of its atoms are neutral (so the density of neutral atoms is $n\eta(r)$). In a given unit time, the number of ionisations in this shell is the product of the number of incident photons $N(r)$ with the probability that a photon is absorbed in the shell $n\sigma\eta(r)\delta r$. Since the star emits $\dot{N}$ photons per second, $N(0) = \dot{N}$. Reusing the derivation above, $N(r)$ obeys

$$N(r) = \dot{N}e^{-\tau}, \tag{3.8}$$

where $\tau = n\sigma \int_0^r \eta(r')dr'$.

The recombination rate over the whole volume of the shell is

$$n_e n_p \alpha_{\mathrm{B}} \cdot 4\pi r^2 \delta r = \left[(1-\eta(r))n\right]^2 \alpha_{\mathrm{B}} \cdot 4\pi r^2 \delta r, \tag{3.9}$$

so the equation describing the ionisation-recombination balance in equilibrium at radius r is

$$\dot{N}\eta(r) n\sigma e^{-\tau} \delta r = \left[(1-\eta(r))n\right]^2 \alpha_{\mathrm{B}} \cdot 4\pi r^2 \delta r, \tag{3.10}$$

$$\implies \frac{\dot{N}}{4\pi r^2}\eta(r) n\sigma e^{-\tau} = \left[(1-\eta(r))n\right]^2 \alpha_{\mathrm{B}}. \tag{3.11}$$

To determine the neutral fraction at equilibrium $\eta(r)$, this equation can be solved numerically, which is most easily done after transforming it into an explicit ODE. In order to do this we first define the dimensionless radius $w \equiv n\sigma r$, so that

$$\tau = \int_0^w \eta(w') dw', \tag{3.12}$$

and then $\eta(w) = \frac{d\tau}{dw}$ from the fundamental theorem of calculus, so we have a quadratic ODE in τ

$$\frac{\dot{N} n\sigma^3}{4\pi w^2} \frac{d\tau}{dw} e^{-\tau} = \left(1 - \frac{d\tau}{dw}\right)^2 \alpha_{\mathrm{B}}. \tag{3.13}$$

In order to allow a simple choice of branch when the roots are taken, we define $y \equiv w - \tau$, so that the ionised fraction $\chi(w) \equiv 1 - \eta = \frac{dy}{dw}$, and the equation becomes

$$A \frac{e^{y-w}}{w^2} \left(1 - \frac{dy}{dw}\right) = \left(\frac{dy}{dw}\right)^2 \alpha_{\mathrm{B}}, \tag{3.14}$$

where we have defined the constant

$$A \equiv \frac{\dot{N} n\sigma^3}{4\pi}. \tag{3.15}$$

This quadratic equation has roots

$$\frac{dy}{dw} = -\frac{Ae^{y-w}}{2w^2} \pm \sqrt{\left(\frac{Ae^{y-w}}{2w^2}\right)^2 + \frac{Ae^{y-w}}{w^2}}, \tag{3.16}$$

where the second term on the RHS is strictly larger in magnitude than the first (for $w > 0$), so the + root is strictly positive and the − root is strictly negative. Since the ionised fraction $\chi(w) = \frac{dy}{dw}$ must be ≥ 0, we can discard the negative root, leaving a single branch in explicit form

$$\frac{dy}{dw} = -\frac{Ae^{y-w}}{2w^2} + \sqrt{\left(\frac{Ae^{y-w}}{2w^2}\right)^2 + \frac{Ae^{y-w}}{w^2}}. \tag{3.17}$$

This is an initial value problem that can be solved numerically by standard methods (e.g. the Runge-Kutta methods mentioned in Chapter 2; Runge 1895; Kutta 1901), with the initial condition $\chi(0) = 1$ (full ionisation adjacent to the star). Figure 3.1 shows solutions to Equation 3.17, that demonstrate the equilibrium ionisation profile converging towards Strömgren's discontinuous step solution as the cross section increases, as expected. The lower the cross section, the less sharp the ionisation front, as radiation penetrates further into the neutral gas. If the radius r_I of the front is defined such that $\chi(r_I) = 0.5$, it can be seen from Figure 3.1 that r_I lies ahead of the Strömgren radius at equilibrium (increasingly so for smaller cross sections). The average cross section of absorption of black-body radiation between 13.6 eV and 24.59 eV (the lowest σ used in the figure) is lower than that of monochromatic 13.6 eV radiation, as the monochromatic photons are perfectly resonant with the ionisation transition (the *Lyman limit*, at 13.6 eV).

The solutions of Equation 3.17 enable a basic test of 3D numerical radiative transfer codes: the ionisation profile of gas in a simulation should match the 1D solution at equilibrium. In the rest of this section I verify that my modified version of Ramses-rt achieves this, and present other Strömgren Sphere inspired tests of the new photochemistry, including a comparison of prescriptions of H_2 self-shielding. Similarly to the testing of photoionisation in Ramses-rt by Rosdahl et al. (2013), I use Strömgren-inspired test cases selected from Iliev et al. (2006) (purely testing radiative transfer and thermochemistry) and Iliev et al. (2009) (which introduce hydrodynamics).

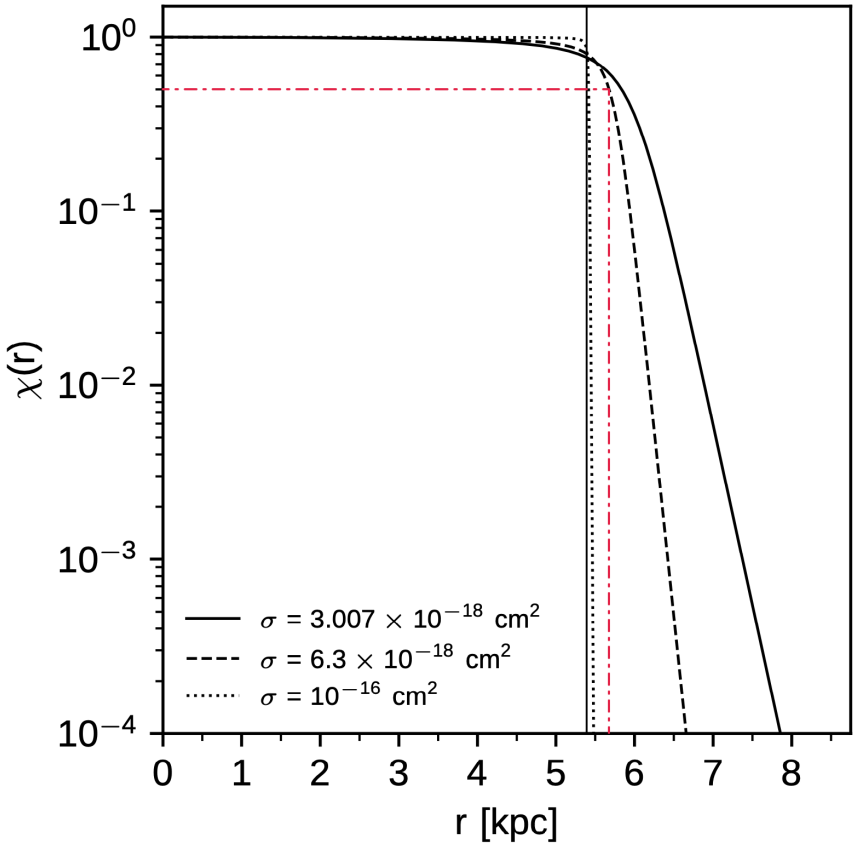


Figure 3.1: Equilibrium ionised gas fractions $\chi(r)$ (calculated by solving Equation 3.17 with the RK45 solver (a 4$^{\text{th}}$ order explicit Runge-Kutta scheme for ODE integration) from the SciPy package (Jones et al., 2001)) around a star with $\dot{N} = 5 \times 10^{48} s^{-1}$ in homogeneous neutral hydrogen with number density $n = 10^{-3}$ cm^{-3}, for three different ionisation cross sections, with the Strömgren radius $r_S = 5.4$ kpc marked by the vertical solid line. 3.007×10^{-18} cm^2 is the average cross section of photons emitted by a 10^5 K black-body between 13.6 eV and 24.59 eV (a close match to the situation in simulations shown later in this chapter); 6.3×10^{-18} cm^2 is the cross section of monochromatic 13.6 eV photons; 10^{-16}cm^2 is chosen arbitrarily to illustrate that the solution tends toward a step function (Strömgren's solution) as the cross section increases. The red dash-dotted line marks the position of the front as defined by $\chi(r_I) = 0.5$.

3.1.1 Tests in Static Atomic Hydrogen

Strömgren's original calculation assumes a static medium of constant density and temperature. Here I verify that my implementation of primordial photochemistry reproduces the analytic predictions shown earlier in this section under these conditions, and is consistent with RAMSES-RT. I also test the non-equilibrium behaviour, checking that the I-front expands and relaxes to equilibrium in the expected time. According to Shapiro et al. (2006) (as seen in Section 2.1.6.2), the radius of the I-front r_I should progress according to

$$\frac{t}{t_{\rm rec}} = \frac{r_I}{ct_{\rm rec}} - \ln\left[1 - \left(\frac{r_I}{r_s}\right)^3\right], \tag{3.18}$$

(where $t_{\rm rec} = \frac{1}{n\alpha_{\rm B}}$) which, in the limit of a low light crossing-time to recombination time ('slow' recombination — which is the case in this test), becomes

$$\frac{r_I}{r_S} = \left(1 - e^{t/t_{\rm rec}}\right)^{\frac{1}{3}}. \tag{3.19}$$

This assumes that the ionisation front is infinitely sharp, and thus fails to account for the extended front that is exhibited in solutions to Equation 3.11. As discussed in the previous section, if the front radius r_I is defined such that $\chi(r_I) = 0.5$, then it would be expected that a realistic front would be slightly ahead of the prediction of the equation above (Pawlik & Schaye 2008; also evident in the solutions to Equation 3.17 in Figure 3.1).

3.1.1.1 Test 1: Monochromatic Source in Static Gas at Constant Temperature

In the first test, a source emitting monochromatic 13.6 eV photons at the rate of $\dot{N} = 5 \times 10^{48}\ {\rm s}^{-1}$ is placed in the corner of a uniform 3D grid of 128^3 cells of width $L = 6.6\,{\rm kpc}$, containing a medium of atomic hydrogen of number density $n = 10^{-3}\ {\rm cm}^{-3}$ and temperature $T = 10^4\ {\rm K}$, in collisional ionisation equilibrium (so that the initial ionised fraction is 1.2×10^{-3}), as in "Test 1" of Rosdahl et al. (2013). Taking $\alpha^B_{\rm HII}(T)$ from reaction 33 of my network (Table 2.2), the Strömgren radius is $r_S = 5.4\,{\rm kpc}$.

Here, every photon is capable of ionising a hydrogen atom, and the monochromatic

spectrum is fully represented, without approximation, by a single photon group with ionisation cross-section $\sigma = 6.3 \times 10^{-18}$ cm^2. Thus, any discrepancy between the simulation's prediction and the equilibrium profile given by Equation 3.11 (which assumes the same constant cross-section for all ionising photons) should be due to Ramses-rt's radiation transport method (which, under the M1 closure, has an approximate form for the pressure tensor). The M1 closure is the least accurate in regions of intermediate opacity, whereas it is exact when the medium is optically-thick (and the radiation is near-isotropic), or when the medium is optically-thin (and the radiation is collimated), so the code is not expected to perfectly resolve radiation around the ionisation front (in which gas transitions from opaque where it is neutral, to transparent where it is ionised).

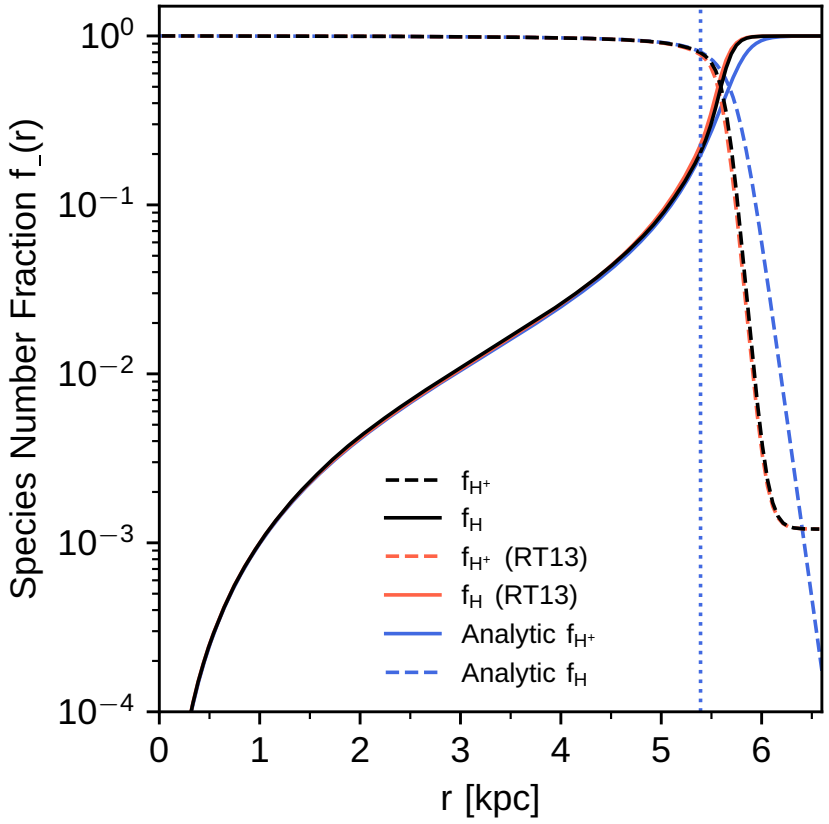


Figure 3.2: *Test 1:* Equilibrium ionised (dashed) and neutral (solid) spherically-averaged gas profiles produced by my modified version of Ramses-rt (black), the public version of Ramses-rt (red, labelled RT13), and the analytic prediction (blue). The Strömgren radius r_S is marked by the vertical blue dotted line. The numerical codes predict a slightly sharper front than the analytic result, but otherwise agree very closely with each other.

Figures 3.2 and 3.3 show the ionisation profile at equilibrium, and the time evolution of r_I respectively. These match the analytic predictions quite well, except that the I-front is slightly sharper than expected at equilibrium, and $r_I(t)$ is slightly ahead of the analytic prediction at all times. The simulated solution also plateaus at $f_{H^+} = 1.2 \times 10^{-3}$ as it is initialised in collisional equilibrium. The overestimated sharpness of the front is most likely due to the approximate nature of the radiation transport scheme, as this region of intermediate opacity (where light is neither isotropic nor perfectly collimated) is the most challenging for the M1 closure. The simulated $r_I(t)$ lies ahead of the analytic prediction, as Equation 3.19 assumes an infinitely sharp front, but the simulation uses a more realistic finite cross section, so it is extended. Defining the position of the extended front as $\chi(r_I) = 0.5$, r_I is expected to lead the analytic prediction by a small amount, as mentioned previously. An initial delay in the I-front expansion is expected, due to the reduced speed of light approximation (where I use $\bar{c} = 0.01c$), but the I-front spends most of its time in the 'slow' phase (low q factor, see Section 2.1.6.2), so the reduced speed of light approximation (RSLA) does not have a large impact here. The light crossing time of the system is ~20 kyr, but the earliest simulation output from this test is at 10 Myr, so the 'fast' phase is not visible here.

The predictions of my code are very close to those of Ramses-rt. This is expected, since the only difference between the codes is in the numerical scheme for the thermochemical step; they otherwise use the same radiation group structure, ionisation cross-section, recombination rates, and radiation transport method (the added H_2 chemistry has no impact in this test, since the gas is purely atomic, and the formation rate is negligible at this density). Ramses-rt's semi-implicit scheme ensures convergence in this step to the point where the change in the ionisation fraction per step is $\sim 10\%$, and, similarly, the adaptive solver dlsodes used in my code automatically chooses its time step to ensure a relative tolerance (the ratio of the error estimated by solver to the magnitude of the solution) of 10^{-4} is reached. These tolerances could potentially be tuned to ensure each code converges more closely to the same solution.

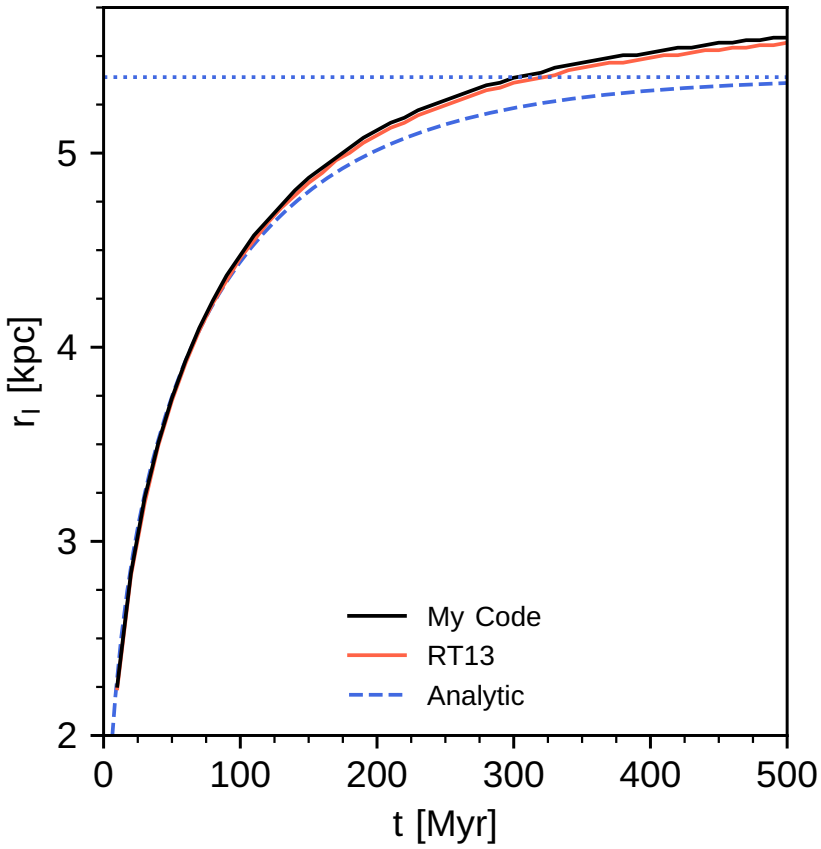


Figure 3.3: *Test 1:* Ionisation front radius r_I as a function of time, for my modified version of RAMSES-RT (black), the public version (red), and the prediction of Equation 3.19 (blue dashed). The analytic prediction lags behind the numerical result when the front is not sharp, if the front position r_I is defined such that $\chi(r_I) = 0.5$. If another ionisation threshold were chosen, r_I could be moved arbitrarily to any point on the extended front. The solution of Equation 3.19 converges to the Strömgren radius, marked by the horizontal blue dotted line. The slight difference between the two codes may be attributed to the difference in the numerical scheme used in the thermochemical step. Small differences in the ionisation profile in the front may lead to relatively large differences in the position r_I, since the front is steep, so the discrepancy observed here is not severe.

3.1.1.2 Test 2: 10^5 K source with Thermal Evolution

In the next test, the grid is set up with the same dimensions and resolution as in Test 1, but the temperature is allowed to vary, accounting for the full suite of heating and cooling processes (Section 2.2.2). The initial temperature is $T_0 = 100$ K, and a 10^5 K black body emitter is used instead of a monochromatic source (with $\dot{N} = 5 \times 10^{48}$ s^{-1}), as in "Test 2" of Rosdahl et al. (2013). Unlike in Test 1, the medium is initially purely neutral, rather than in collisional equilibrium. Three radiation groups are used with the public version of Ramses-rt (with bin edges on each of the ionisation energies of hydrogen and helium), but now the full 7-group structure (Table 2.1) is used with my code. This means results will differ more significantly from Ramses-rt than in the previous test, but it is important to test that discretising the spectrum more finely, and including more photoreactions, has not adversely affected results.

The analytic predictions of Equations 3.11 and 3.19 do not apply to a medium with changing temperature, as the recombination coefficient α_B varies with temperature, and cooling via non-ionising photons is allowed. However, even if verification against analytic results is not possible, a close match with Ramses-rt is reassuring, since this has been tested favourably against other radiation transport codes in Rosdahl et al. (2013) (via Iliev et al. 2006, 2009). Figure 3.4 shows that the results of my code closely match those of Ramses-rt, with slight differences due to the extended group structure in my code, and the use of dlsodes instead of the semi-implicit scheme for the thermochemical step.

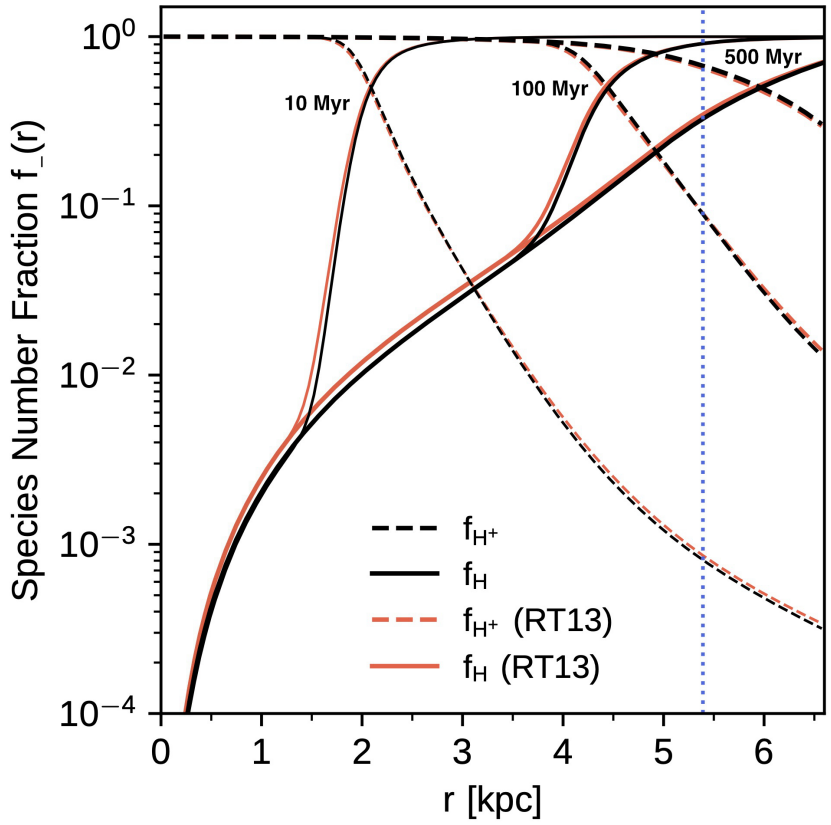


Figure 3.4: *Test 2:* Non-equilibrium ionised (dashed) and neutral (solid) spherically-averaged gas profiles produced (without hydrodynamics) by my modified code (black), and the public version of RAMSES-RT (red, labelled RT13), at 10 Myr (thinnest), 100 Myr and 500 Myr (thickest). The Strömgren radius r_S is marked by the vertical blue dotted line. The difference between my code and RAMSES-RT is barely visible in the ionised fraction, but appears to be more significant in the neutral fraction due to the use of a log scale; these differences are of the order $\sim 10^{-3}$, and are more strongly present in the rapidly ionising region, rather than the tail extending into the neutral gas.

3.1.2 Tests in Non-Static Atomic Hydrogen

Once hydrodynamics is accounted for, the behaviour the I-front changes. Ionising photons may impart kinetic energy when absorbed by neutral hydrogen atoms if their energy E is greater than the ionisation energy $E_0 = 13.6\,\text{eV}$, which means that the gas around a star can be heated, driving it through thermal pressure, as well as being ionised, depending on the source spectrum and conditions in the medium. The dynamics of the I-front can be found by considering the mass, momentum and energy conservation across it, in the frame comoving with the front, assuming that the gas ahead is fully neutral, and the gas

behind is fully ionised (Kahn 1954; these are the Rankine-Hugoniot conditions, seen in Section 2.1.2, but now with a source term due to energy absorbed from ionising photons):

$$\rho u = \rho_i u_i, \tag{3.20}$$

$$p + \rho u^2 = p_i + \rho u_i^2, \tag{3.21}$$

$$e + \frac{p}{\rho} + \frac{1}{\mu m_H}\left(\bar{E} - E_0\right) = e_i + \frac{p_i}{\rho_i}, \tag{3.22}$$

where quantities subscripted with i are associated with the ionised gas behind the front, and the non-subscripted quantities with the neutral gas. μ is the average mass of particles ahead of the I-front in atomic units, which is 1 for pure, neutral atomic hydrogen. $\left(\bar{E} - E_0\right)$ is the average kinetic energy imparted by a single photoionisation, in which $\bar{E}$ is the average energy of photons participating in photoionisations. The hydrogen (both ionised and neutral) can be considered to be an ideal gas with adiabatic index $\gamma = \frac{5}{3}$, so

$$\rho\left(e - \frac{1}{2}u^2\right) = \frac{p}{\gamma - 1}, \tag{3.23}$$

and the energy equation becomes

$$\frac{5p}{2\rho} + \frac{1}{2}u^2 + \frac{1}{m_H}\left(\bar{E} - E_0\right) = \frac{5p_i}{2\rho_i} + \frac{1}{2}u_i^2. \tag{3.24}$$

Given the initial state of the neutral gas around the star, Equations 3.20, 3.21 and 3.24 can be solved for the state behind the front. These can be reduced to a single quadratic equation in u, and it can be shown that real-valued solutions only exist for values of u above a critical speed $u_R > c$, or below another critical speed $u_D < c < u_R$, where c is the speed of sound in the neutral medium. If the medium is initially static, u is also the speed of the I-front, unless the I-front becomes subsonic and a shock has been able to propagate ahead, causing $u \neq 0$ when the I-front arrives. The ionising flux $J \equiv \frac{\dot{N}}{4\pi r^2}$ is the only possible source of ions behind the front (disregarding collisional ionisation by e.g. a shock), thus one can write

$$\rho u = \rho_i u_i = m_H J. \tag{3.25}$$

J therefore relates the critical speeds to critical densities at a given radius

$$\rho_R(r) = \frac{m_H J(r)}{u_R}, \tag{3.26}$$

$$\rho_D(r) = \frac{m_H J(r)}{u_D}. \tag{3.27}$$

$$\tag{3.28}$$

If an I-front propagates into a medium with $\rho < \rho_R$, it is R-type (*rarefied*), and expands supersonically, dominated by ionisation and recombination, rather than hydrodynamic effects. Since the radiation, rather than gas dynamics, dominates, R-type fronts expand approximately according to Equation 3.19 (Spitzer, 1978).

In typical ISM conditions, the initial expansion of the I-front is R-type, until the radius of the front approaches $r_I \approx r_S$, where the flux J is diminished such that $\rho > \rho_D$ and the front transitions to D-type (*dense*). Then the front is subsonic with respect to the ionised gas behind it, and proceeds outwards (now behind the hydrodynamic shock) approximately according to Spitzer's solution

$$\frac{r_I}{r_S} = \left(1 + \frac{7}{4}\frac{c_I(t - t_S)}{r_S}\right)^{\frac{4}{7}}, \tag{3.29}$$

where c_I is the speed of sound in the ionised region behind the front, and t_S is the time at which the front transitions from R-type to D-type. As opposed to the solution in static gas, once hydrodynamics is included the I-front is allowed to expand beyond r_S, as the ionised gas behind the front is at higher pressure than the neutral medium, so it exerts a force on the neutral gas ahead of the I-front, pushing it away from the source.

In the transitional region where $\rho_R < \rho < \rho_D$, there can be no single front across which ionisation, density, temperature and velocity all jump discontinuously. That is, a shock necessarily separates from the I-front.

3.1.2.1 Test 3: 10^5 K Source with Hydrodynamics

In this test, I use the same initial conditions and source as test 2, but switch on hydrodynamics, as in "Test 5" of Rosdahl et al. (2013), with a larger $L = 15\,\mathrm{kpc}$ box, as the I-front can now expand far beyond the Strömgren radius, so a larger box size is needed to capture it. Figure 3.5 shows that my code closely matches Ramses-rt, and Figure 3.6 shows that the expected behaviour in R- and D-type regimes is approximately reproduced. The front diverges from the exact R-type behaviour before reaching the Strömgren radius, as a hydrodynamic shock begins to separate from the I-front. After the front passes r_S, it proceeds quite closely to Spitzer's prediction, with the shock propagating ahead of the I-front (shown in Figure 3.7).

Exact agreement with the Spitzer solution is not expected, as this fails to account for the fact that the D-type ionisation front passes through material that is already moving due to the shock that preceded it (Raga et al., 2012). As in previous tests, my code closely agrees with Ramses-rt, with small differences due to the different radiation group structure.

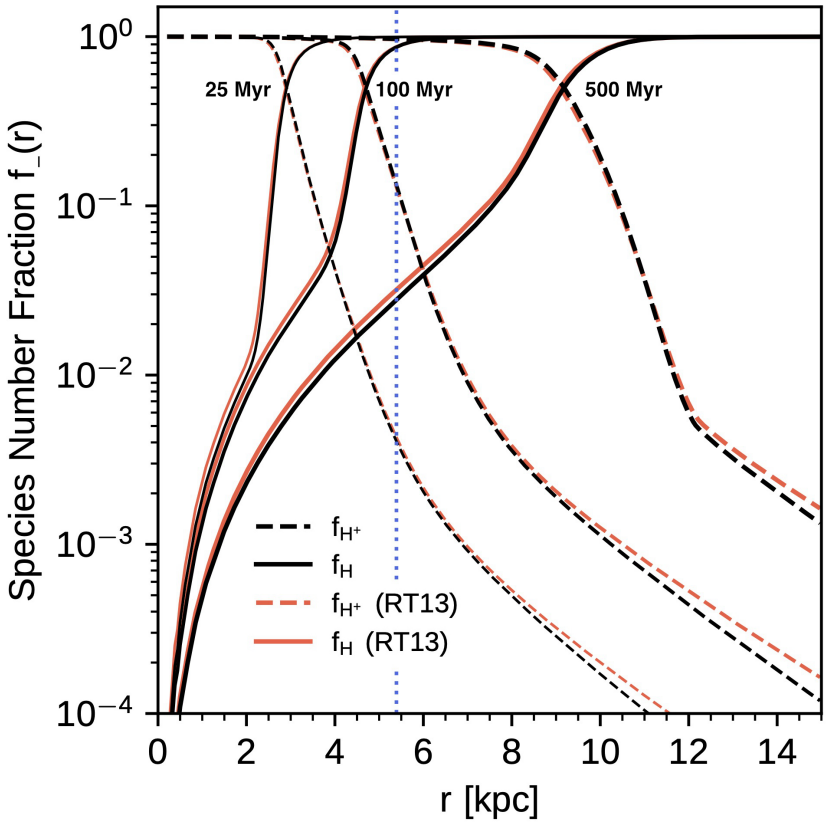


Figure 3.5: *Test 3:* Non-equilibrium ionised (dashed) and neutral (solid) spherically-averaged gas profiles produced (with hydrodynamics) by my modified code (black), and the public version of RAMSES-RT (red, labelled RT13), at 10 Myr (thinnest), 100 Myr and 500 Myr (thickest). The Strömgren radius r_S is marked by the vertical blue dotted line.

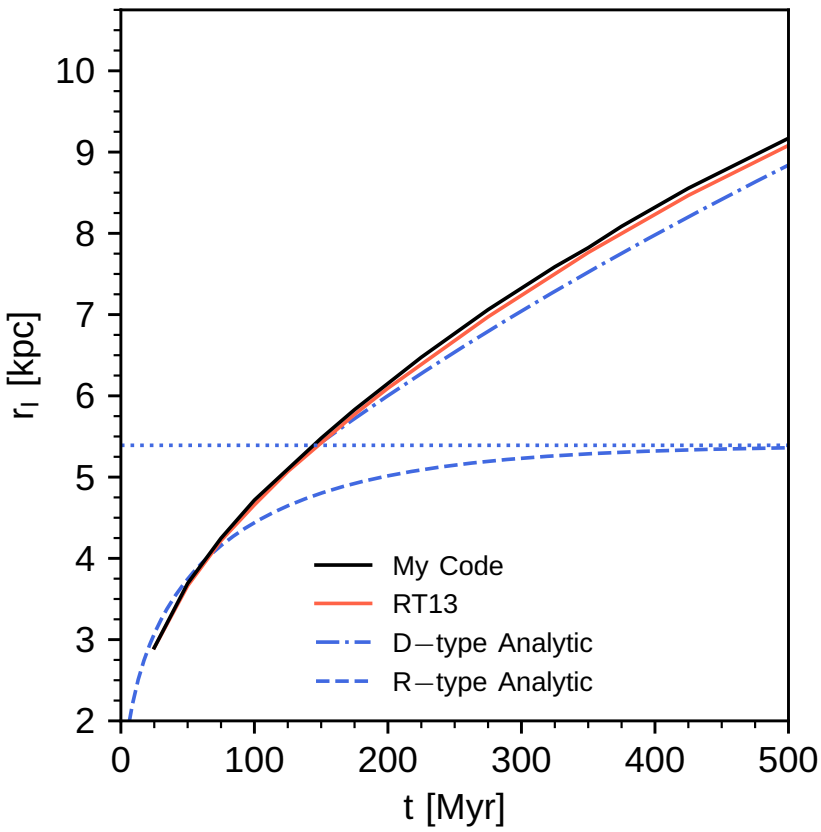


Figure 3.6: *Test 3:* Radius of the ionisation front vs. time for my modified version of RAMSES-RT, and the public version. Analytic solutions are plotted in blue for the R-type and D-type regimes, with a horizontal dotted line marking the Strömgren radius, assuming a 10^5 K medium.

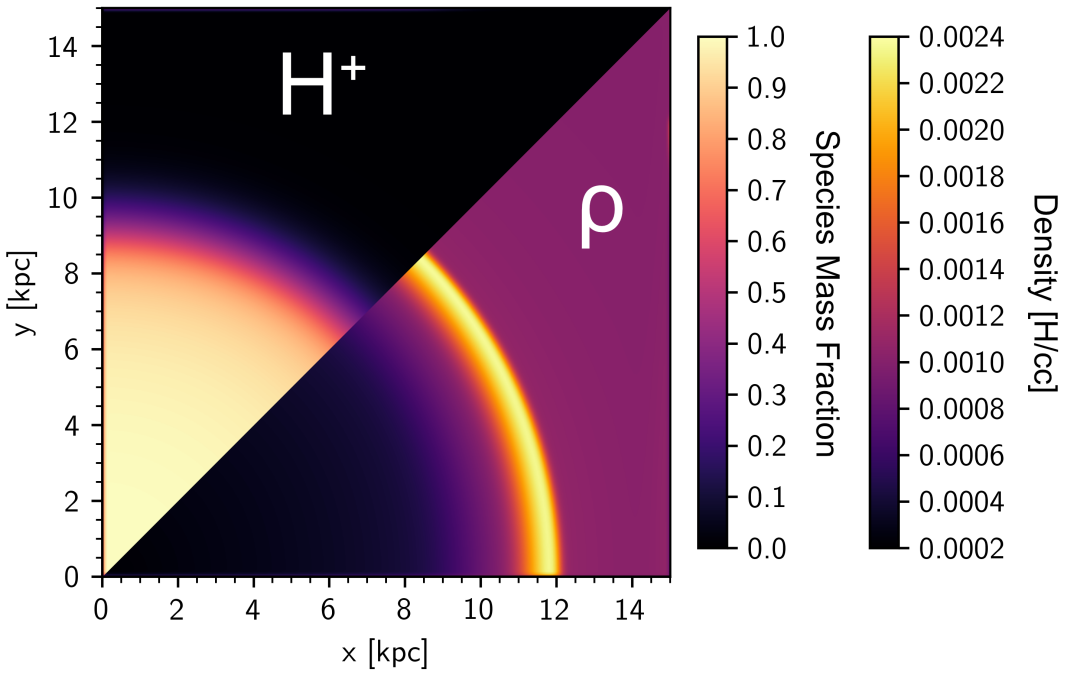


Figure 3.7: *Test 3:* 2D maps of f_H and ρ at $t = 500\,\mathrm{Myr}$. As predicted by the analysis in Section 3.1.2, the hydrodynamic shock (clearly visible in the ρ map, where gas has been swept up by the shock) has propagated ahead of the I-front.

3.1.3 Tests in Molecular Hydrogen

Now I consider a source in a medium of (hydrodynamically static) pure molecular hydrogen in order to test the additions to the photochemical network involved in forming, dissociating and ionising H_2. Strömgren's reasoning can be extended to a molecular medium by considering the number of dissociating photons and formation rate of H_2, similarly to the ionisation and recombination of atomic hydrogen (Nickerson et al., 2018). An important difference between the treatment here and that of Strömgren is that now neutral hydrogen only exists behind the dissociation front, so ionisation can only occur after dissociation; similarly, H_2 reformation can only occur in the neutral gas; thus, the ionisation and dissociation problems are tightly coupled.

As illustrated in Figure 3.8, a dissociation front initially expands through the medium, leaving a band of atomic hydrogen behind it which is then itself ionised. The ionised region expands out to the Strömgren radius, and the dissociation front also eventually stops at r_{H_2} where the number of dissociating photons emitted per unit time $\dot{N}_{\mathrm{diss}}$ in the sphere balances the number of H_2 molecules formed, $\alpha_{\mathrm{H}_2} n_H^2$ (where α_{H_2} is the formation rate coefficient of H_2 in atomic hydrogen). Since the ionised hydrogen cannot directly form H_2, only the shell of neutral hydrogen outside the Strömgren radius forms H_2. Therefore, with $\dot{N}_{\mathrm{ion}} \leq \dot{N}_{\mathrm{diss}}$ ionising photons emitted per unit time (and assuming that neutral H, H^+ ions and free electrons do not interact with H_2-dissociating photons), we have

$$\dot{N}_{\mathrm{ion}} = \alpha_{\mathrm{B}} n_H^2 \left(\frac{4}{3} \pi r_S^3 \right), \tag{3.30}$$

$$\dot{N}_{\mathrm{diss}} = \alpha_{\mathrm{H}_2} n_H^2 \left[\frac{4}{3} \pi \left(r_{\mathrm{H}_2}^3 - r_S^3 \right) \right], \tag{3.31}$$

where n_H is the number of hydrogen nuclei in the medium (equal to the number of protons or electrons in the ionised gas, and double the number of molecules in the surrounding medium). So, assuming infinitely sharp ionisation and dissociation fronts, the equilibrium radius of the dissociation front is

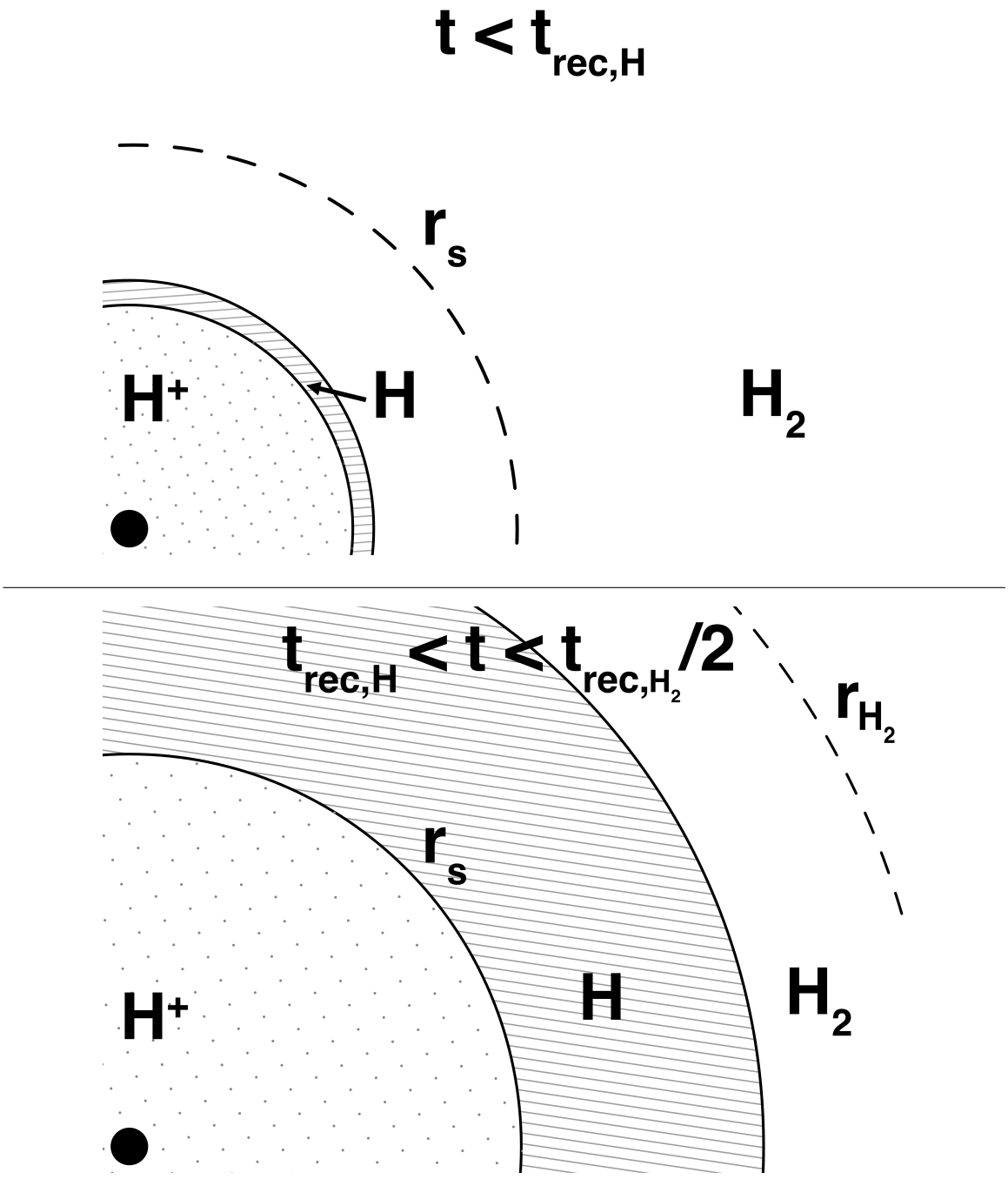


Figure 3.8: Schematic of dissociation front (outer border of the hatched region) and ionisation front (outer border of the dotted region) propagation around a star in an initially pure molecular hydrogen medium. The dotted, hatched and clear regions represent ionised, neutral atomic, and molecular hydrogen respectively. The Strömgren radius is labelled by r_S, and the equilibrium radius for the dissociation front is labelled by $r_{\mathrm{H_2}}$. **Top:** when $t < t_{\mathrm{rec,\,H}}$, the dissociation and ionisation fronts have not yet passed the Strömgren radius, so neutral hydrogen (produced by the dissociation front) is rapidly ionised, only existing in a thin shell between the two fronts. **Bottom:** when $t > t_{\mathrm{rec,\,H}}$, the ionisation front does not progress beyond r_S, and the dissociation front carries on, leaving a large shell of neutral hydrogen behind it; the dissociation front eventually comes to rest at $r_{\mathrm{H_2}}$.

$$r_{\mathrm{H}_2} = \left(r_S^3 + \frac{3\dot{N}_{\mathrm{diss}}}{4\pi\alpha_{\mathrm{H}_2} n_H^2} \right)^{\frac{1}{3}}. \tag{3.32}$$

Similarly to the ionisation front in atomic hydrogen, in the slow-formation regime (analogous to slow-recombination, where the light crossing time is much lower than the recombination time), the radius of the dissociation front $r_D(t)$ decays exponentially toward equilibrium

$$\frac{r_D(t)}{r_{\mathrm{H}_2}} = \left(1 - e^{-2t/t_{\mathrm{rec,H}_2}}\right)^{\frac{1}{3}}, \tag{3.33}$$

where $t_{\mathrm{rec,H}_2} = \alpha_{\mathrm{H}_2} n_H$, and the factor 2 in the exponent is due to the fact that a 'recombination' into H_2 removes 2 hydrogen atoms from the neutral atomic hydrogen region. This follows from assuming that ionising/dissociating photons instantly stream out to the (infinitely sharp) ionisation/dissociation fronts respectively, where they are immediately absorbed by neutral/molecular hydrogen. There are $\dot{N}_{\mathrm{diss}}$ dissociations per second, $\dot{N}_{\mathrm{ion}}$ ionisations per second, the number of H_2 molecules formed per second is $\frac{4}{3}\pi \left(r_D(t)^3 - r_I(t)^3\right) \alpha_{\mathrm{H}_2} n_H^2$, and the number of neutral H nuclei formed by recombination per second is $\frac{4}{3}\pi r_I(t)^3 \alpha_{\mathrm{B}} n_H^2$. Thus, the total number of neutral H atoms $\frac{4}{3}\pi \left(r_D(t)^3 - r_I(t)^3\right) n_H$ obeys

$$\frac{d}{dt}\left[\frac{4}{3}\pi \left(r_D(t)^3 - r_I(t)^3\right) n_H\right] = 2\dot{N}_{\mathrm{diss}} - \dot{N}_{\mathrm{ion}} - 2\cdot\frac{4}{3}\pi \left(r_D(t)^3 - r_I(t)^3\right) \alpha_{\mathrm{H}_2} n_H^2 + \frac{4}{3}\pi r_I(t)^3 \alpha_{\mathrm{B}} n_H^2. \tag{3.34}$$

Assuming that photoionisation equilibrium $\left(\frac{dr_I}{dt} = 0\right)$ is reached quickly compared to the timescale of change of $r_D(t)$, i.e. that $n_H\alpha_{\mathrm{B}} >> n_H\alpha_{\mathrm{H}_2}$, so $r_I(t) = r_S$ immediately, we have

$$\frac{d}{dt}\left[r_D(t)^3\right] = \frac{3\dot{N}_{\mathrm{diss}}}{2\pi n_H} - 2\left(r_D(t)^3 - r_S^3\right)\alpha_{\mathrm{H}_2} n_H. \tag{3.35}$$

This can be solved by separation of variables

$$\int \frac{d(r_D^3)}{\frac{3\dot{N}_{\rm diss}}{2\pi n_H} - 2\left(r_D(t)^3 - r_S^3\right)\alpha_{\rm H_2} n_H} = \int_0^t 1dt', \tag{3.36}$$

$$\implies --\frac{1}{2\alpha_{\rm H_2} n_H} \ln\left[\frac{3\dot{N}_{\rm diss}}{2\pi n_H} - 2\left(r_D(t)^3 - r_S^3\right)\alpha_{\rm H_2} n_H\right] + C = t, \tag{3.37}$$

$$\implies \ln\left[\frac{3\dot{N}_{\rm diss}}{2\pi n_H} - 2\left(r_D(t)^3 - r_S^3\right)\alpha_{\rm H_2} n_H\right] = C' - 2\alpha_{\rm H_2} n_H t, \tag{3.38}$$

$$\implies \frac{3\dot{N}_{\rm diss}}{2\pi n_H} - 2\left(r_D(t)^3 - r_S^3\right)\alpha_{\rm H_2} n_H = Ae^{-2\alpha_{\rm H_2} n_H t}, \tag{3.39}$$

$$\implies r_D(t)^3 - r_S^3 = \frac{3\dot{N}_{\rm diss}}{4\pi\alpha_{\rm H_2} n_H^2} - A'e^{-2\alpha_{\rm H_2} n_H t}, \tag{3.40}$$

and since $r_D(t)^3 = r_S^3 + \frac{3\dot{N}_{\rm diss}}{4\pi\alpha_{\rm H_2} n_H^2} \equiv r_{\rm H_2}^3$ as $t \longrightarrow \infty$, we recover

$$\frac{r_D(t)^3}{r_{\rm H_2}^3} = 1 - e^{-2\alpha_{\rm H_2} n_H t} = 1 - e^{-2t/t_{\rm rec,H_2}}. \tag{3.41}$$

As the formation rate of H_2 is much lower per atom than the recombination rate, if $\dot{N}_{\rm ion} \lesssim \dot{N}_{\rm diss}$, $r_{\rm H_2}$ is much larger than r_S, and the expansion of r_D is faster at all t than r_I, so there is always pure neutral atomic hydrogen just ahead of the ionisation front, and the expansion of $r_I(t)$ is unchanged by the presence of molecular hydrogen in the medium (proceeding according to Equation 3.19), as long as formation and recombination times are slow compared the light crossing time of the system.

3.1.3.1 H_2 Self-Shielding

The molecular nature of H_2 allows it to be excited rotationally and vibrationally (with transitions in the microwave and infrared frequency ranges), and also allows a much greater variety of electronic transitions (in the optical and UV frequency ranges) leading to the presence of a dense collection of lines in the Lyman-Werner (LW) band (11.2 eV to 13.6 eV). LW photon absorptions put a H_2 molecule into an excited state, which only decays into the vibrational continuum (causing dissociation) $\sim$ 10% of the time (Stecher & Williams, 1967) in optically thin gas, whereas the more probable decay channel leads to fluorescence in the infrared, leaving the molecule intact. As the density of H_2 increases in a column,

it becomes fully opaque around line centres, so that only radiation in the 'wings' around each line propagates further into the column. This remaining light in the wings is much less likely to dissociate H_2 molecules when it is absorbed (as it is further from the line centres, and may be absorbed instead by overlapping lines; Abgrall et al. 1992), so the fraction of photons absorbed by the column that lead to dissociations is greatly reduced (steeply and non-linearly as a function of density). As the column density increases further, these absorption lines are broadened, eventually strongly absorbing across the LW band, such that no light in the band can penetrate further into the gas.

Since, in dense clouds of H_2, there are many fewer dissociations than there are photons absorbed in total, these clouds can very efficiently shield themselves from relatively high incident fluxes of LW radiation, staying intact where less dense clouds would have been destroyed. It is therefore highly important to capture this behaviour in simulations of star-forming clouds, but exact treatment of the complex line structure is difficult with Ramses-rt's scheme for radiative transfer, as discretising the spectrum finely enough to resolve the lines would be extremely computationally expensive (and would still fail to resolve the Doppler-shifting and broadening of lines on a per-cell basis). Instead, the entire Lyman-Werner band can be treated approximately with a single radiation group.

As introduced in Section 2.2.7, Draine & Bertoldi (1996) propose a formula for the reduction of the dissociation rate of H_2 by Lyman-Werner photons per molecule in a column of gas with column density N_{H_2}

$$f_{\mathrm{shd}} = \frac{1-\omega}{(1+x)^2} + \frac{\omega}{(1+x)^{1/2}} \exp\left(-8.5\times10^{-4}\sqrt{1+x}\right), \tag{3.42}$$

where $x = N_{H_2}/5\times10^{14}\ \mathrm{cm}^{-2}$. This is the modified form suggested by Gnedin et al. (2009), in which ω is tuned to better fit observations; they find that $\omega = 0.2$ best matches UV observations of the Milky Way and its satellites by the Far Ultraviolet Spectroscopic Explorer and Copernicus orbiting observatory by Wolfire et al. (2008). The original formula of Draine & Bertoldi (1996) differs only in the value of ω used, and the coefficient of x in the first term, and is obtained by a fit to detailed radiative transfer simulations of the

rovibrational and electronic level populations of H_2 in a column illuminated by an interstellar radiation field. The first term captures the first-order behaviour (approximately a power law at high N_{H_2}), and the second term ensures the effective line width does not exceed its approximate maximum value at saturation of the LW band at $N_{H_2} \gtrsim 10^{21}\text{cm}^{-2}$ (without this the dissociation rate is overestimated at these densities). Taking $N_{H_2} \to \infty$, Equation 3.42 returns $f_{\text{shd}} = 0$, as expected — dense columns fully shield the gas behind them from dissociating radiation, so there are close to zero dissociations per molecule; under the optically-thin limit $N_{H_2} \to 0$, $f_{\text{shd}} = 1$, i.e. there is no shielding, as required.

In Eulerian radiative transfer simulations, it is not immediately clear how to define the column density N_{H_2}, since, given a radiation field in a cell, it is not necessarily known which sources it originated from — there is no non-local information available. In my tests, following Katz et al. (2017), I assign a shielding factor to each cell according to Equation 3.42 (taking $\omega = 0.2$), rather than each source (or ray) illuminating a cell, approximating the column density as $N_{H_2} = n_{H_2}\Delta x$ (the column density experienced by a ray travelling a distance Δx through the cell). Other approximations for the column density may be used, usually in the form $N_{H_2} = n_{H_2}L$, where L is an approximation for the scale of self-shielding clouds. These include the "Sobolev-like" treatment of Gnedin et al. (2009), in which $L = \frac{\rho}{|\nabla\rho|}$ (the length scale over which the density changes), or the true Sobolev length $L = \frac{c_s}{|\vec{\nabla}\cdot\vec{u}|}$, or setting L to be the Jeans length (Equation 1.38), all considered by Safranek-Shrader et al. (2017).

Since f_{shd} increases super-linearly with density, Equation 3.42 underestimates shielding when the significant contributions to the column density 'seen' by a ray are spread over more than one cell. A more exact treatment would trace the density along rays from each source to the place it is absorbed. Indeed, ray-tracing schemes can do this, but these become more computationally expensive in proportion to the number of sources, and so may not be practical in simulations with many stars (or self-emitting cells) without further approximation (for example, Baczynski et al. 2015 ignore self-emitting cells, and 'prune' their rays after they exceed a given optical depth, to decrease computational cost).

In this prescription, as the column density increases, the dissociation rate decreases sharply, as desired, but more Lyman-Werner photons are able to penetrate the gas as

the absorption rate decreases proportionally, which is not realistic. Nickerson et al. (2018) propose an alternative scheme, where the photon destruction rate, rather than the Lyman-Werner dissociation rate, is modified. The rate of photon destruction in the Lyman-Werner band is multiplied by a constant factor $S = 400$ (without a corresponding increase in dissociation), which has been determined by calibration against one-dimensional radiative transfer models and observations of HI-H_2 transitions in the Milky Way (Bialy et al., 2017). In this prescription, there is no need for an approximation of the column density from the local volume density, instead it takes advantage of an inherent property of radiative transfer: radiation absorbed by a column depends non-linearly on its optical depth ($\sim$ exponentially at a given constant opacity, according to Equation 3.7), and the enhanced absorption due to line-overlap is approximated to be of the same form, but increased by S times.

3.1.3.2 Test 4: 4.3×10^4 K Black-Body Source in Static H_2 at Constant Temperature

In the first test in a molecular medium, I see to what extent my code agrees with the analytic predictions of Equations 3.32 and 3.41 (concerning the equilibrium radius of the dissociation front, and speed of expansion to it) in static, constant-temperature gas. Following Nickerson et al. (2018), I use a 4.3×10^4 K black-body source (similar to an O-class star) emitting a total of $\dot{N} = 5 \times 10^{48}$ photons per second into the groups in Table 2.1, such that the number of dissociating photons and ionising photons emitted per second are similar ($\dot{N}_{\rm diss} = 1.785 \times 10^{48}$ and $\dot{N}_{\rm ion} = 1.232 \times 10^{48}$ respectively). The density and temperature of the medium are set to $n_H = 10^{-3}{\rm cm}^{-3}, T = 3000\,{\rm K}$ (consistent with the tests by Nickerson et al. 2018; the absolute density is not important, as this test is only sensitive to the ratio of the H_2 formation and dissociation rates to the photon production rate) respectively, so the Strömgren radius is $r_S = 2.78\,{\rm kpc}$.

In pristine (non-metallic) gas, assuming H_2 formation is carried out via the major H^- and H_2^+ channels, with an equilibrium concentration of these intermediate reactants (using rates from my network, Table 2.2), the H_2 formation rate is extremely slow at 3000 K

(and H_2 would be destroyed rapidly by collisions at $T \gtrsim 1 \times 10^4$ K); under these conditions, the dissociation radius would be $r_{H_2} = 46.2$ Gpc, with a formation time of $t_{\mathrm{rec},H_2} = 3.12 \times 10^{29}$ yr. So, to more easily resolve the equilibrium dissociation radius in a practical simulation, I initialise the medium with an artificially-high metallicity of $Z = 100\,Z_\odot$ (100 times the metallicity in the test of Nickerson et al. (2018)), so that the dissociation radius is $r_{H_2} = 6.59$ kpc (with a dust formation coefficient of $\alpha_{H_2} = 3.5 \times 10^{-17} \cdot 10 \cdot (1 - f_{H^+}) \cdot Z$, taking $f_{H^+} = 0$), relatively close to r_H, and $t_{\mathrm{rec},H_2} = 905$ Myr. The test is run with a uniform grid[1] of 128^3 cells, in a box with side length $L = 8$ kpc to capture both equilibrium radii. I do not enable a self-shielding model in this test, as it is not accounted for in the simple analytic model in Section 3.1.3.

Figures 3.9 and 3.10 show the chemical state at an intermediate time $t_{\mathrm{rec,H}} < t < t_{\mathrm{rec},H_2}$, and Figure 3.11 shows the time evolution of the ionisation and dissociation fronts. As expected (and illustrated in Figure 3.8), the ionisation front equilibrates first near r_H, and the dissociation front moves out ahead, leaving a band consisting mostly of neutral hydrogen, but the dissociation front reaches equilibrium slightly short of r_{H_2}. The equilibrium radii reached by the simulated fronts are not in perfect agreement with the calculated values, since these fail to account for significant secondary H and H_2 destruction processes. The early-time evolution of both the ionisation and dissociation front radii is also not as expected from the analytic formulae — Eqs. 3.19 and 3.41 — since the equation for $r_I(t)$ assumes that the medium is pure neutral hydrogen, rather than molecular, so initially the ionisation rate depends on $\dot{N}_{\mathrm{diss}}$ (instead of $\dot{N}_{\mathrm{ion}}$) as the H_2 must be dissociated before the atomic hydrogen it becomes can be ionised; furthermore, the equation for $r_D(t)$ assumes that r_I reaches its equilibrium value immediately, and that $r_{H_2} >> r_H$, neither of which is a valid assumption in this case.

Agreement is not excellent in this test, but this is mostly due to the assumptions that

[1] Though a refined grid could resolve the behaviour of the front over a large length scale without using a prohibitively large number of cells (since gradients are less strong further from the source), it does not mitigate the large number of time steps that would be needed to reach equilibrium if I were to attempt a simulation of the dissociation front in a low-metallicity medium, which would need to be much larger to resolve the increased dissociation front radius at equilibrium. The fine cells require shorter time steps, in proportion to their size, but these must still be updated for the long period of time after which the dissociation front has passed them. AMR is not as beneficial in this problem as it may be in others, since the time scale increases as the cube of the spatial scale.

were made in order to make the analytic calculations tractable. I was unable to run a simulation with $r_{\mathrm{H}_2} >> r_H$, as the computational cost was prohibitive. That the qualitative behaviour shown in the schematic Figure 3.8 is reproduced is reassuring, as is the imprecise agreement in the equilibrium radii, considering that there are collisional destruction reactions not accounted for in the analysis.

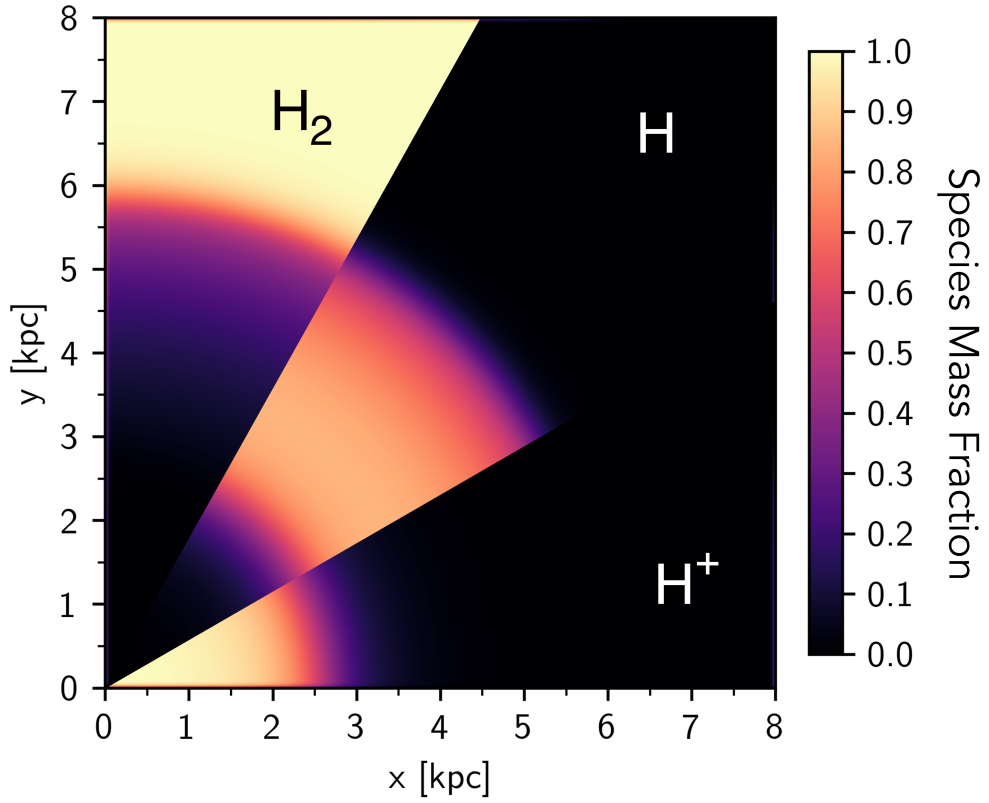


Figure 3.9: *Test 4:* 2D slices of mass fractions of ionised, neutral and molecular gas at 1500 Myr. Here $t_{\mathrm{rec,H}} < t < t_{\mathrm{rec,H}_2}$, so the system resembles the schematic in the bottom part of Figure 3.8, with the ionisation front in equilibrium at r_H, and the dissociation front still moving out to r_{H_2}.

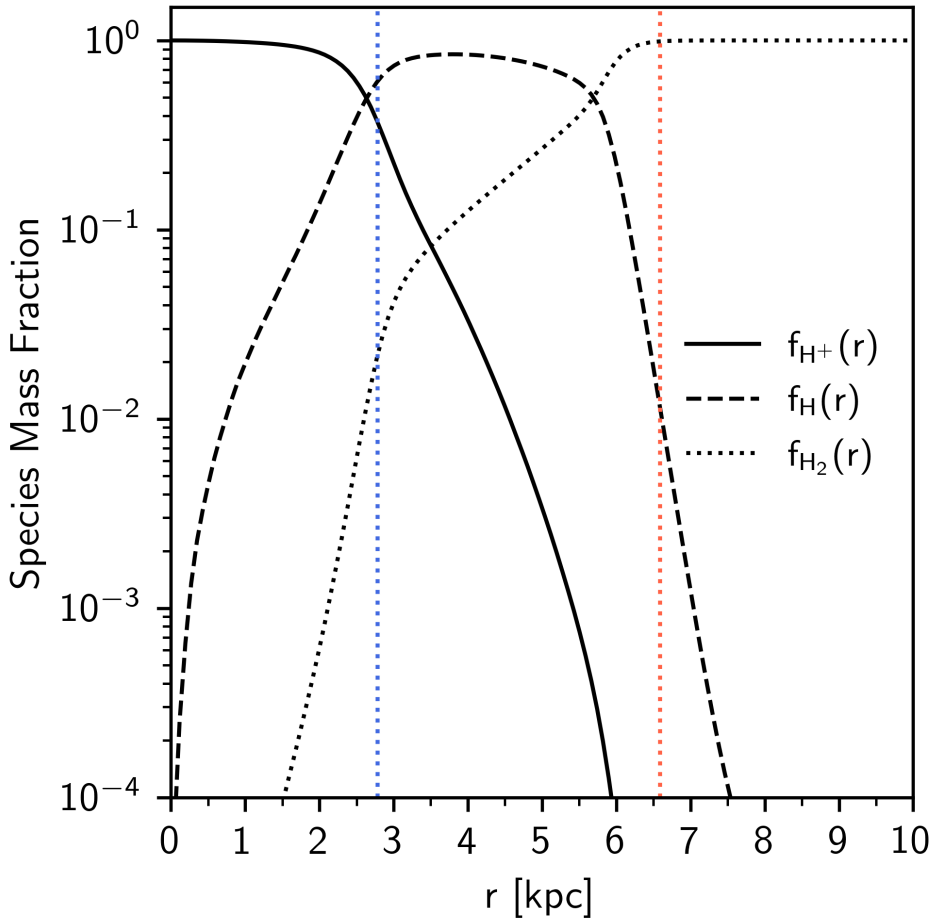


Figure 3.10: *Test 4:* Ionised, neutral and molecular gas profiles at 1500 Myr. Since $t_{\mathrm{rec,H}} < t < t_{\mathrm{rec,H_2}}$, the ionisation front has reached equilibrium at the Strömgren radius (marked by the vertical blue dotted line), but the dissociation front is still moving out toward $r_{\mathrm{H_2}}$ (vertical red dotted line). The profile of $f_{\mathrm{H_2}}(r)$ rises sharply from $r = 0$ as the concentration of neutral hydrogen (from which H_2 can form) increases, but beyond $r = r_H$ the gradient is reduced as $f_H(r)$ is saturated (and thus so is the H_2 formation rate, which is $\propto f_H n_H^2$) and the equilibrium concentration of H_2 is determined almost entirely by the balance of the destruction by Lyman-Werner photons against the formation rate on dust.

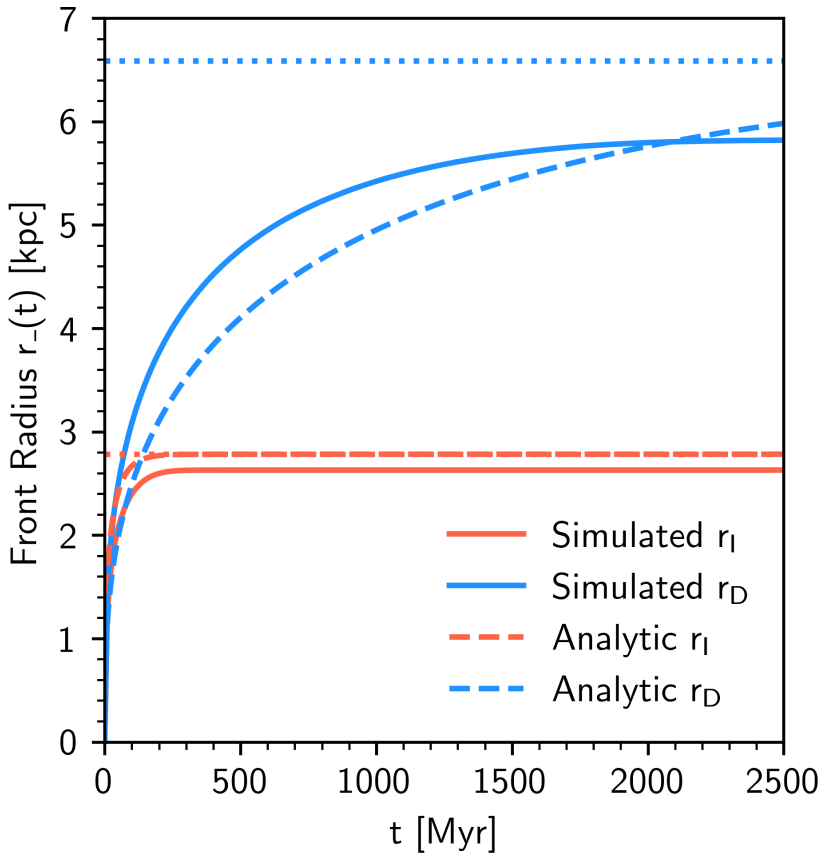


Figure 3.11: *Test 4:* Positions of the ionisation (red) and dissociation (blue) fronts vs. time, and the analytic predictions (dashed). A horizontal red dotted line marks the Strömgren radius, and a horizontal blue dotted line marks r_{H_2}. The simulated ionisation and dissociation fronts do not converge exactly to r_H and r_{H_2}, since, in the system with molecular hydrogen, there are secondary reactions that contribute to the destruction of neutral hydrogen (particularly the formation of H_2 on dust), and the destruction of H_2 (particularly collisions with electrons and neutral hydrogen) that have not been accounted for in the calculation of the equilibrium radii (which balance only the primary formation channels against photodestruction).

3.1.3.3 Test 5: Comparison of Self-Shielding Prescriptions

Now I compare the behaviour of different prescriptions for H_2 self-shielding in a test with hydrodynamics and thermal evolution enabled. I use a box sized $L = 10\,\mathrm{kpc}$, resolved by 128^3 cells, in a medium with density $n_H = 1 \times 10^{-3}\,\mathrm{cm}^{-3}$ chosen such that $x \approx 482.2$ in Equation 3.42, so $f_{\mathrm{shd}} = 0.0089$, which ensured that the effects of self-shielding are strong and clearly visible. Notably, the f_{shd} of Gnedin et al. (2009), with the choice of column density $N_{\mathrm{H}_2} = n_{\mathrm{H}_2}\Delta x/5 \times 10^{14}\,\mathrm{cm}^{-2}$, is resolution dependent, so not only the density, but the box size and resolution have been tuned to achieve a particularly low value. Nonetheless, this test is meant to qualitatively demonstrate the differences between the self-shielding prescriptions of Gnedin et al. (2009) and Nickerson et al. (2018), and does not represent fully realistic conditions in astrophysical gas.

The radiation source has the same spectrum as in Test 4, but the photon emission rate is reduced proportionally to the reduction in density, so that the dissociation front travels at a similar speed. I run 3 different simulations with this setup, one with no self-shielding, one with the self-shielding prescription of Gnedin et al. (2009) (Equation 3.42), and one with the recipe of Nickerson et al. (2018). Figure 3.12 shows that even though the two self-shielding prescriptions lead to similar positions for the dissociation fronts over time (which is a coincidence of the particular test initial conditions), the dissociation fronts have a different structure, with the prescription of Gnedin et al. (2009) allowing more LW radiation to penetrate the medium, causing the neutral fraction (solid red line) to be higher at larger radii than in the other prescriptions, due to the increased dissociation rate beyond the front.

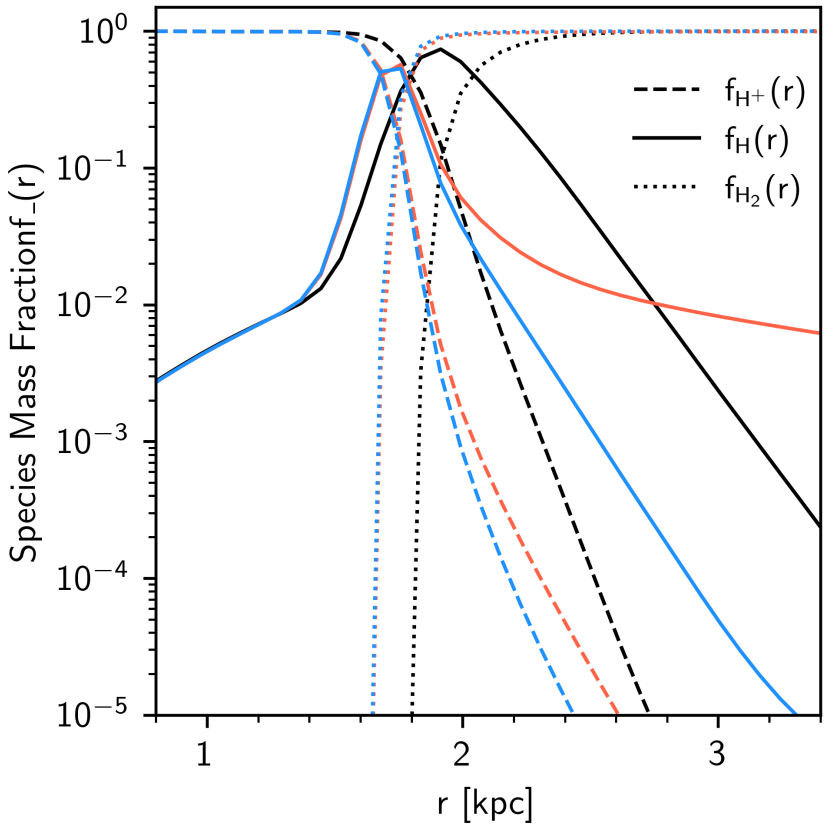


Figure 3.12: *Test 5:* Profiles of ionised, atomic and molecular gas for simulations with three different prescriptions for H_2 self-shielding. **Black:** No self-shielding, **Blue:** Gnedin et al. (2009) with $N_{H_2} = n_{H_2}\Delta x/5 \times 10^{14}\,\mathrm{cm}^{-2}$, **Red:** Nickerson et al. (2018) with $S = 400$.

3.2 Spherical Cloud Collapse

Radiative cooling enables Jeans-stable clouds of gas in the early universe to collapse further and eventually form stars. My implementation of primordial chemistry aims to accurately predict the cooling of such clouds due to H_2, so here I test the effect of the non-equilibrium chemical model on the state of a collapsing cloud.

Omukai (2000) have used a one-zone (or one-'cell') model where a chemical network (on which mine is partially based) is evolved as the cell collapses under its own gravity (with a fixed density evolution $\rho(t)$ that is unaffected by the chemistry), to investigate the chemical evolution of protostellar clouds. I run a similar one-zone simulation with my network, and compare to both the results of Omukai (2000), and the evolution of the centre of a spherical cloud in a 3D self-gravitating hydrodynamics simulation run with

my modified version of RAMSES-RT.

3.2.1 One-zone Model

Following Omukai (2000), I consider a cloud formed of a single 'cell', whose density evolves according to

$$\frac{d\rho}{dt} = \frac{\rho}{t_{\rm ff}}, \tag{3.43}$$

where

$$t_{\rm ff} \equiv \sqrt{\frac{3\pi}{32G\rho}} \tag{3.44}$$

is the *free-fall time*: the time taken for a cloud to collapse to a singularity from rest, neglecting pressure. The specific energy density

$$e = \frac{k_{\rm B}T}{(\gamma - 1)\mu m_p} \tag{3.45}$$

obeys

$$\frac{de}{dt} = -P\frac{d}{dt}\left(\frac{1}{\rho}\right) - \mathcal{L}, \tag{3.46}$$

where the first term represents heating due to compression,

$$P = \frac{\rho k_{\rm B}T}{\mu m_p} \tag{3.47}$$

is the pressure of an ideal gas, and $\mathcal{L}$ is the radiative cooling rate (due to the processes discussed in Section 2.2.2). The density evolution (Equation 3.43) is solved separately to the thermochemistry by a forward Euler scheme

$$\rho^{n+1} = \rho^n + \Delta t \cdot \frac{\rho^n}{t_{\rm ff}}, \tag{3.48}$$

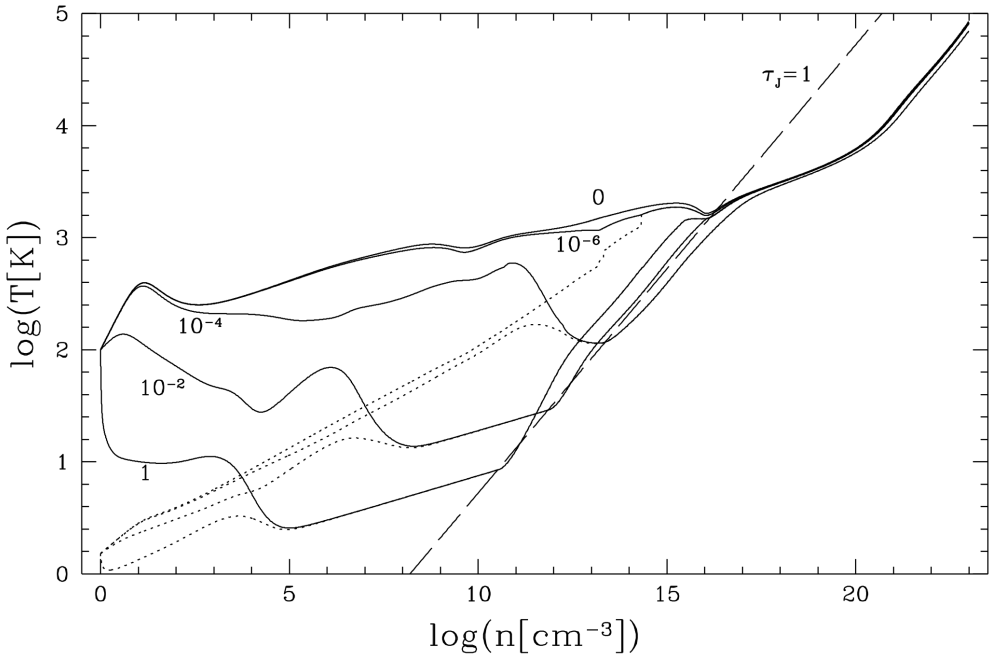
but with a small time step $\Delta t \propto \frac{1}{\rho}$ that decreases as the density increases, so that as the compression accelerates, the thermochemical evolution is still resolved. After each

density update, I solve the thermochemical system with DLSODES (Hindmarsh, 1983), the ODE solver bundled with KROME, with the same global time step.

To compare with the results of Omukai (2000), I initialise the one-zone model with $T = 100\,\mathrm{K}$, $n_\mathrm{H} = 1\,\mathrm{cm}^{-3}$, $n_{\mathrm{H}^+} = 10^{-4}\mathrm{cm}^{-3}$, $n_{\mathrm{H}_2} = 10^{-6}\mathrm{cm}^{-3}$, with the concentration of all other species initially zero, and set a constant metallicity Z. Figure 3.13 shows how the temperature varies as a function of the density in the one-zone model with both the network of Omukai (2000) and my model. I do not run the tests with my model up to densities beyond $n = 1 \times 10^{10}\,\mathrm{cm}^{-3}$ as this is far beyond the densities expected to occur in cosmological simulations with $\sim$0.7 pc physical (as opposed to comoving) resolution, as performed in Chapter 4, and my network has not been designed to include all of the reactions that become relevant at higher densities. The evolution is qualitatively similar at $Z = 0$, though my network reaches a higher peak temperature on the initial rise. This peak temperature is controlled by the H_2 formation rate (as H_2 is the major coolant in this regime); my network, with more up-to-date reaction rates from the UMIST database (McElroy et al., 2013), has formed less H_2 at this density, and so the cloud reaches a higher temperature/density before beginning to cool. Exact agreement with Omukai (2000) is not expected, especially above $n \sim 10^6\ \mathrm{cm}^{-3}$, as I, unlike Omukai (2000), do not include reactions relevant to densities higher than this, since I do not expect the cosmological simulations performed in this thesis to exceed it. Similar features should be present below this density, as I use a subset of the reactions in Omukai (2000), but with different rate coefficients, so agreement should not be quantitative.

As the metallicity increases, my model exhibits much stronger cooling than that of Omukai (2000), and the clouds with $Z/Z_\odot = 10^{-4}, 10^{-3}$ undergo runaway cooling until they reach the temperature floor at 10 K. This is unsurprising given the differences in our treatments; Omukai (2000) explicitly track and evolve the concentrations of carbon, oxygen and hydrogen atoms (and ions), and simple molecules (and molecular ions) formed of these (such as $\mathrm{H}_2\mathrm{O}$ and CO), and use detailed cooling rates that depend on the concentrations of these molecules and metal ions. However, I only track the non-equilibrium chemistry involved in the formation of H_2 in pristine gas, and use a tabulated

metal cooling rate (including the cooling due to metals other than C and O, based on solar-average relative metal abundance ratios) produced by CLOUDY (Ferland et al., 2013, 2017) — a highly-detailed 1D radiative transfer code.



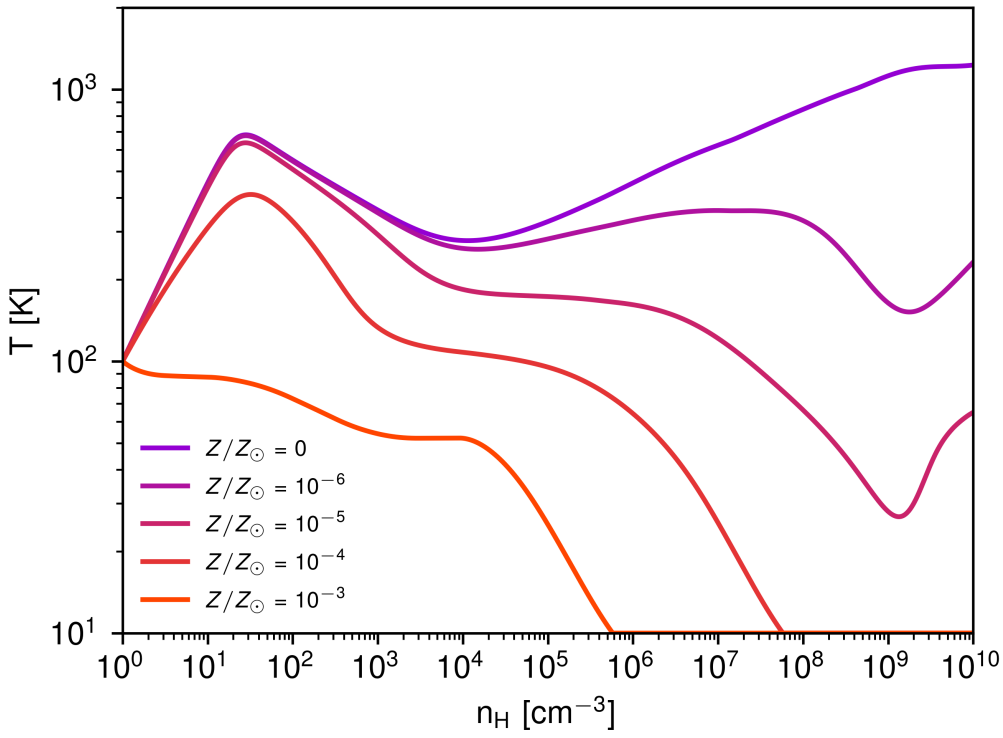


Figure 3.13: Top: Temperature vs. number density (of hydrogen nuclei) for the one-zone model of Omukai (2000) at various metallicities (in solar units). The straight dashed line marks the threshold beyond which the cloud becomes opaque to its thermal continuum radiation emission. Dotted lines show the dust grain temperatures at each metallicity. **Bottom:** Temperature vs number density at various metallicities for the one-zone model with my network and cooling suite.

3.2.2 Comparison with 3D Simulations

I try to reproduce the results of the one-zone model with a 3D hydrodynamics simulation of a collapsing cloud using the same chemical network and radiative cooling functions, at various metallicities. I begin with a spherical homogeneous cloud of the same density, temperature and chemical composition as above, in pressure equilibrium with an atomic medium around it of density $n = 10^{-2}\mathrm{cm}^{-2}$ and temperature $T = 10\,000\,\mathrm{K}$. The radius of the cloud is chosen to be 0.11 kpc, just above the Jeans length $\lambda_J \approx 0.1\,\mathrm{kpc}$, so that the cloud is only just Jeans-unstable — so its collapse is driven by cooling, not purely by gravity. The density at the core of the cloud should then supposedly evolve similarly to that predicted by Equation 3.43 (though not necessarily identically, since the one-zone free-fall model is approximate).

The cloud is placed in the centre of a 5 kpc box, on a 3D grid 64 cells wide at the coarsest level ($l = 6$), which is statically-refined a further 7 times incrementally in spheres of decreasing size around the centre, reaching a maximum resolution of 0.610 pc at the highest level ($l = 13$). Periodic boundary conditions are used on the faces of the box — which is chosen to be much larger than the cloud itself — minimising the gravitational effect of the medium outside the cloud. In testing, the chosen mesh resolution was found to allow the core of the cloud to reach densities of $\sim 10^5\mathrm{cm}^{-3}$, with evolution consistent with simulations with higher spatial resolution, before the resolution prevents further collapse. Thus, the test has numerically converged at this density. I only consider densities up to this value, as higher densities are not expected to occur in the cosmological simulations in the next chapter. Additionally, my modified version of Ramses-rt does not simulate the transport of the infrared radiation emitted by H_2 and metals that is primarily responsible for cooling the cloud, and so cannot capture the correct behaviour when the cloud is no longer optically thin to this at higher densities.

To verify that the chemical network is correctly coupled to the hydrodynamics in my modified version of Ramses-rt (that numerical problems or bugs in the code do not affect results), I test that the 3D simulations of collapsing spherical clouds are consistent with the one-zone model described above. The one-zone model has a density evolution

independent of the chemistry, so the time step can be controlled to verify that the model converges to a solution as the step length decreases. If my implementation of thermochemistry is correct, and the density evolution of the core of the cloud is similar, it should produce similar results to the one-zone model. I also run a metal-free cloud with the version of RAMSES-RT by Kimm et al. (2017) for comparison.

I modify RAMSES-RT (applying the same modification to that of Kimm et al. (2017)) so that it outputs the hydrodynamic and thermochemical state of the central eight cells of the cloud every coarse time step. There are thousands of coarse time steps in each run, so this provides data with much higher time-resolution than what could be extracted from full 3D snapshots (which are too large to output so frequently).

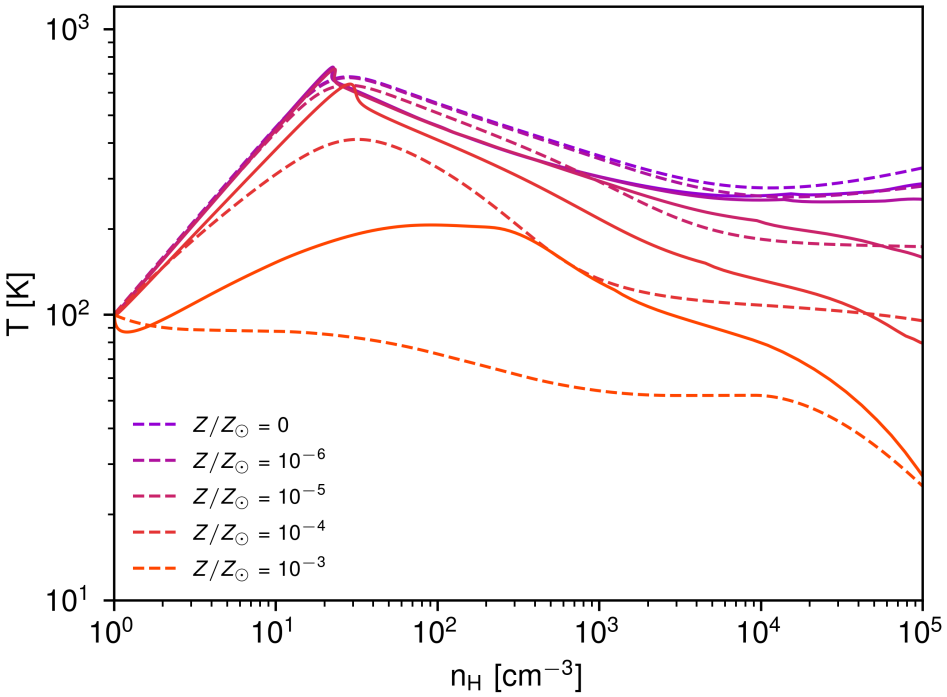


Figure 3.14: Temperature against density of the one-zone model with my network and a free-fall density evolution (dashed), and the core of the 3D cloud (solid), for various metallicities (in solar units).

Figure 3.14 shows the temperature vs. number density (of hydrogen nuclei, which is also ρ/m_H in a pristine cloud) at the core of the cloud as it collapses, for different metallicities, as predicted by the one-zone model and 3D simulations. For the $Z/Z_\odot = 0, 10^{-6}$ cases, the agreement is close but not perfect. These low metallicity one-zone models agree very

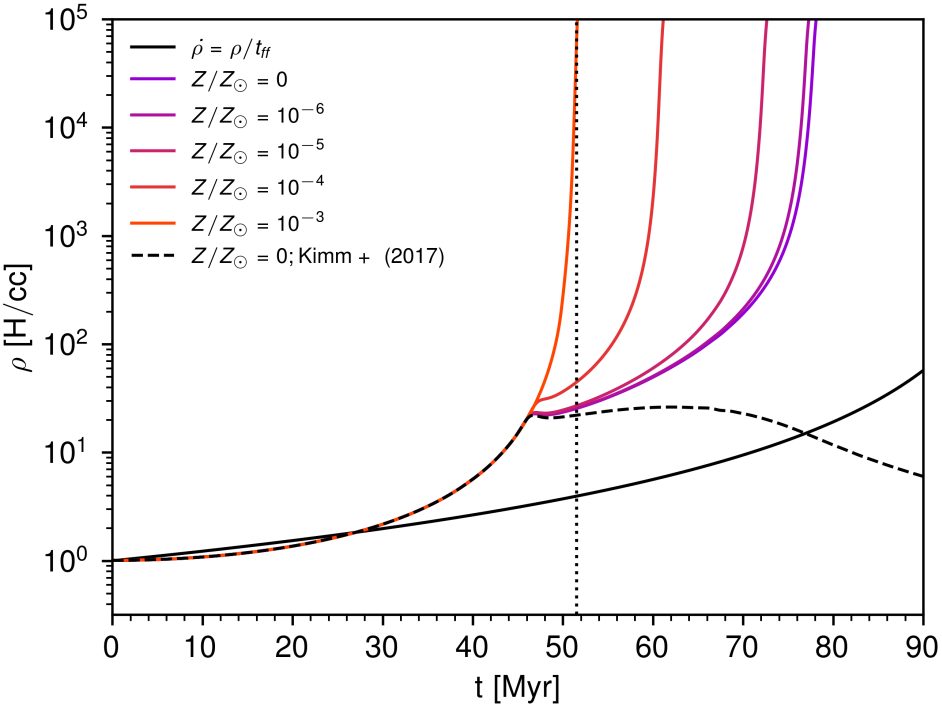


Figure 3.15: Core density evolution for the 3D cloud (coloured solid lines) for various metallicities (in solar units), the one-zone free-fall model (black solid line), and the version of RAMSES-RT of Kimm et al. (2017) (dashed) with $Z = 0$. The free fall time is marked by the vertical dotted line.

well with 3D clouds during the initial collapse, in which radiative cooling is negligible in comparison to the compressional heating, leading to a self-similar temperature evolution, as the compressional heating term in $\frac{dT}{dt}$ is proportional to the temperature, with no density dependence. Omukai (2000) sees this $T(\rho)$ power law in their simulations independently of starting temperature. This power law continues in all low-metallicity models until the density is high enough for molecular cooling to become significant at $n_H \approx 20$ cm^{-3}. As the density increases from here up to $n_H \sim 10^4$, there is a low H_2 fraction (and thus cooling rate) in the low-metallicity clouds (shown in Figure 3.17 for the 3D collapse), but the increased density leads to an increased H_2 formation rate, and thus an increased heating rate due to the formation, as energy is released from the binding of two H atoms. The low-metallicity clouds (with $Z/Z_\odot \leq 10^{-6}$) then begin runaway heating as they pass $n_H \sim 10^4$, as the compressional heating dominates. H_2 formation is also the major effect preventing runaway cooling for clouds with $10^{-4} \leq Z/Z_\odot \leq 10^{-6}$. For

more metallic clouds (with $Z/Z_{\odot} \geq 10^{-5}$) the T vs. n curves for the two models diverge immediately and remain very discrepant as the cloud collapses further.

The primary reason for the poor agreement between the one-zone and 3D models is evident in Figure 3.15. None of the 3D clouds collapse according to the free-fall relation (Equation 3.43) of the one-zone model, although the evolution is somewhat similar for the first 30 Myr. The lower metallicity one-zone models agree better than the more metallic ones up to $n_H \approx 20$ cm^{-3} since they are dominated by compressional heating which is a function of only temperature, not density. At higher metallicities the metal cooling rate is relatively high, and begins to overcome the heating from compression and H_2 formation.

The metal-free cloud 'bounces' back before its free-fall is complete, and then begins to fall back again. This also occurs in the simulation with the version of Ramses-rt by Kimm et al. (2017), which uses a semi-equilibrium model of H_2 formation, and a different scheme for solving the thermochemical ODEs, so it appears that this bounce is not due to a deficiency in my scheme in particular. The cloud run with the chemical model of Kimm et al. (2017) does not cool as much as those run with mine in this test, so the pressure support prevents the collapse entirely on the first bounce. More metallic clouds cool more rapidly, delaying the 'bounce' (as pressure support is reduced), or eliminating it entirely (for metallicities equal to or higher than $Z/Z_{\odot} = 10^{-3}$) up to the maximum density I consider (so the cloud collapses as completely as the simulation resolution allows within one free-fall time $t_{ff} = 51.51$ Myr). Due to the difference in the density evolution, the compressional heating term and density-dependent reaction rates in the one-zone model are not consistent with the 3D simulation, so the thermochemical state should not be expected to be consistent.

3.2.2.1 Forcing the Density Evolution

In order to verify that the thermochemistry implementation is correctly coupled to the hydrodynamics, the density evolution of the central cells of the 3D clouds (which I have output at high time resolution in my modified version of Ramses-rt) can be input into the one-zone model directly, instead of using the free-fall relation (Equation 3.43). In the following tests in this subsection, the state of the core of each 3D cloud is interpolated

in log space to fix the density as a function of time in the one-zone model, and a finite difference is taken to calculate the derivative in the compressional heating term, $-P\frac{d}{dt}\left(\frac{1}{\rho}\right)$ of Equation 3.46.

Figure 3.16 shows the temperature evolution of clouds at different metallicities, as in 3.14 but with the density of the one-zone model fixed to be consistent with the corresponding 3D simulations (as described above). Now the agreement is near-perfect at all metallicities, suggesting that the coupling between thermochemistry and hydrodynamics is correctly implemented, and converges despite the operator splitting in Ramses-rt. That is, in the case of a collapsing spherical cloud, the time step chosen by Ramses-rt is small enough that the coupling between the operator-split steps is well resolved. There is also no evidence in this test of numerical instability caused by the interactions between hydrodynamics, gravity, radiation and chemistry operators.

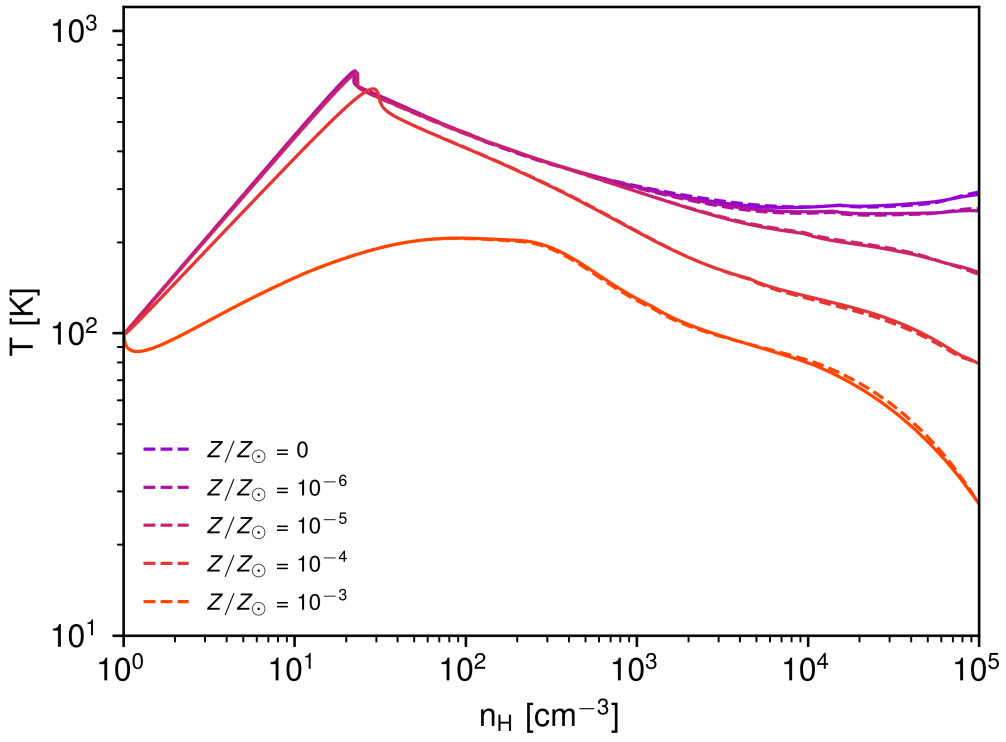


Figure 3.16: Temperature vs. density of the core of the 3D cloud (solid) and the one-zone model with a density evolution consistent with the 3D simulation (dashed), for various metallicities (in solar units).

Figure 3.17 shows the mass fraction of molecular hydrogen as the cloud collapses.

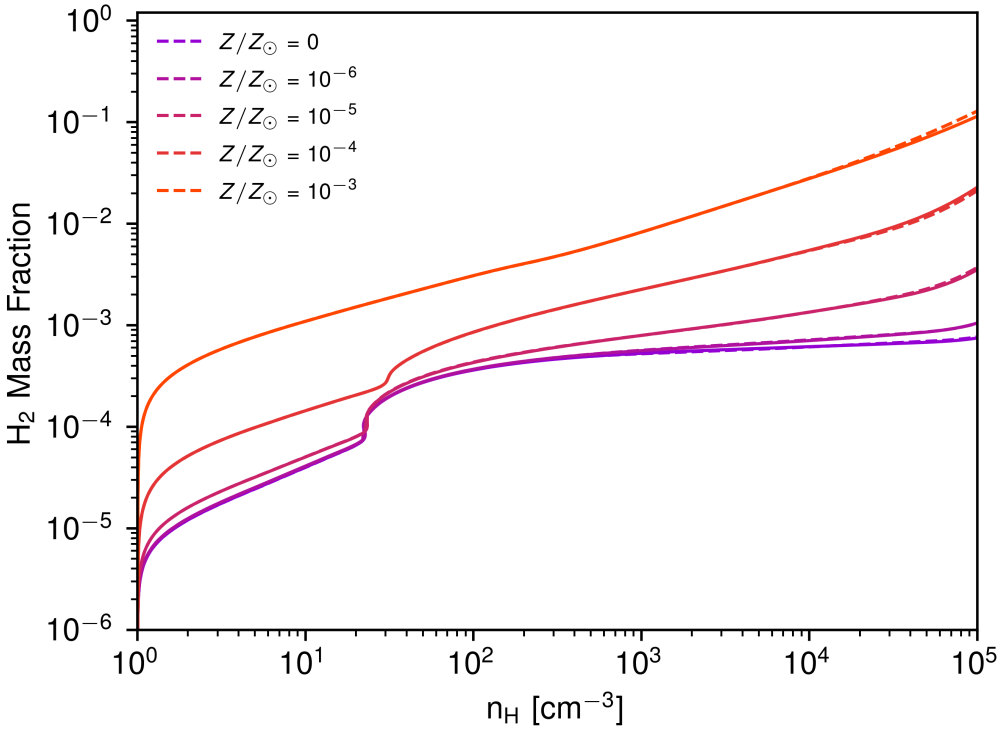


Figure 3.17: Mass fraction of molecular hydrogen vs. density of the core of the 3D cloud (solid) and the one-zone model with a density evolution consistent with the 3D simulation (dashed), for various metallicities (in solar units).

Again, agreement is near-perfect between the 3D and one-zone models. In all but the $Z/Z_\odot = 10^{-3}$ cloud, the 'bounce' after $\sim 1 t_{\rm ff}$ causes a sharp increase in the H_2 concentration, with the formation rate much reduced as the collapse is temporarily halted. More metallic clouds cool more quickly and thus begin forming H_2 sooner, but below 100 K the formation rate does not change much with temperature, so the clouds that begin to runaway cool have similar formation rates at higher densities (similar slopes in Figure 3.17).

Self-consistently modelling the non-equilibrium formation of H_2 in collapsing gas (of various metallicities) is one of the main goals of my modified code. Agreement between the thermochemistry of the simulated 3D cloud and the one-zone model (when forced to have a consistent density evolution) suggests that the thermochemical solver is correctly implemented in my modified version of RAMSES-RT, and that coupling between hydrodynamics, gravity and thermochemistry is well-resolved in collapsing clouds despite

operator splitting.

Chapter 4

Non-equilibrium Primordial Chemistry in Simulations of Dwarf Galaxies in the Early Universe

'Yueh! Yueh! Yueh!' goes the refrain. 'A million [CPU hours] were not enough for Yueh!'

— 'A Child's History of Muad'Dib' by The Princess Irulan

4.1 Introduction

Before any stars form in a galaxy, the abundance of elements in the gaseous medium is as it was after Big Bang Nucleosynthesis: $\sim$ 75% hydrogen, $\sim$ 25% helium (Planck Collaboration et al., 2020), with deuterium making up $(2.55 \pm 0.02) \times 10^{-5}$ of hydrogen nuclei (Riemer-Sørensen & Jenssen, 2017), also containing a trace amount of Lithium. The very first (*Population III*, or Pop. III) stars then coalesce out of this metal-free 'pristine' gas. The supernovae caused by these stars are responsible for the initial enrichment of the galactic medium with metals, allowing the birth of the next generation of stars.

However, there have been no direct detections of Pop. III stars, and their properties are not well constrained. The initial mass function (IMF) is uncertain, with disagreement between theoretical studies (Greif et al., 2011; Hirano et al., 2015; Stacy et al., 2016; Hosokawa et al., 2016; Hirano & Bromm, 2017; Sharda et al., 2019), whose predictions

of initial mass functions span many decades, from $\sim 0.1 M_{\odot}$ to $\sim 1000 M_{\odot}$. Depending on their mass, these stars end their lives in different ways, with those below $40 M_{\odot}$ producing type-II core-collapse supernovae, and those above $140 M_{\odot}$ producing extremely energetic pair-instability supernovae, injecting their entire metal content into the galactic medium, leaving no remnant (Heger et al., 2003). Thus, understanding the formation of Pop. III stars is critical to understanding the evolution of galaxies in the early universe.

One prospect for the detection and study of Pop. III stars is that galaxies containing Pop. III stars may exhibit detectable spectral features (Zackrisson et al., 2011; Grisdale et al., 2021; Woods et al., 2021; Katz et al., 2023). However, these may be degenerately associated with other properties of the host galaxy and stellar population, so to detect or study the stars indirectly via this method, a model of galaxy formation must be used to disentangle the Pop. III signal from other sources. In order to properly predict the properties of the early galaxies that contain Pop. III stars, such a model must self-consistently capture the interaction between these stars and the gas in their hosts, as they can strongly influence their surroundings via pair-instability supernovae (as well as other forms of *feedback*, including type II supernovae, radiation and stellar winds), and should additionally describe the formation and feedback of Pop. II stars, which start to come into being once the first stars have enriched their environment with metals. The onset of Pop. II star formation, seeded by the deaths of Pop. III stars, has been studied in cosmological simulations by Tornatore et al. (2007) (with a detailed metallic enrichment model, but without self-consistent radiation transport or chemistry), Ricotti et al. (2008) (with radiation transport using the scheme of Gnedin & Abel 2001, and a non-equilibrium treatment of H_2 chemistry from Galli & Palla 1998, but run at $\sim 200\times$ coarser resolution than the simulations in this work, and missing mechanical supernova feedback) and Wise et al. (2012) (at much higher resolution than previous work, and with mechanical supernova feedback, but with monochromatic radiation transport, rather than multi-group).

Carrying out the self-consistent modelling of Pop. III stars and their host haloes in a truly predictive way is challenging. In addition to the large uncertainty in the IMF (and other properties) of Pop. III stars, the processes that lead to their formation – even in non-metallic gas — are complex and highly non-linear, subject to the coupled effects of

gravity, radiation, and magnetohydrodynamics. Additionally, since primordial gas clouds are unable to efficiently radiatively cool below $\sim 10^4$ K by H and He atomic line cooling alone, the chemistry involved in the formation of the molecules H_2 and HD (which can emit much more efficiently in the infrared when collisionally excited) is critical to the formation of the first stars.

Cosmological hydrodynamics simulations are required to model the entire system, but self-consistently simulating the chemistry can be prohibitively costly. It is common to instead assume equilibrium concentrations for chemical species in order to calculate cooling rates as a function of temperature, density and metallicity ahead of time (with metal-line cooling rates tabulated by a comprehensive 1D radiative transfer code such as CLOUDY (Ferland et al., 2013, 2017)). However, in collapsing star-forming clouds, or in the CGM where hot outflows interact with cold inflows (Mitchell et al., 2018) and the hot phase coexists with molecular gas (Péroux et al., 2019), this equilibrium is not reached rapidly enough for the assumption to be valid. The position of this equilibrium also depends on the radiation field in general, especially in optically thin gas which may be driven far out of collisional equilibrium by an intense field. Thus, a non-equilibrium treatment is necessary to properly predict radiative cooling rates in simulations, in which the system of equations associated with the chemical network is solved alongside the radiation-hydrodynamics on the same timescale.

Previous work has demonstrated the possibility of solving complex non-equilibrium chemical systems on-the-fly in simulations of galaxies (in addition to those already mentioned). Richings & Schaye (2016) run simulations of idealised galaxies (with exponential disc, rather than cosmological, initial conditions) at metallicities of $Z > 0.01 Z_\odot$, observing significantly enhanced H_2 mass in cool outflowing gas, and a factor ~ 2 different inferred conversion factor between H_2 column density and CO line intensity with non-equilibrium chemical evolution compared to simulations that assume equilibrium. Kimm et al. (2017) use a semi-equilibrium model of H_2 formation (based on the implementation of Baczynski et al. (2015), which assumes an equilibrium concentration for H^- in the H_2 formation rate), coupled to self-consistent radiation transport in cosmological simulations of primordial dwarf galaxies to study their contribution to reionisation. Pallottini et al. (2019) use a full

non-equilibrium model of H_2 formation in a cosmological radiation-hydrodynamics simulation to study a $\sim 10^{11} M_\odot$ halo, running until redshift $z \sim 8$ and reaching a maximum resolution of 30 pc. Lenoble et al. (2024) run high resolution (~ 0.14 pc) hydrodynamics simulations (without radiation) that self-consistently solve the chemical network involved in the formation of primordial molecules H_2 and HD, coupled to a model of Pop. III star formation, up to high redshifts (early times) of $z = 18$ to investigate the properties of Pop. III star-forming haloes of a similar mass to those studied here.

Despite the frequent use of non-equilibrium chemical treatments, as yet, the effect and importance of radiation-coupled non-equilibrium H_2 formation chemistry in simulations of galaxies forming Pop. III stars is not fully understood. In this work I compare the structure and evolution of two simulated dwarf galaxies under five different physics prescriptions for chemistry, metal enrichment, radiation and star formation, each treating non-equilibrium physics to a different degree, in order to investigate its importance in the early stages of galaxy formation. In Section 4.2 I describe the physics prescriptions and summarise the numerical methods used to produce the simulations. Section 4.3 presents the simulations, and compares the outcomes of using each physics prescription. Finally, I discuss my findings in Section 4.4.

4.2 Methods

4.2.1 Cosmological Simulations

I run a suite of cosmological simulations using the same initial conditions as those used in haloes H8 (which I shall call Halo 1 here) and H2 (Halo 2) of (Kimm et al., 2017) — chosen as these are respectively the least and most massive of the set simulated by Kimm et al. (2017) — forming dwarf galaxies with mass $\sim 10^8 M_\odot$ at redshifts $z = 8.1, 10.6$ respectively. H8 is thus expected to form stars relatively later than H2. These initial conditions were generated using MUSIC (Hahn & Abel, 2013), with cosmological parameters from the 9$^{\text{th}}$ year results of the WMAP mission (Hinshaw et al., 2013): $\Omega_m = 0.288$, $\Omega_\Lambda = 0.712$, $\Omega_b = 0.045$, $H_0 = 69.33\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $n_s = 0.971$ and $\sigma_8 = 0.830$.

Group	Photon energy range / eV	Left edge significance	Present in Prescriptions
1	[0.75 – 2.65)	H^- detachment	I
2	[2.65 – 11.2)	H_2^+ dissociation	I
3	[11.2 – 13.59)	H_2 dissociation	I, II
4	[13.59 – 15.42)	H ionisation	I, II
5	[15.42 – 24.58)	H_2 ionisation	I, II
(4-5)	[13.59 – 24.58)	H ionisation	III only
6	[24.58 – 54.41)	He ionisation	I, II, III
7	[54.41 – ∞)	He^+ ionisation	I, II, III

Table 4.1: The fixed photon groups used in the simulations in this chapter, labelled by the prescriptions they are present in. The left edges are chosen to align with minimum ionisation/dissociation energies. Group 3 is the *Lyman-Werner Band*, containing H_2 dissociating photons. Groups 4-5 account for H ionising radiation, with the right edge at the first ionisation energy of He, and are split by an edge at the H_2 ionisation energy in prescriptions with a non-equilibrium H_2 treatment.

All simulations are run with 128^3 cells at the coarsest level, and are refined up to a maximum physical (not comoving) resolution of 0.7 pc, following Kimm et al. 2017, and using the same strategy of a combination of quasi-Lagrangian and Jeans refinement criteria (see Section 2.1.3), in which a cell is refined if either its contained mass is $> 8\times$ the dark matter particle mass ($M_{\rm DM} = 90 M_\odot$ here), or if the Jeans length is resolved by fewer than 32 cells. In order to maintain the desired physical resolution, a new refinement level is added for every doubling of the scale factor a after the simulation begins. Pop. II star particles have mass $M_{\rm II}^\star = 91 M_\odot$, each representing a population formed with the IMF of Kroupa (2001), using the star formation criteria of Rosdahl et al. (2018) (based on that of Federrath & Klessen 2012), described in Section 2.1.7.1.

For each halo, I run five different simulations, each with a prescription for incrementally more non-equilibrium physics (listed in Table 4.2), using modified versions of the radiation-hydrodynamics code Ramses-rt (Rosdahl et al., 2013). Simulations that use radiation transport are performed with a reduced speed of light of $\bar{c} = 0.003$, and groups listed in Table 4.1, with average cross sections determined on the fly based on stellar spectral evolution models from Bruzual & Charlot (2003) for Pop. II stars, and Schaerer (2002) for Pop. III stars.

Prescription I is implemented by my custom version of Ramses-rt, which solves a large non-equilibrium chemical network (described in Section 2.2.8) of the most important

chemical and photo reactions involved in the formation and destruction of the primordial molecules H_2 and HD. The thermochemical scheme in RAMSES-RT is mostly replaced here by KROME, which generates code to solve the ODE system associated with the network with the implicit, adaptive solver DLSODES, which is efficient for sparse, stiff systems. The chemistry is self-consistently coupled to the radiative transfer, by including photon group densities, and the absorption/emission reactions that evolve them in the chemical network. The self-shielding scheme of Nickerson et al. (2018) is adopted here, increasing the destruction of Lyman-Werner (LW) Band photons to approximate the effect of line-broadening in H_2 as column density increases, allowing cells to shield their neighbours from dissociating radiation.

Prescription II is that of Kimm et al. (2017) and Katz et al. (2017), featuring a semi-equilibrium treatment of H_2 formation (described in Section 2.2.7) that assumes a (collisional) equilibrium concentration for the primary primordial reactant H^-, which is also coupled to the radiation transport (though the assumed H^- equilibrium does not account for the radiation field). This prescription uses the self-shielding prescription of Gnedin et al. (2009), in which the absorption of LW photons is decreased to achieve an approximately realistic local reduction in H_2 dissociation.

Prescriptions I and II model the formation of individual Pop. III stars (as described in Section 2.1.7.1) and supernova feedback (via both type II and pair-instability supernovae) using the scheme of Wise et al. (2012), as used in (Kimm et al., 2017).

Prescription	Chemistry	Radiation	Pop. III	Z Floor
I (Mine)	Full non-eq. H_2 & HD	M1 RT (7 groups)	•	×
II (Kimm et al., 2017)	Semi eq. H_2	M1 RT (5 groups)	•	×
III (Rosdahl et al., 2018)	H & He Ion.	M1 RT (3 groups)	×	•
IV (Kimm et al., 2017, no RT)	Semi eq. H_2	UV Background	×	×
V (Dubois et al., 2021)	H & He Ion.	UV Background	×	•

Table 4.2: My physics prescriptions. I has the most complete non-equilibrium models (using my newly-modified code), and then each subsequent set has incrementally less non-equilibrium physics. Prescriptions II-V correspond to physics models used in other published work (or slight modifications thereof).

Prescription III is identical to that used in SPHINX (Rosdahl et al., 2018), a cosmological simulation with comoving side length 10 Mpc/h, tracking only the non-equilibrium evolution of the ionisation states of hydrogen and helium, coupled to 3 photon groups in the radiation transport. To compensate for the lack of a molecular cooling model, the simulations are initialised with a metallicity floor of $Z = 3.2 \times 10^{-4} Z_{\odot}$ to artificially enhance the cooling rate (calibrated to cause the formation of the first stars at $z \sim 15$).

Prescription V mimics NewHorizon (Dubois et al., 2021), a zoom-in cosmological simulation (into a spherical region of the larger Horizon-AGN simulation; Dubois et al. 2014) with comoving radius $\sim$ 10Mpc/h. These simulations are initialised with the same metallicity floor as Prescription III, and track the non-equilibrium evolution of hydrogen and helium, but use a homogeneous UV background instead of a radiation transport model.

Prescription IV is identical to II except that the radiation transport is disabled, replaced by the homogeneous UV background model of Faucher-Giguère et al. (2009), with the shielding treatment of Rosdahl & Blaizot (2012) in which the incident radiation is attenuated in proportion to the density (Section 2.2.4). Prescription IV also does not model the formation of Pop. III stars, and thus may be similarly thought of as V with non eq. H_2 formation instead of a metallicity floor.

The simulations run with Prescription I are $\sim 10\times$ more expensive than those run with Prescription II, so I was only able to run Halo 1 down to $z \approx 9.88$ and Halo 2 down to $z \approx 9.76$ with my fully non-equilibrium code (due to the much larger cost of solving the chemical system than in the other prescriptions). In Halo 1 all other prescriptions were run down to at least the same redshift as I, but in Halo 2 the simulation with Prescription IV eventually became more expensive (only run down to $z \approx 11.49$) due to forming a large number of stars, which shock and heat the medium frequently via supernova feedback, putting a strict constraint on the numerical time step.

Unfortunately, a numerical or software bug with the modified version of Ramses-RT[1] used to run Prescription III (SPHINX) caused a catastrophic failure of the simulation shortly after the first star forms in either set of ICs, in which a bubble around the star

[1]This was kindly provided by Harley Katz, who was most likely not responsible for the bug, but nonetheless tried to assist in its diagnosis.

expands at the *grid speed* $\frac{\Delta x}{\Delta t}$ (which is always superluminal, compared to the reduced speed of light used), strongly heating and ionising the gas inside it. It was not possible to resolve the bug, so I omit most of the results from these runs as it is impossible to interpret them beyond the formation of the first star.

4.2.2 Halo Finding

The implementation by Tweed et al. (2009) of the AdaptaHOP algorthim (Aubert et al. 2004; summarised in Section 2.3), HaloMaker, is used to find haloes in the simulation snapshots, and then TreeMaker (Tweed et al., 2009) is used to construct merger trees, associating progenitor haloes with their combined products. These haloes and merger trees are found in only the dark matter of the snapshots, but the gas (and thus stars) are not necessarily distributed the same way as the dark matter (though the correlation is strong). The gas associated with a halo is defined to be that contained in the virial radius $r_{\rm vir}$, which is found by integrating the mass due to dark matter, gas and stars in a sphere out to a radius such that the average density in the sphere of radius $r_{\rm vir}$ is

$$\bar{\rho}(r_{\rm vir}) = 178\rho_0(1+z)^3, \tag{4.1}$$

where ρ_0 is the average density of the universe today. The centre of the sphere from which this integration is performed is taken to be either the "densest" particle found by HaloMaker (approximately that with the most neighbours in its vicinity), or, if stars are present in the halo, it is taken to be the centre of mass of those stars. Stars are considered to be present in a halo if they are contained within the convex hull of all dark matter particles associated with the halo by HaloMaker. The *main branch* of the merger tree is defined to be the one containing the most massive progenitor halo at each level in the tree.

Due to a significant data loss, it was not possible to robustly identify merger trees in all simulations — such that each one had a similar tree, with the same subhaloes identified in each snapshot, so that they can be compared on equal terms — by using TreeMaker

alone, though the backed-up data was at sufficient time resolution for the rest of the analysis. Since the dark matter evolution, and thus the subhalo locations are mostly identical between prescriptions, I was able to resolve this by rerunning one of the simulations with Prescription II (which I shall call the 'reference' simulation) for each set of initial conditions, and using the branches of the merger tree as a 'guess' for progenitor identification in other simulations. For each of the other simulations, progenitor branches are found by interpolating the comoving position of the halo in the corresponding branch of the reference simulation, and using the interpolated 'guess' as the starting point for a 'Shrinking Spheres' (Power et al., 2003) iteration to find the updated centre of the subhalo, in which the centre of mass of a sphere of radius $r_{\rm vir}$ (also interpolated from the reference branch) centred on the initial guess, then for a sphere of radius $r_{\rm vir} f_{\rm shrink}$ (where I take the 'shrink factor' to be $f_{\rm shrink} = 0.975$) around the new centre of mass, and so on recursively for successive spheres until the mass contained is 1% of the initial mass. This procedure is vital for my analysis, replacing the unusable merger history from the damaged data with one that is as clean as the reference simulation in all of my other runs.

4.3 Results

The analysis in this chapter is centred on the impact of a full non-equilibrium treatment of radiation-coupled H_2 formation and destruction, compared to alternative chemical and radiation prescriptions from other published work. Since the most sophisticated of the other prescriptions under test is II (that of Kimm et al. 2017) the differences between Prescriptions I and II are emphasised, though comparisons to the other prescriptions serve to demonstrate the importance of non-equilibrium treatments of radiation and chemistry in general.

The additional non-equilibrium H_2 physics present in Prescriptions I, II and IV is most important in gas with little to no metal enrichment, which can go on to form Pop. III stars. I therefore focus on the properties of these stars in particular, and the gas in their host haloes at early times, before significant metal enrichment has occurred.

In the rest of this section I summarise the important global properties of the two simulated haloes (and their evolution), comparing these under each physics prescription. I then analyse the effect of differing physics on both Pop. III and Pop. II star formation, and study the significance of my results in the context of the inherent stochasticity of the star formation models used (and unavoidable, non-deterministic rounding errors). Finally, I examine the deviation from photochemical equilibrium in gas in haloes hosting Pop. III stars, close to the time of their formation, to demonstrate the importance of using a fully non-equilibrium treatment.

4.3.1 Properties of the Simulated Haloes

Since I am most interested in star-forming regions, I consider the properties of the star-forming *subhaloes* (found using HaloMaker, as described above) of the simulated objects. The final redshifts of the simulations are presented in Table 4.3, along with the integrated masses of dark matter, gas and stars of the top 8 most massive subhaloes at the minimum redshift that all runs have reached for each of the two sets of initial conditions. At redshift 11.48, the subhaloes of Halo 2 have accrued slightly less dark matter, but slightly more gas than those in Halo 1 at redshift 9.87. In Halo 1, Prescription IV has formed the most stars (and produced the greatest mass of metals), followed by I, V, then II. Halo 2 is similar, except that now V has formed more stars than I; this may be explained by the fact the UV background of Faucher-Giguère et al. (2009) — present only in IV and V — is more intense at lower redshift, so Halo 1 experiences stronger photoionisation and heating at its end time than Halo 2, suppressing star formation by destroying cool, star-forming gas.

It can be seen in Figure 4.1 that the UV background has a stark effect on the structure of the haloes in Prescription V (shown in the bottom row). Here, In Halo 1, all filamentary structure that was not sufficiently dense to shield itself has been destroyed, as have the smaller structures of Halo 2. This has not occurred visibly in Prescription IV (3rd row), as these simulations form H_2 very efficiently everywhere, even in the rarefied medium outside subhaloes (note the high projected H_2 fraction in the voids in the lower panels of Figure 4.4 vs. the other prescriptions), with no radiation from stars to destroy it, maintaining a much cooler IGM than Prescription V (evident in the third row of Figure 4.2)

ICs	Pr.	$z_{\rm end}$	$z_{\rm PopIII}$	$\log M_{\rm halo}$	$\log M_{\rm gas}$	$\log M^\star$	$\log M^{\star\ \rm init}_{\rm PopIII}$	$\log M^{\star\ \rm init}_{\rm PopII}$	$\log M_Z^{\rm zoom}$
H. 1	I	**9.87**	16.11	8.17	6.88	5.73	3.41	5.83	3.75
	II	8.53	17.34	8.17	6.94	4.70	3.84	4.80	3.08
	III*	9.87	-	8.17	6.86	-	-	-	-
	IV	8.96	-	8.17	7.01	5.87	-	5.96	3.98
	V	8.99	-	8.17	6.70	5.42	-	5.47	3.52
H. 2	I	9.76	19.94	8.05	6.99	5.36	3.39	5.43	3.49
	II	9.50	20.51	8.05	7.08	5.29	3.69	5.34	3.46
	III*	14.66	-	-	-	-	-	-	-
	IV	**11.48**	-	8.05	7.01	5.81	-	5.96	4.10
	V	9.29	-	8.05	6.81	5.62	-	5.73	4.03

Table 4.3: Integrated properties of the top 8 most massive subhaloes in every simulation (with the first column showing the initial conditions — Halo 1 or 2, and the second showing the prescription used in the run). The properties are given at the maximum redshift ($z_{\rm end}$) reached by all runs with prescriptions I, II, IV, and V (highlighted in **bold**) for a given set of ICs. I do not show values for Prescription III in some columns as it crashes immediately after forming stars (except in Halo 1 where it reaches $z = 9.87$ without forming any stars). $z_{\rm PopIII}$ is the redshift of formation of the first Pop. III star (for the runs that have them); $M_{\rm halo}$, $M_{\rm gas}$, $M^\star$ are the dark matter, gas, and stellar masses of the subhaloes; $\log M^{\star\ \rm init}_{\rm PopIII}$ and $\log M^{\star\ \rm init}_{\rm PopII}$ are the sums of the initial masses of Pop. III and Pop. II stars, i.e. the mass of each population of stars that has *ever* formed; $\log M_Z^{\rm zoom}$ is the mass of metals in the entire zoom region (not the top 8 subhaloes, as for the other properties).

due to cooling from H_2. Accordingly, the IGM in Prescription V is highly ionised, whereas it is almost entirely neutral with IV in either Halo (Figure 4.3), except for the supernova 'bubbles'. The amount of H_2 formed in Prescription IV, and the strong cooling as a result, is completely unrealistic, leading to a continuously high rate of star formation. This is due to the fact that the dissociation due to the background radiation field is not accounted for — it only couples to the gas via a heating term, which is completely counteracted by the strong cooling from H_2 that should have been destroyed.

Figure 4.1 also demonstrates the differences in where stars (clusters of which are indicated by white star icons in the figure) form between the Prescriptions. Each has a varying degree of clustering, with V forming all of its stars (in either Halo) in only one subhalo, and IV forming them across many. Stars in Prescription I are comparatively more clustered than those of II; this is studied in more detail in the next section (4.3.2).

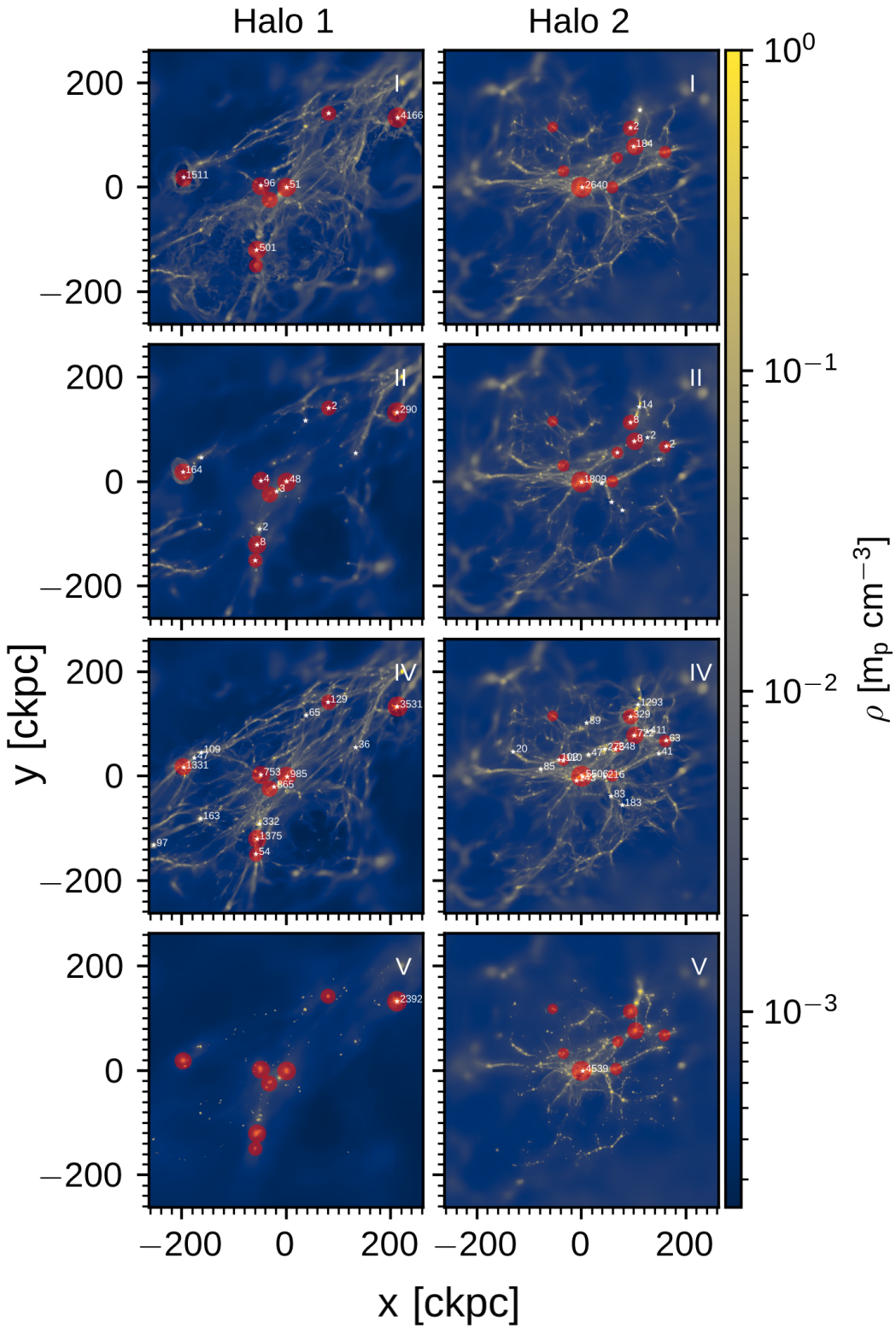


Figure 4.1: Maps of the average gas density along the line of sight (weighted by density) for each of the physics prescriptions (excluding III), for each set of ICs (Halo 1 and Halo 2). The 8 most massive subhaloes are marked by red circles, with the size of the circle (arbitrarily) proportional to the virial radius. Clusters of stars are marked by white symbols at their centre of mass, with adjacent numbers indicating the count of star particles (where solitary particles are not marked with a count).

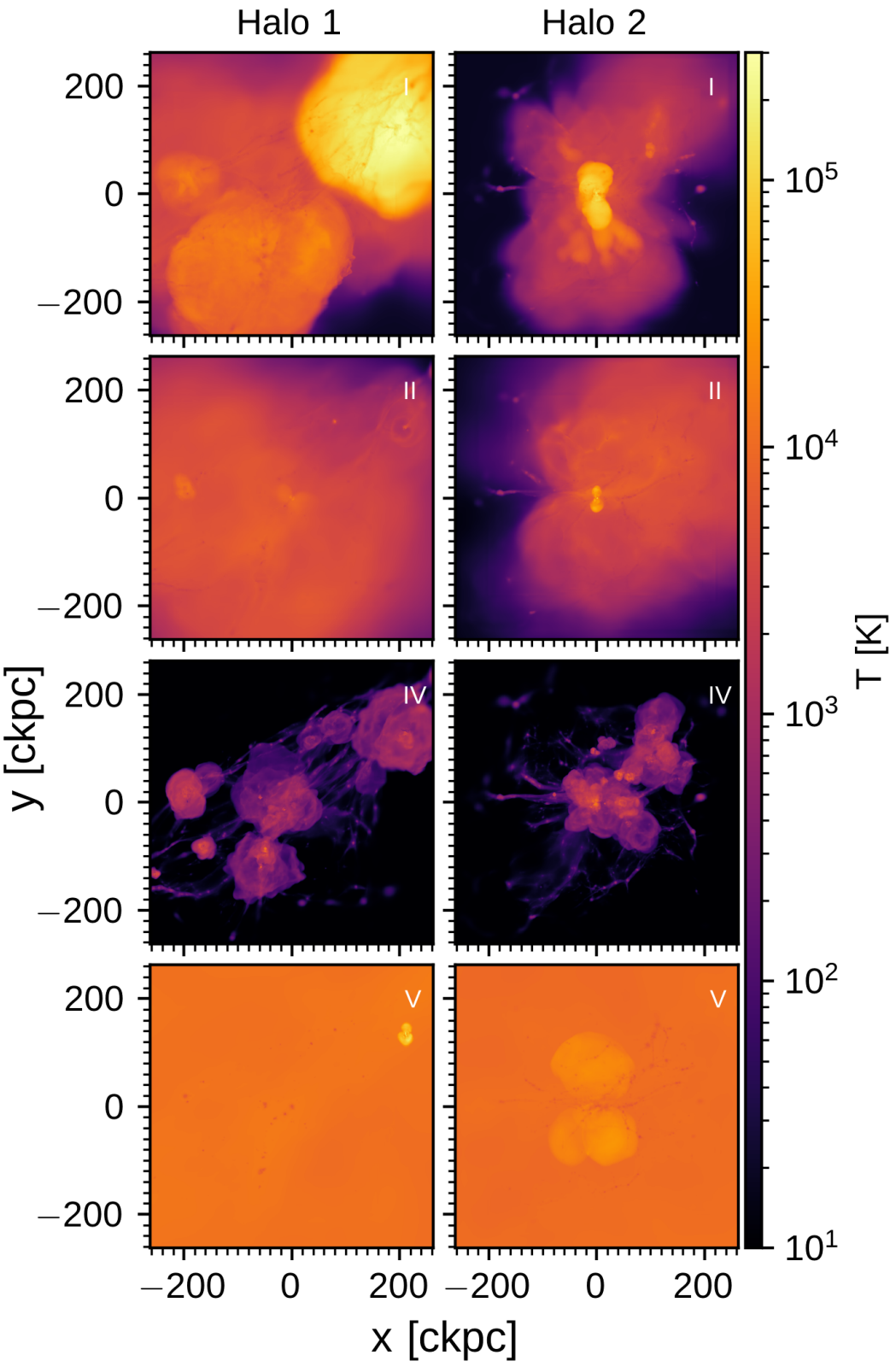


Figure 4.2: Maps of the average gas temperature along the line of sight (weighted by density) for each of the physics prescriptions (excluding III), for each set of ICs (Halo 1 and Halo 2).

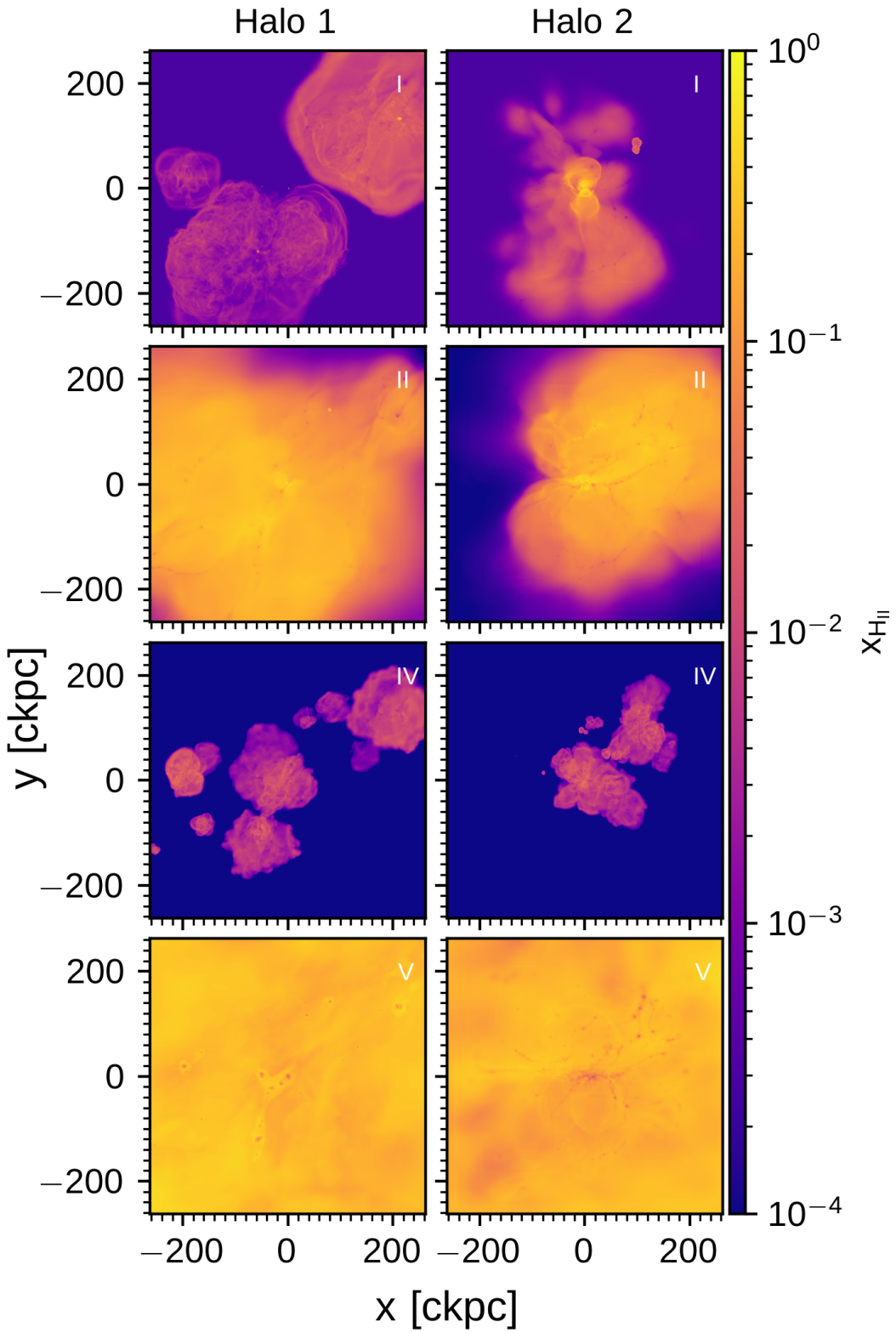


Figure 4.3: Maps of the average hydrogen ionisation fraction along the line of sight (weighted by density) for each of the physics prescriptions (excluding III), for each set of ICs (Halo 1 and Halo 2).

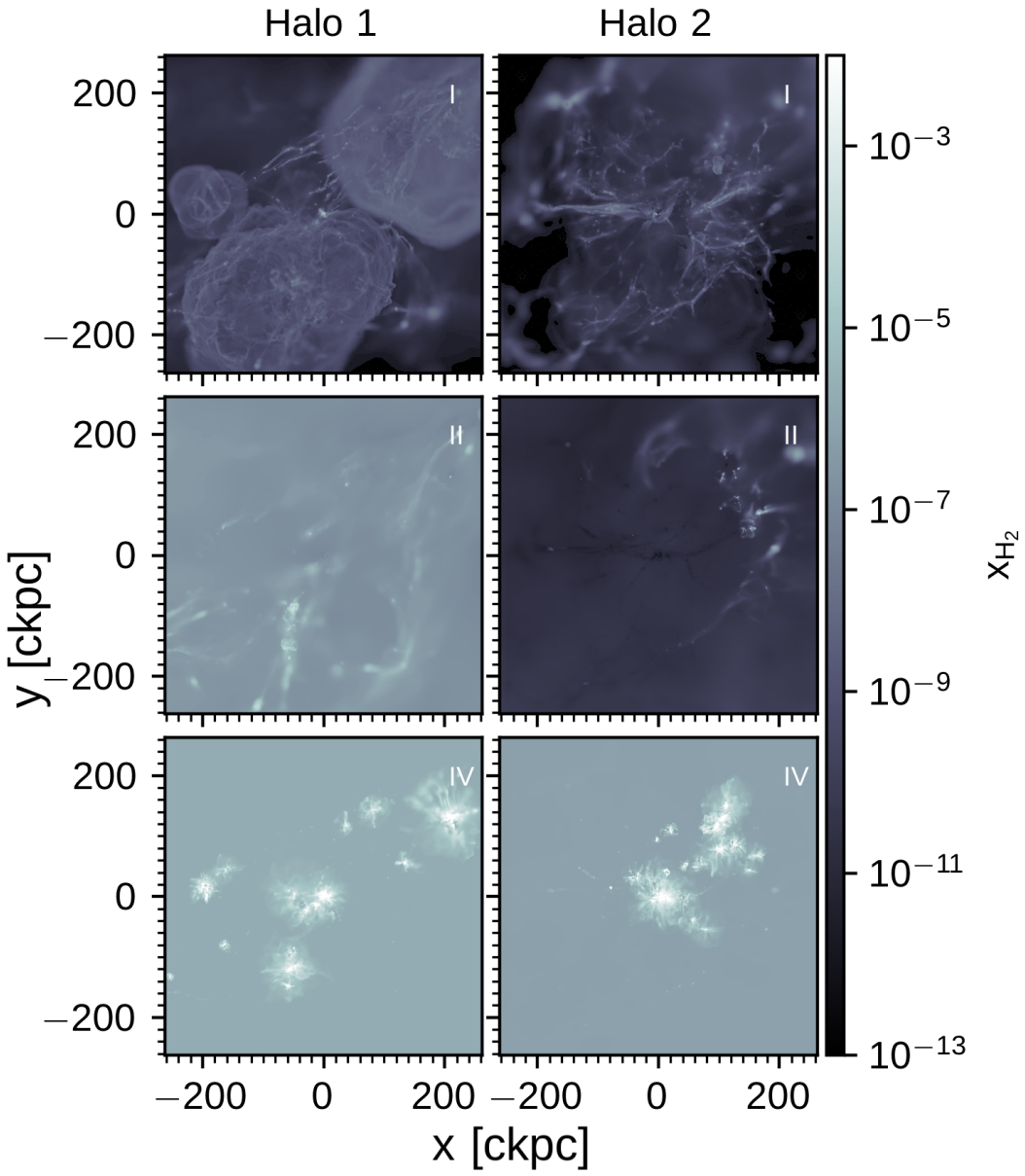


Figure 4.4: Maps of the average molecular hydrogen mass fraction along the line of sight (weighted by density) for each of the physics prescriptions (excluding III and V), for each set of ICs (Halo 1 and Halo 2).

4.3.1.1 Final Thermochemical State

2D histograms of the subhalo gas state in (ρ, T) and (ρ, H_2) space (Figures 4.5 and 4.6) show large differences between the prescriptions, giving insights into the chemistry. All prescriptions but IV exhibit a 'plateau' of abundant gas at $\sim 10^4$ K, which is the limit below which purely atomic hydrogen cannot cool efficiently. This is not present in IV as the gas forms so much H_2 so uniformly that the gas in such a plateau would be rapidly cooled below the limit. In IV, there is an abundance of gas between 10-100 K across the whole range of densities, which is expected given the large abundance of H_2, present even in rarefied gas with $\rho < 1$ H/cc (unlike in the other prescriptions). Though Prescription V has a coolant in its rarefied gas in the form of the metallicity floor of $Z = 3.2 \times 10^{-4} Z_\odot$, this is not as strong, and only dense gas that can shield itself from the UV background can stay cool. The two prescriptions with radiation transport (I and II) do not form significant amounts of H_2 except in relatively dense gas, as, not only is the formation rate the highest here, H_2 is more able to self-shield from dissociating stellar radiation, so only this gas can cool below $\sim 10^4$ K. 'Relatively dense' here has a different meaning for Prescriptions I and II — in Figure 4.6, it can be seen that the runs with Prescription I contain gas with a H_2 fraction above 10^{-3} at densities $\gtrsim 1$ H/cc, whereas the majority of mass above this H_2 fraction in Prescription II is at densities $\gtrsim 10^2$ H/cc. In either I or II, Halo 1 has a greater proportion of gas with a H_2 fraction above 10^{-3} at densities between 1-10 H/cc, so both runs have a larger fraction of gas below 10^4 K at these densities than Halo 2.

Only runs with Prescription I track the HD abundance and capture its radiative cooling; Figures 4.8 and 4.7 respectively show the average HD abundance along lines of sight, and the 2D histogram of the gas state in (ρ, x_{HD}) space of the most massive subhaloes of each simulation with Prescription I. Like H_2, HD only forms significant amounts in dense, cold gas, but due to the low primordial abundance of deuterium ($\sim 2.5 \times 10^{-5}$ of hydrogen nuclei), it is $\sim 10^4\times$ less abundant than H_2 in cells with $\rho > 10^2 m_p$ cm^{-3}.

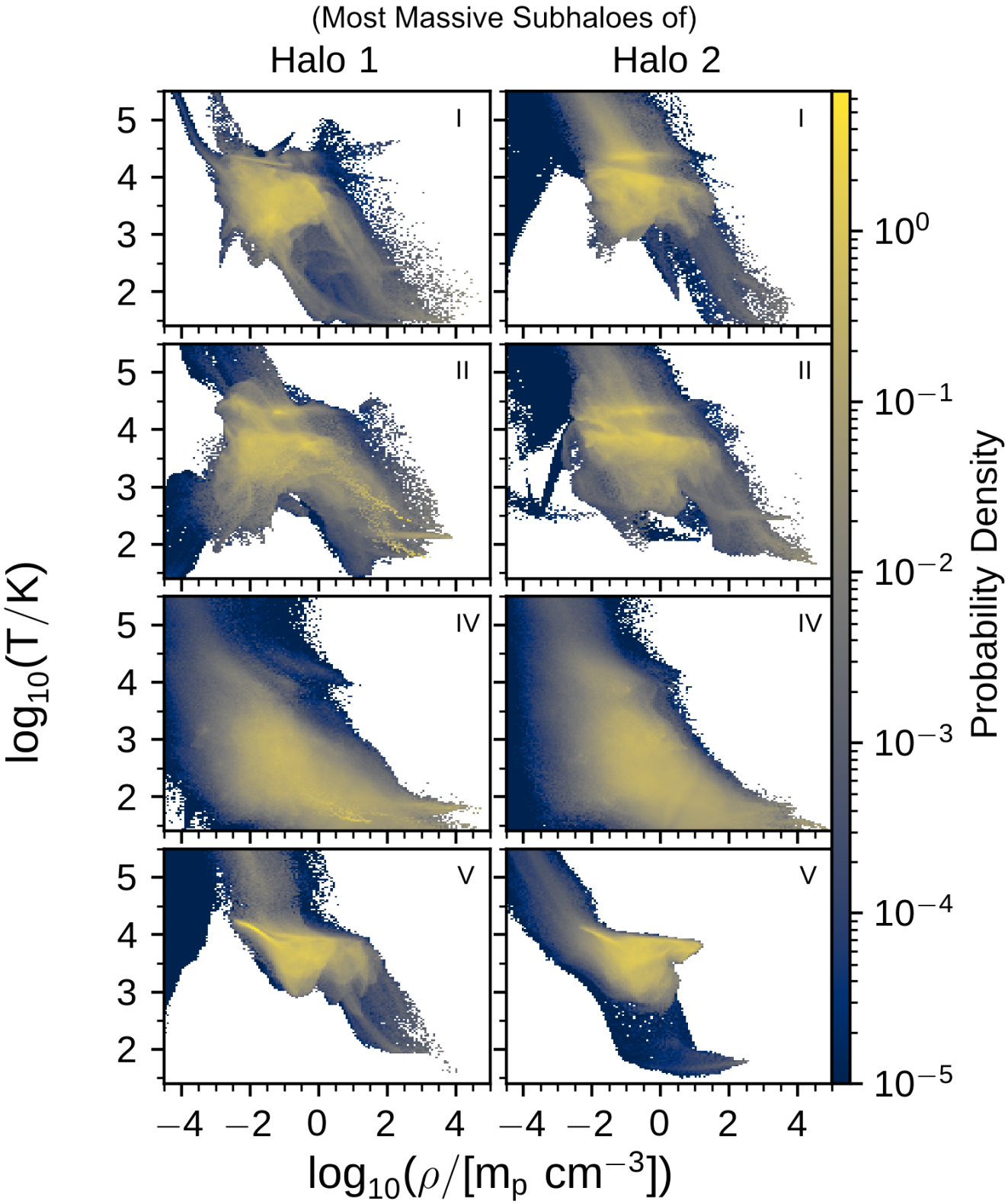


Figure 4.5: Gas phase histograms in (ρ, T) space for each of the physics prescriptions (excluding III), for each set of ICs (Halo 1 and Halo 2).

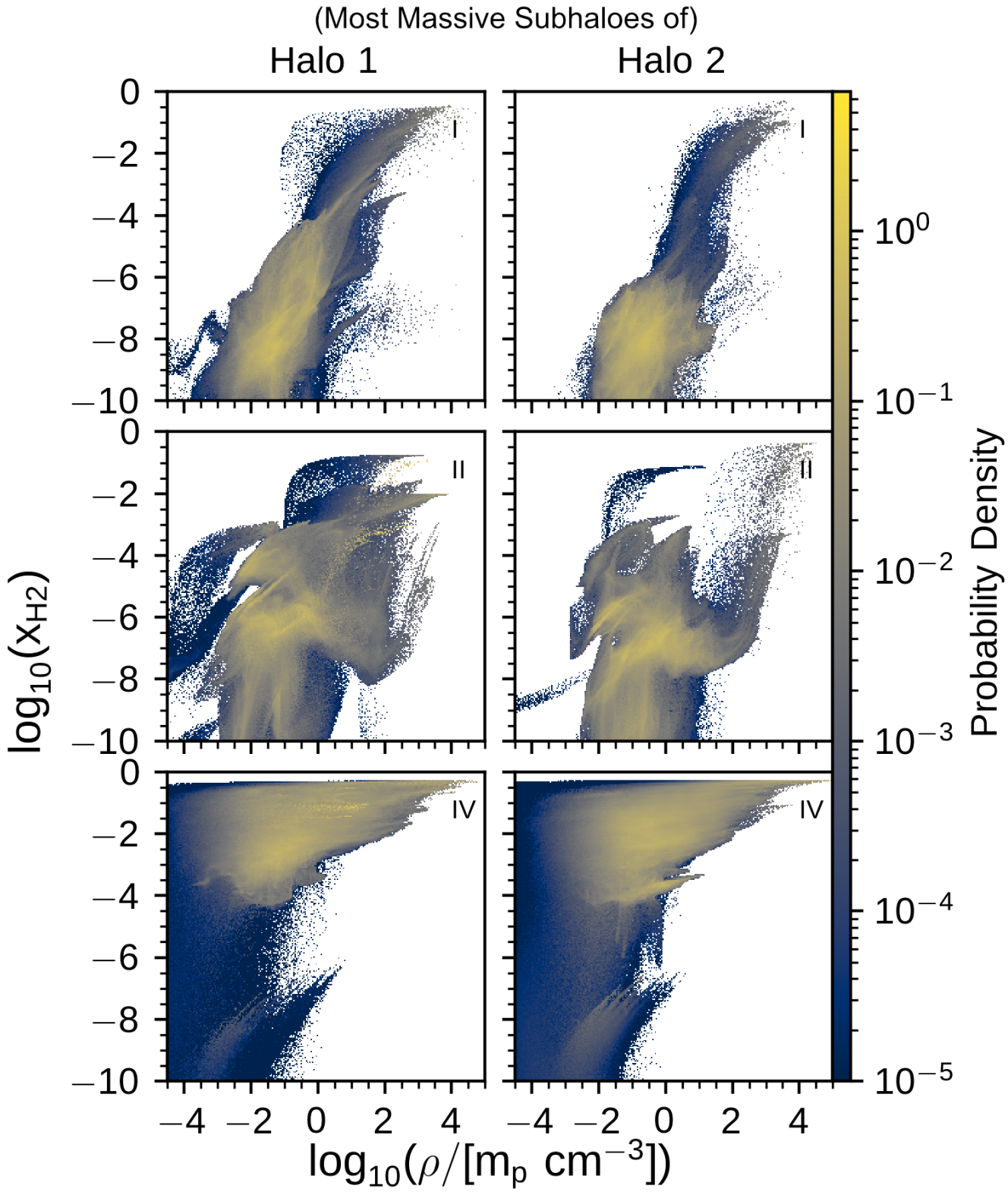


Figure 4.6: Gas phase histograms in (ρ, x_{H_2}) space (where x_{H_2} is the gas mass fraction of H_2) for each of the physics prescriptions (excluding III), for each set of ICs (Halo 1 and Halo 2).

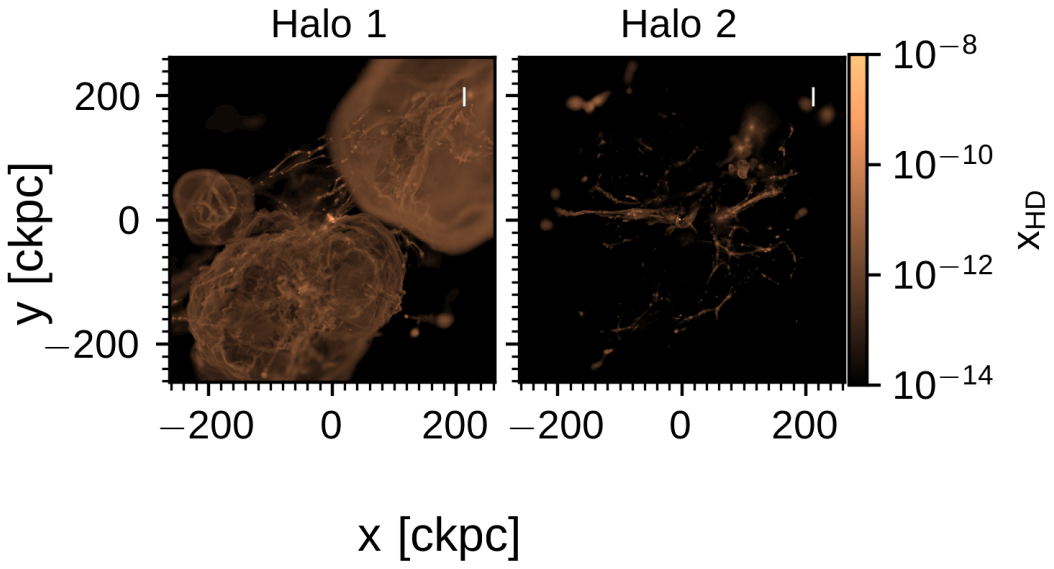


Figure 4.7: Maps of the average HD mass fraction along the line of sight (weighted by density) for Prescription I, for each set of ICs (Halo 1 and Halo 2).

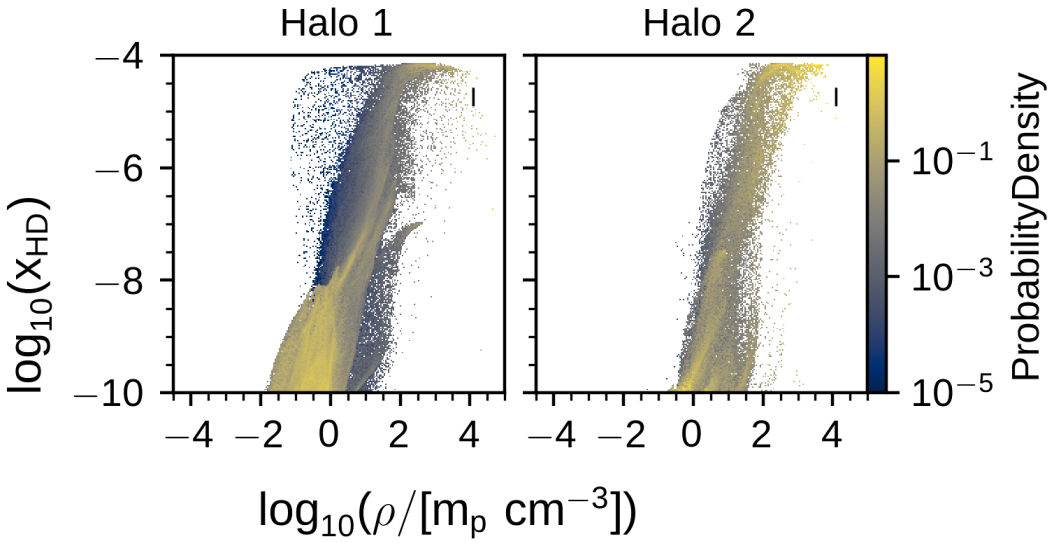


Figure 4.8: Gas phase histograms in (ρ, x_{HD}) space (where $x_{\mathrm{H_2}}$ is the gas mass fraction of HD) for Prescription I, for each set of ICs (Halo 1 and Halo 2).

4.3.1.2 Stellar Mass – Halo Mass Relation

All subhaloes, across all prescriptions, in both sets of ICs, have a stellar mass ($M^\star$) – halo mass (M_h) (SMHM) sequence (plotted in Figures 4.9) lying above the fit to observed galaxies at $z = 0$ by Behroozi et al. (2019). In Prescriptions IV and V, the subhaloes form

$\sim$ 3 orders of magnitude more stars than expected from the fit to observed local galaxies (though they are not expected to match exactly; the line plotted on these figures is an extrapolation to much lower halo masses than those examined by Behroozi et al. 2019, who also accept a strong deviation from the $z = 0$ fit at $z > 5$). It is unsurprising that subhaloes simulated with Prescription IV form the most stars, since, as shown previously, this Prescription leads to the largest abundance of cool molecular gas. Prescription V is more strongly affected by the UV background heating, so only the most massive subhalo is able to shield itself and form stars; however, this subhalo (for each set of ICs) still forms a relatively large mass of stars compared to its dark matter content, in line with Prescription IV. The two non-radiative Prescriptions IV and V are able to form large, highly-clustered, groups of stars efficiently, as they do not account for radiative feedback, whereas in I and II, stars may ionise and expel the star-forming gas in their neighbourhood to some extent (though clusters form with these prescriptions nonetheless, they are mostly not as massive relative to their host subhaloes).

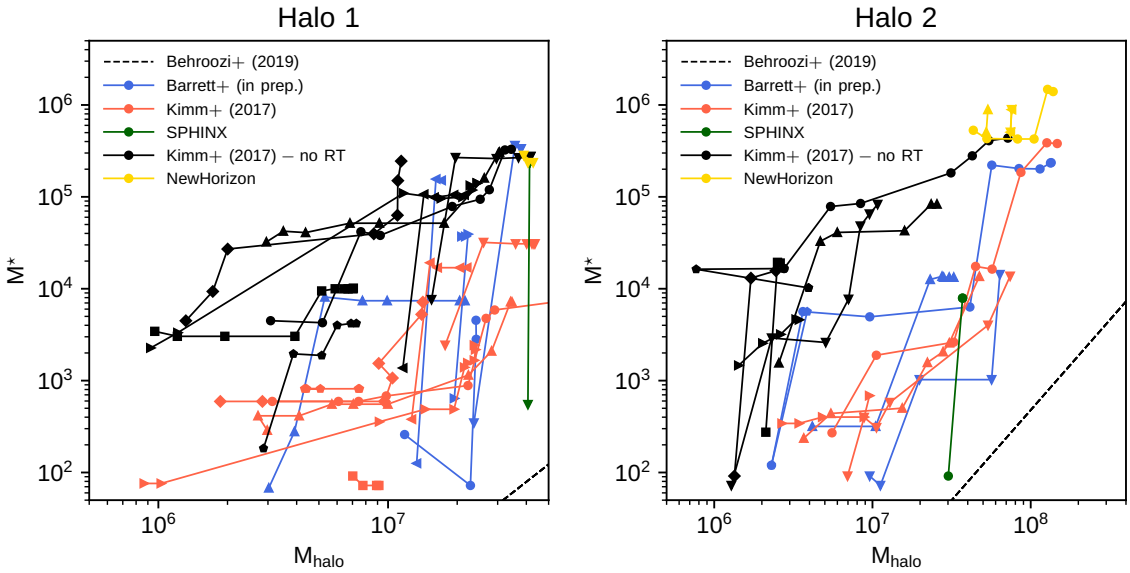


Figure 4.9: Stellar vs. halo masses of subhaloes that contain stars (plotted over many snapshots) in each physics suite for Halo 1 (**left**) and Halo 2 (**right**). Each subhalo is identified (consistently between runs, across all figures) with a different marker style on a separate line. The fit of Behroozi et al. (2019) to the observed Stellar Mass – Halo Mass relation at $z = 0$ is shown by the dashed line.

The SMHM sequence of subhaloes of runs with Prescriptions I and II lie closer to,

but still above, the fit of Behroozi et al. (2019). It appears that the SMHM sequences of subhaloes in I are 'noisier' than those of II (with greater variance between the separate lines), aligning with the fact that I forms more strongly-clustered groups of stars (with some of the most massive subhaloes not forming any stars) than II, showing that stars form more homogeneously in II; this could be explained by a decreased sensitivity of star-forming regions to their environment (indeed, II assumes equilibrium concentrations of H_2-forming reactants, regardless of the radiation field).

4.3.1.3 Metal Enrichment and Chemical Evolution

After stars form, they influence their environment via radiative feedback (dissociating and ionising gas around them), supernova feedback (injecting energy, momentum, and metals) and through stellar winds (which are not modelled here). The differences in the treatments of radiation and chemistry between prescriptions lead to differences in the star formation, and thus also in the resultant radiation field, metal enrichment, and chemical state.

Figure 4.10 shows the mass fraction of hydrogen that is ionised in each simulation as a function of redshift. Runs with prescription IV cannot ionise a large fraction of their neutral hydrogen, as the radiation from stars is not accounted for, and the IGM stays cool despite the UV background, so ionisation is mainly a result of gas heated by supernovae. In Prescription V, since no stars form until $z \approx 11$ in Halo 1 and $z \approx 13$ in Halo 2, the only significant source of heating before then is the UV background, so the ionisation histories for each halo are identical, with negligible ionisation before $z \approx 13.5$. Runs using Prescription I are consistently less ionised than those using II at all redshifts, and also have a much more 'bursty' ionisation history.

As it forms the greatest mass of stars, Prescription IV also leads to the greatest enrichment of the IGM with metals (shown in Figure 4.11). In V, the metallicity floor at $Z = 3.2 \times 10^{-4} Z_\odot$ means that these runs begin with a much higher mass of metals in the zoom region than the others. In Halo 1, runs with Prescription I produce $\sim 5\times$ more metals than with II, but end with a similar mass of metals (and IGM metallicity) in Halo 2. The metallicities (and metal masses) of the subhaloes that contain stars (presented in

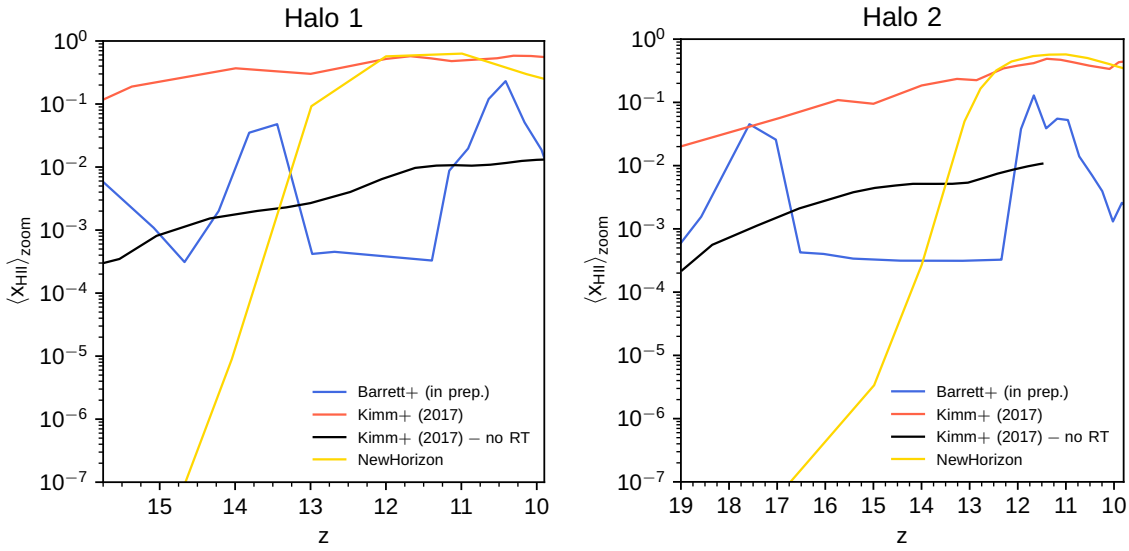


Figure 4.10: Mass-weighted average hydrogen ionisation of the simulation zoom region with each physics suite for Halo 1 (**left**) and Halo 2 (**right**). As the NewHorizon prescription does not track non-equilibrium ionisation, this is calculated after the fact by assuming collisional ionisation equilibrium (equal recombination and collisional ionisation rates) given the gas temperature in the snapshot.

Figure 4.12) do not follow the same trend, with those of Prescription II and IV staying at an approximately constant level, with II at $\sim 10^{-3} Z_\odot$ and IV at $\sim 10^{-2} Z_\odot$, at $z > 13$ for Halo 1 and $z > 15$ for Halo 2. The subhalo metallicity in runs with Prescription I, like the ionisation history, is significantly more bursty, likely indicating either burstier star formation and/or supernova feedback histories (studied in the next subsection).

Since there is no radiative feedback in Prescription IV, subhaloes in these simulations have the highest fraction of H_2 (shown in Figure 4.13). The H_2 fraction as a function of redshift, like the metallicity history, is noisier with Prescription I than II, but has no discernibly different trend across both sets of ICs, except at the final redshifts (both near $z \sim 9.8$), where the H_2 fraction becomes very low. This is most likely not coincidental: the simulations are much slower to run during periods of rapid star formation, as the high temperatures/pressures created by radiative and supernova feedback mean that a shorter time step must be chosen to satisfy the CFL condition (Eq. 2.59), and in these same periods the rate of destruction of H_2 is highest; thus, after running the simulations for a given wall clock time (as I have done), rather than to a given redshift, it is probable that they would

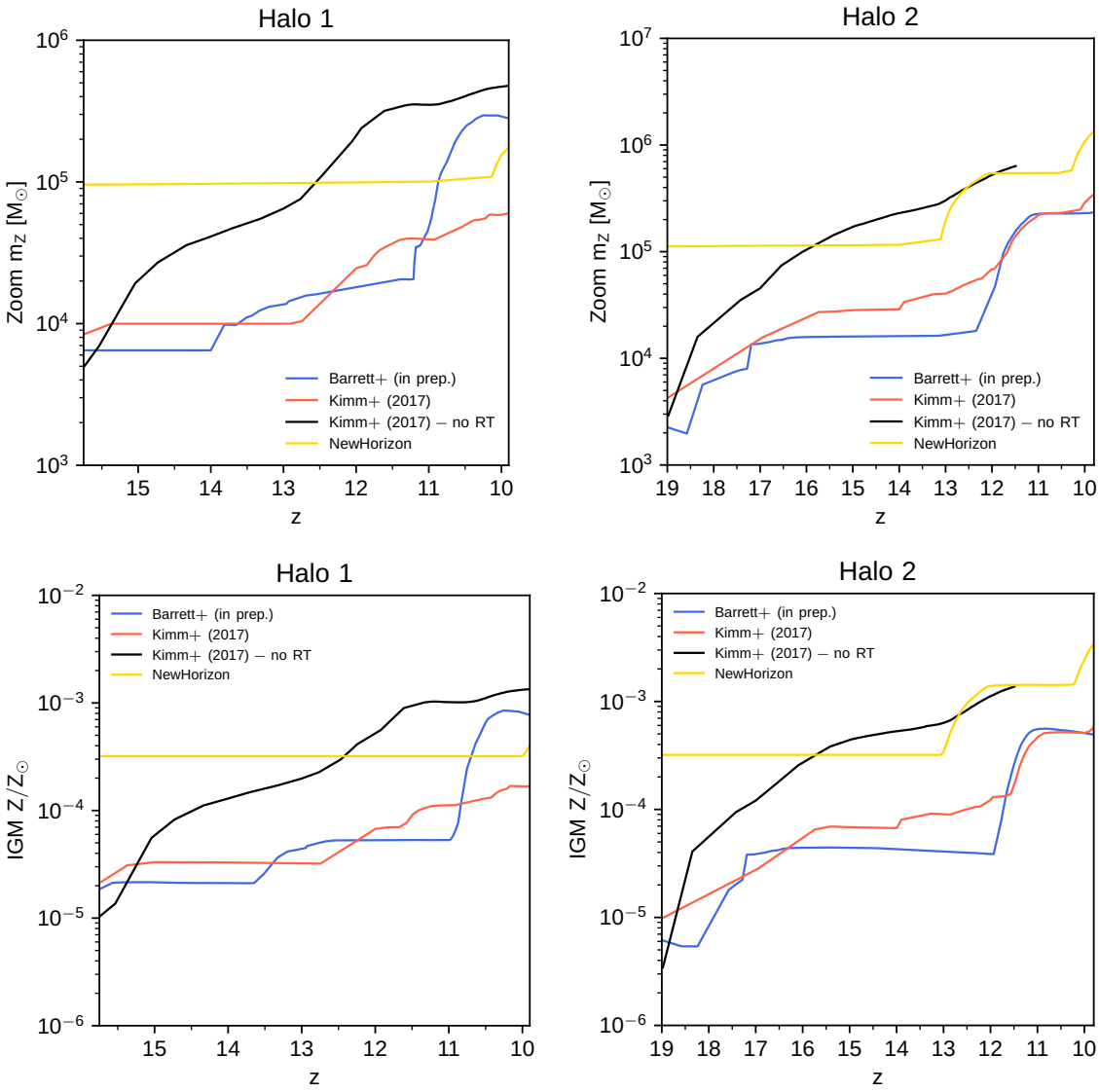


Figure 4.11: Top: Total mass of metals in the zoom region for Halo 1 (**left**) and Halo 2 (**right**) for each of the physics prescriptions. **Bottom:** Mass-weighted average IGM metallicity for Halo 1 (**left**) and Halo 2 (**right**). Here, the IGM is taken to be the volume inside the zoom region, excluding the spheres within the virial radii of each of the 8 most massive progenitor subhaloes.

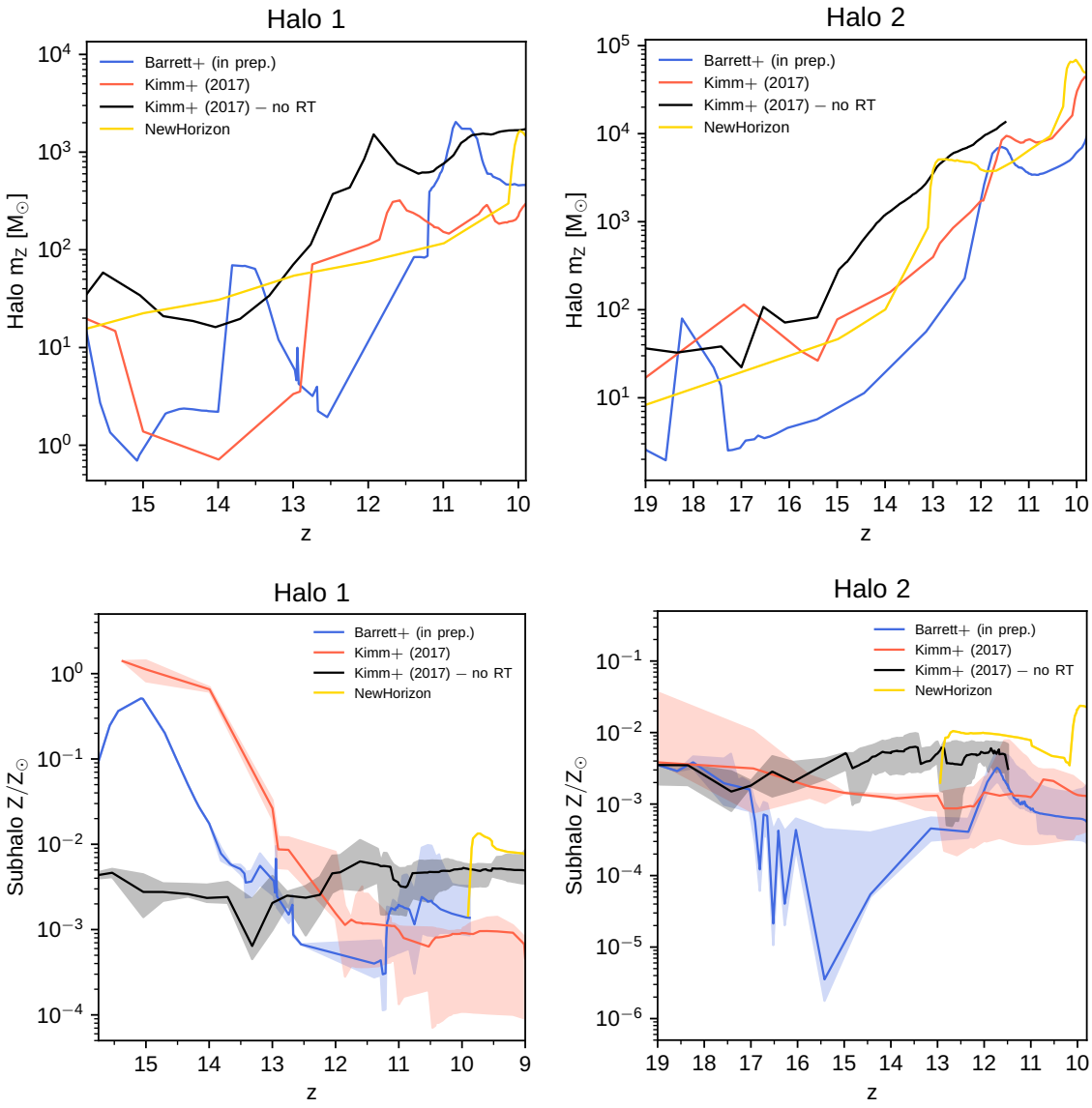


Figure 4.12: Top: Total metal mass in subhaloes in Halo 1 (**left**) and Halo 2 (**right**) for each of the physics prescriptions. **Bottom:** Median mass-weighted subhalo metallicity for Halo 1 (**left**) and Halo 2 (**right**). Of the 8 most massive subhaloes, only those with stars are included in the average. Shaded regions are bounded by the upper and lower quartiles.

end during a period of rapid star formation (assuming that such periods are equally likely as periods of slow star formation in simulated time).

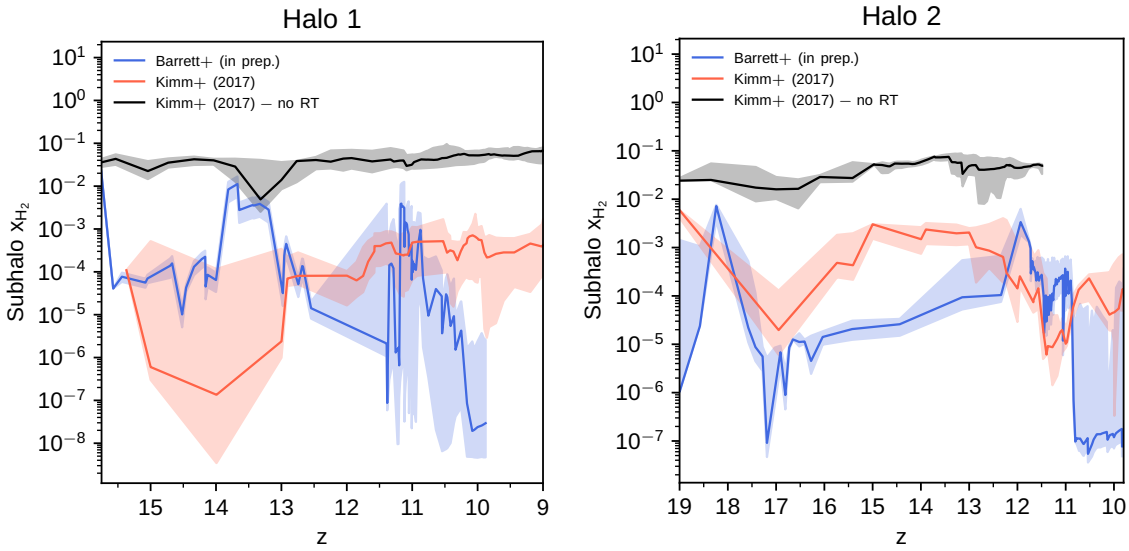


Figure 4.13: H_2 mass fractions in the 8 most massive subhaloes for each physics prescription with non-equilibrium H_2 formation, in Halo 1 (**left**) and Halo 2 (**right**). Shaded regions are bounded by the upper and lower quartiles.

4.3.2 Star Formation and Supernova Feedback

As stars inject metals and energy (in the form of radiation or supernova feedback) into their environments, many of the properties of gas in subhaloes shown in the previous subsection are closely coupled to the time and quantity of star formation in the subhaloes. In this subsection, in order to find the cause of the differences in subhalo properties between physics prescriptions, I investigate and compare star formation and supernova feedback in each of the simulations. Since Pop. III stars are most sensitive to the difference in H_2 chemistry between the prescriptions, these are analysed separately from Pop. II stars.

Figure 4.14 shows the star formation rates in each simulation over time. The onset of star formation is latest with Prescription v (at $z \approx 11.3$ in Halo 1 and $z \approx 13.3$ in Halo 2), as expected given the lack of primordial coolants H_2 and HD; the metallicity floor of $Z = 3.2 \times 10^{-4} Z_\odot$ is calibrated to form the first stars at $z \sim 15$ in SPHINX (Rosdahl et al., 2018), and is not intended to approximate the primordial cooling in the prescriptions that

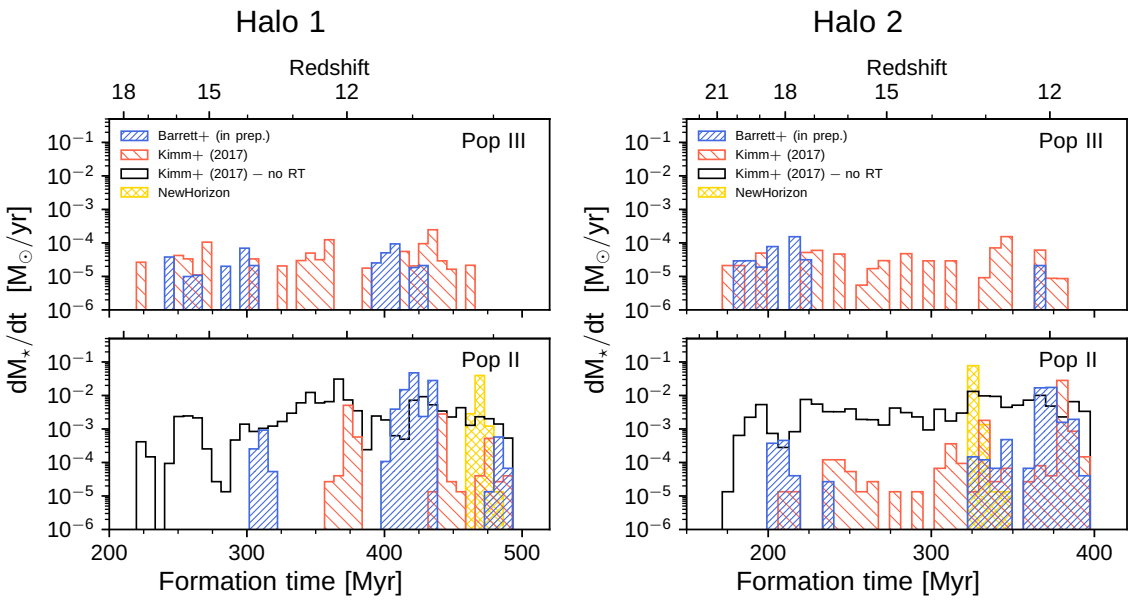


Figure 4.14: Mass of population III (**top**) and II (**bottom**) stars formed per year in each physics suite for Halo 1 (**left**) and Halo 2 (**right**).

use it, nor is the tuned redshift value of 15 universal across simulation configurations and initial conditions. The simulations with prescriptions II and IV both begin to form stars at $z \approx 17.5$ in Halo 1 and $z \approx 20.5$ in Halo 2, (unsurprising, given that their H_2 model is the same without the presence of radiation). Simulations with Prescription I begin star formation slightly later than those with II.

To quantify the 'burstiness' of formation, I use the *variance-to-mean ratio* (VMR) of the inter-star-formation-event times (in Myr, labelled τ)

ICs	Prescription	Pop. III Formation τ VMR	Pop. II Formation τ VMR
Halo 1	I	34.36	44.39
	II	10.17	26.37
	IV	-	1.79
	V	-	0.86
Halo 2	I	94.23	47.62
	II	5.68	7.70
	IV	-	0.47
	V	-	3.87

Table 4.4: Variance to mean ratio of inter-formation-event times (in Myr) in each run for Pop. III and Pop. II stars. Higher numbers indicate more bursty formation histories.

$$\mathrm{VMR} \equiv \frac{\mathrm{Var}(\tau)}{\mathrm{E}(\tau)}. \tag{4.2}$$

The more bursty the formation, the higher the VMR; an exactly periodic history of formation events would have a VMR of zero, and an exponential distribution would have a VMR equal to its mean. The VMRs of Pop. III and Pop. II star formation for each run are presented in Table 4.4. Prescription I, and II cause significantly more bursty formation (in both populations) than IV and V, as shown in Figure 4.14. The VMRs of formation in I and II are difficult to compare given the stochastic nature of the star formation model (and inaccuracy introduced by floating point rounding errors), which effectively lead to 'noise' in the results. This is discussed further in Section 4.3.3, addressing the comparison of this result, and others, under Prescriptions I and II quantitatively.

It can be seen in Figure 4.14, Figure 4.15 (a histogram of the initial masses of star particles), and Table 4.3 that Prescription II forms $\sim 2.5\times$ the number and mass of Pop. III stars than Prescription I in both sets of ICs. Other than in the normalisation, there is no significant difference between the Pop. III IMF observed in any of the runs, which is expected given that the mass of each Pop. III star is sampled from a prescribed Salpeter-like IMF with the same prescription for the efficiency (Section 2.1.7.1), and the only influence the host cell has on the mass of a star that it forms is in enforcing that the star cannot form with more mass than already exists in the cell.

The metallicity distribution of stars $p(Z_\star)$ (i.e. the distribution of the metallicity of the gas they formed from) is shown in Figure 4.16. In Prescription IV, stars form with a $p(Z_\star)$ resembling a power law, cut off at $Z_\star \approx 0.025 Z_\odot$, with almost every bin populated. Stars in Prescription V cannot, of course, form with a metallicity below the floor. Other prescriptions follow a similar distribution, but with 'gaps' at the lower metallicity end, except in the lowest metallicity bin (which is occupied by Pop. III stars in prescriptions I and II), which is well-populated (since every subhalo is initially metal-free). Prescriptions I and II do not fully-populate the low metallicity bins as, once the first stars form, feedback suppresses formation in areas that cannot cool rapidly enough to counteract the heating. The distribution of stellar metallicities in Prescription II is cut off at $Z_\star \approx 0.01 Z_\odot$,

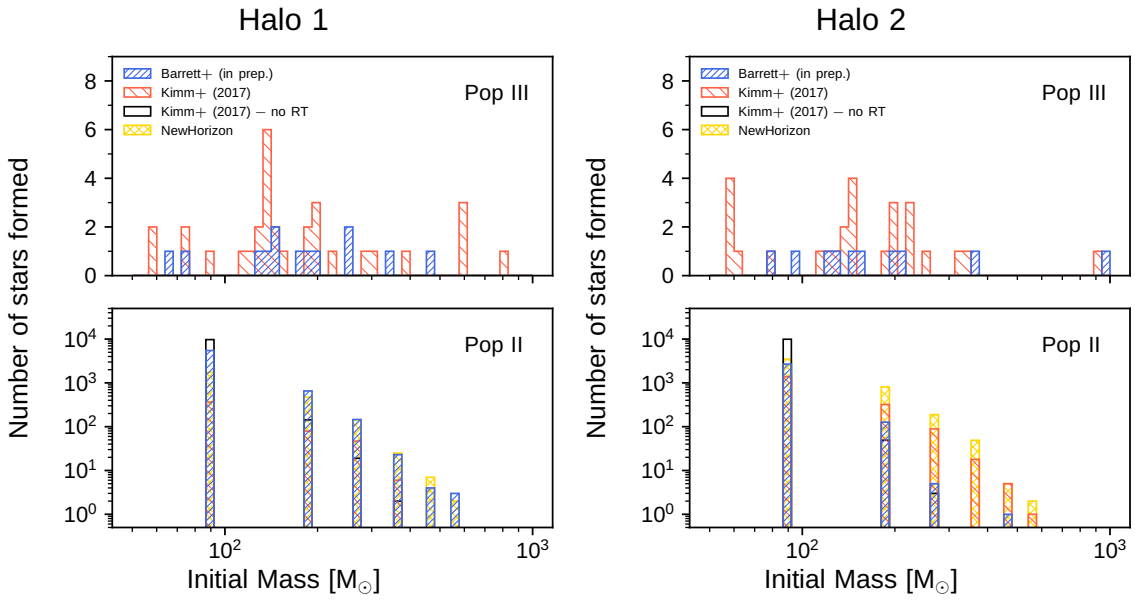


Figure 4.15: Initial masses of all population III (**top**) and II (**bottom**) star particles formed in each physics suite for Halo 1 (**left**) and Halo 2 (**right**). Pop. III particles represent individual stars, whereas Pop. II particles represent a population of equivalent mass, so the bottom panel does not represent a physical IMF, but rather the initial representative masses of populations. The initial mass of a Pop. II particle is an integer multiple of the minimum mass, where this integer is sampled from a Poisson distribution such that the expected value of the particle mass is consistent with the mass given by the Schmidt law in the star formation model (Section 2.1.7.1).

$\sim 2.5\times$ lower than the others, possibly indicating that it is less likely for stars to form adjacent to previously-existing clusters (where the medium is more likely to be enriched by supernovae).

In Figure 4.17, the difference in the distribution of stars between the prescriptions (seen earlier, qualitatively, in Figure 4.1) is shown in a histogram of the separation of all pairs of stars, for Populations III and II. Additionally, Table 4.5 shows the VMR of the comoving distance between all star pairs in both populations for each run. Prescription I appears to lead to more clustered Pop. III stars than II, but the difference is slight, and not consistently apparent for both sets of ICs in Pop. II stars (Section 4.3.3 addresses this comparison statistically). In Prescription V, the stars are present in only a single cluster (their distribution of separations is unimodal), but are still spread out quite uniformly in this

ICs	Prescription	Pop. III pair-separation VMR	Pop. II pair-separation VMR
Halo 1	I	107.03	224.00
	II	75.38	202.68
	IV	-	115.89
	V	-	0.17
Halo 2	I	71.13	136.62
	II	54.53	162.62
	IV	-	76.94
	V	-	0.92

Table 4.5: Variance to mean ratio of comoving distance (in kpc) between pairs of Pop. III and Pop. II stars in each run. Higher numbers indicate more tightly clustered stars (where the variance in their separations is larger at a given mean).

cluster, so it has by far the lowest separation VMR. Stars in Prescription IV, despite having separation distributions that are relatively uniform in appearance (seemingly much more so than those of I and II), have a VMR that is at least of the same order of magnitude as those in I and II.

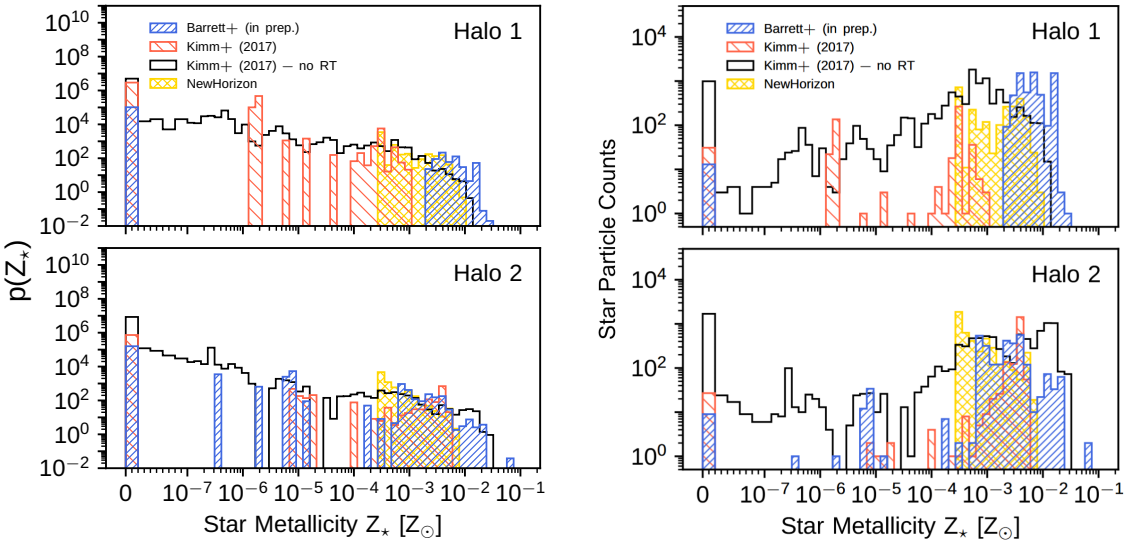


Figure 4.16: Probability density function of metallicities of stars (**left**) and counts of star particles in each metallicity bin (**right**; as in the left panel, but no longer dividing by bin width) formed in each physics suite for Halo 1 (**top**) and Halo 2 (**bottom**). In Prescriptions that model the formation of Pop. III stars (I and II), these are contained in the leftmost bin.

Though supernovae originate from the death of stars, they do not necessarily occur in an identical environment to that which their host star formed in. Figure 4.18 shows

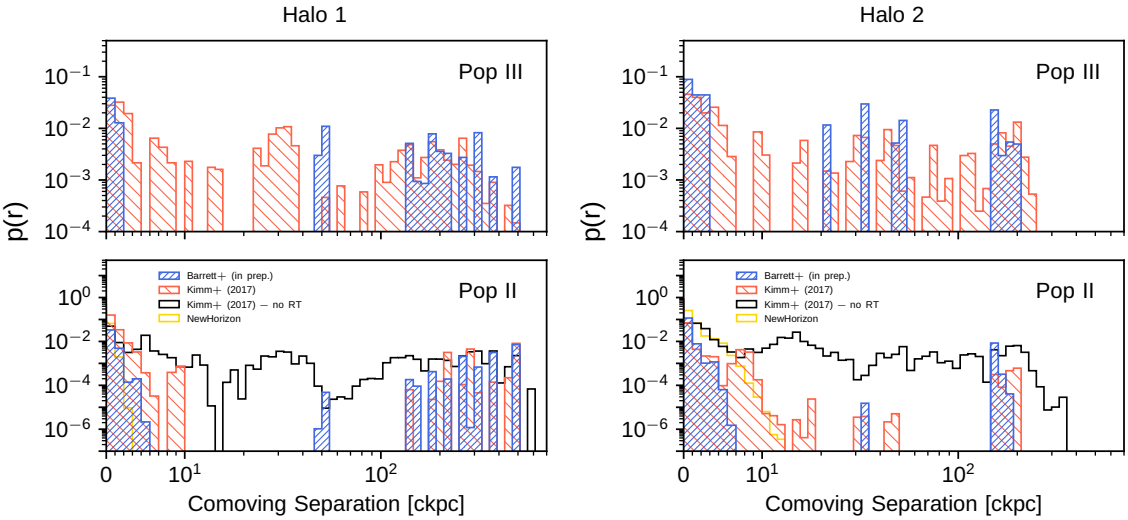


Figure 4.17: Histograms showing the probability density for the comoving separation r of pairs of Pop. II (**bottom**) and Pop. III stars (**top**) in Haloes 1 (**left**) and 2 (**right**) in the latest snapshots for each run.

that the majority of supernovae occur in gas with density $\rho < 10^2 m_p$ cm^{-3} (with similar distributions for each prescription for a given Halo), whereas stars only form above $\rho \gtrsim 10^4 m_p$ cm^{-3} in all prescriptions. This occurs because radiation from stars heats and drives out the gas in their environments (along with supernovae from other nearby stars), so that when a supernova occurs, it may be in a region whose star-forming gaseous material has been evacuated. The distribution of star-forming gas density is very similar for Prescriptions I and II in each Halo, with approximately-equivalent cut-offs at either end. In Halo 1, the density distributions of Prescriptions IV and V are also similar to the rest, though with slightly different cut-offs at the upper end. In Halo 2 there are more significant differences, with Prescription V's distribution extending out beyond $\rho > 10^7 m_p$ cm^{-3}, and much less populated at $\rho \lesssim 10^{5.5} m_p$ cm^{-3} — suggesting that gas at below this density may be overheated compared to other prescriptions.

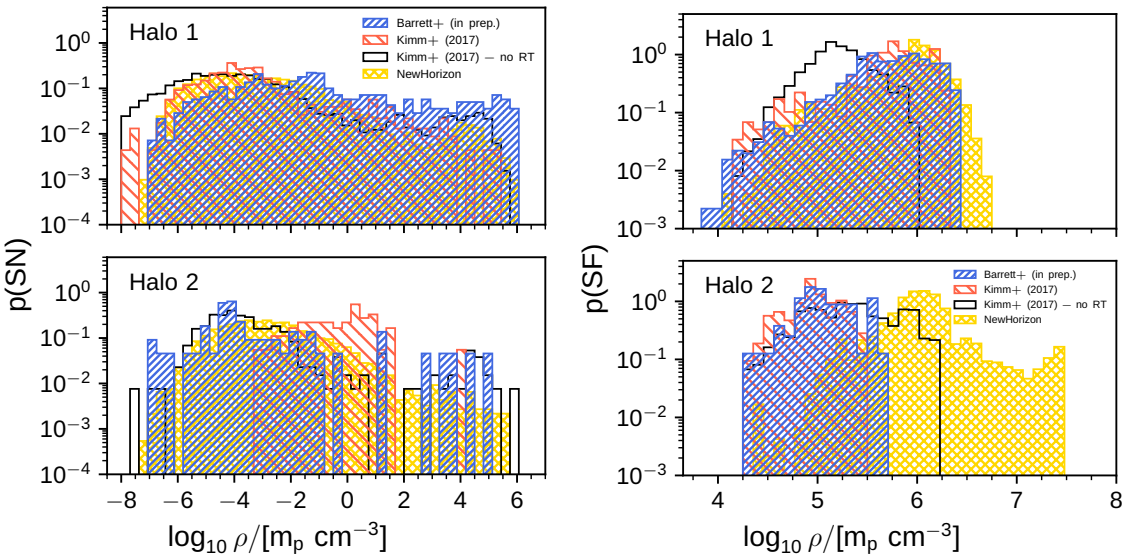


Figure 4.18: Probability density function of gas density at supernovae sites (**left**) and star formation sites (**right**) in each physics suite for Halo 1 (**top**) and Halo 2 (**bottom**).

4.3.3 Chaos and the Stochasticity of Star Formation

There are two potential sources of non-deterministic behaviour in the simulations in this chapter (Keller et al., 2019), despite the fact the physical model being solved is a deterministic one. The first is that the order of application of binary operations in parallel reductions is not guaranteed to be the same between any two invocations of the same program. Floating point addition/subtraction is non-associative (i.e. $x + (y + z) \neq (x + y) + z$), incurring a *rounding error* that is different depending on the order of operations. These non-deterministic rounding errors may accumulate over many parallel reductions, leading to large differences in the output of simulations of sufficiently chaotic systems with identical initial conditions (Genel et al., 2019). This amplification of very small differences into macroscopic ones by chaotic systems is famously known as the *butterfly effect*.

The second source of non-determinism is the random sampling of star masses from the IMF (described in Section 2.1.7.1). This is done using a pseudo random number generator, or PRNG, (since programs cannot produce true random numbers without external random input) which are usually seeded differently in each parallel process (to avoid spurious correlations). Thus, in Ramses-rt, where the spatial domain is partitioned between processes, star formation in each part of the domain is governed by a differently-seeded

PRNG. This means that if a simulation is run with a different number of processes, or restarted from a snapshot mid-run (which resets the seeds), stars can form with different masses or at different times from the original simulation.

The combination of these effects makes comparing simulations, even with the same initial conditions and physical models, difficult. Each simulation can be considered a sample from the distribution of all possible simulations with the same initial conditions (for every possible order of operations, and random number generated for the star formation model); thus, their results are susceptible to *noise*. Even with perfect knowledge of the distribution of possible results, it is difficult to tell whether any two simulations are different due to differences in their physics treatments, or due to only noise.

To address this, I have run additional simulations using Prescription II with the initial conditions of Halo 1. These are two simulations (prefixed "ORD") with identically-seeded PRNGs — continuous runs (not restarted) with the same number of processes — and another three (including the original run presented in the previous subsections) with differing initial seeds (prefixed "RNG", where the original run is RNG3). I choose to run these with Prescription II as I am concerned mainly with the differences between I and II, and II is far cheaper to run. I do not carry out an equivalent study with Halo 2 due to limited computational resources, which may be better spent exploring the distribution of possible simulations of Halo 1, than splitting the samples across both sets of ICs. Results are presented for the latest common snapshot of simulations of Halo 1, $z = 9.87$, to allow direct comparison.

In Figure 4.19 maps of the density, temperature, ionisation, and molecular gas in each run are shown at $z \approx 9.87$ (the lowest common redshift reached in the main suite of simulations). Though they look similar, with the major gas overdensities present in the same place in each run, there are differences in the smaller structures. RNG2 (in the 4th column), and to a lesser extent RNG1 (in the 3rd column), have a string of dense clouds that have formed along a filament (on the upper side of the density maps in the first row), which are visibly suppressed in other runs. Supernova remnant 'bubbles' (especially visible as hot spots in the temperature maps, second row) are also present in different locations, or with different sizes, in each run. Star clusters (marked by white icons on the density maps)

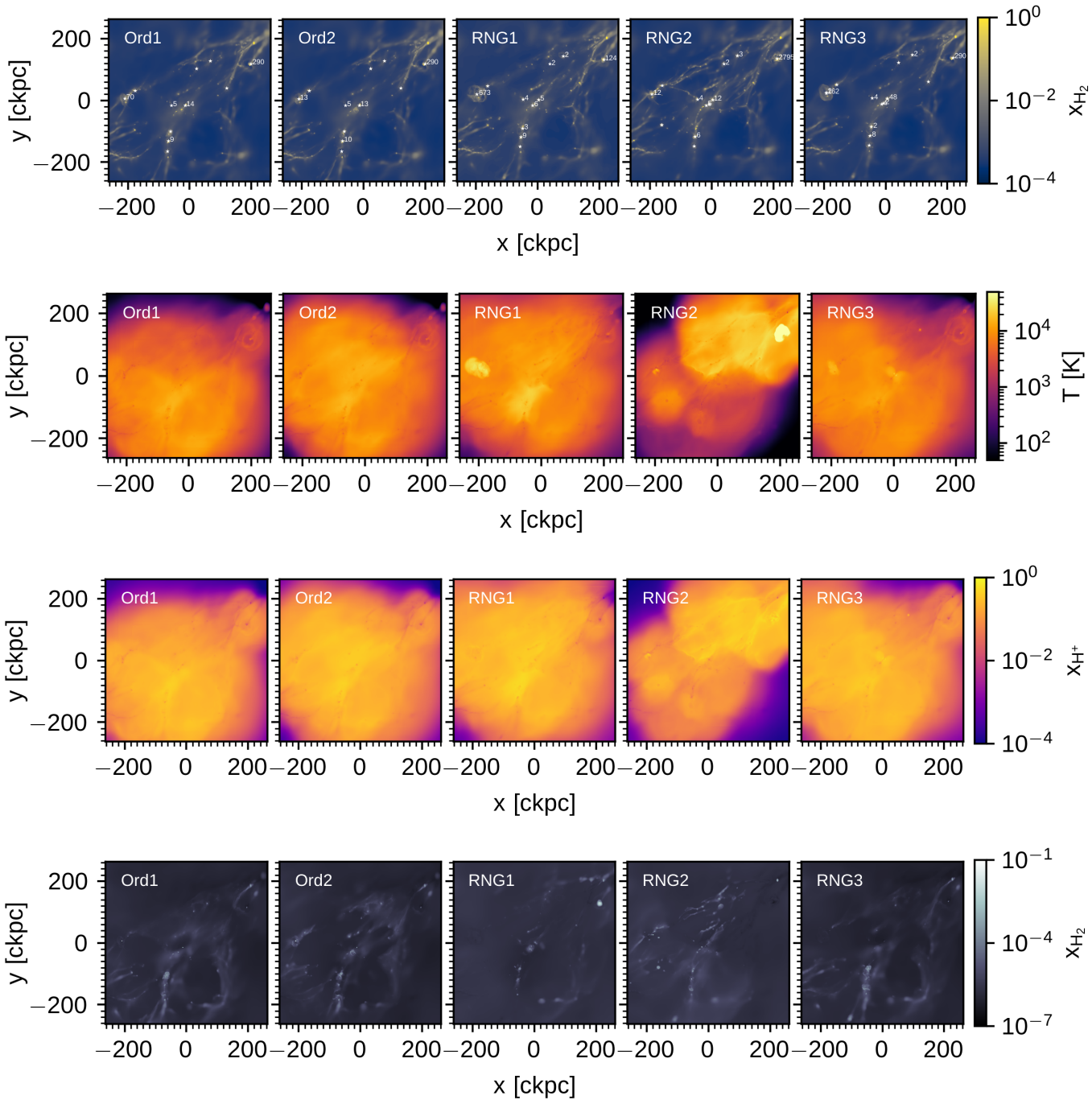


Figure 4.19: Maps of the average gas density, temperature, hydrogen ionisation fraction, and molecular gas mass fraction, along the line of sight (weighted by density) for each of the additional runs with Prescription II of Halo 1 at $z = 9.87$. Clusters of stars are marked by white symbols at their centre of mass, with adjacent numbers indicating the count of star particles (where solitary particles are not marked with a count).

are universally present in the same locations, but these clusters are formed of different numbers of particles, even in the Ord runs with identically-seeded PRNGs. Differences in the ionisation and temperature maps appear to be mostly due to the fact that supernovae have occurred at different times and locations (but the different distribution of stars also must have an effect), and this will also have heavily affected the molecular gas distribution; H_2 is destroyed in gas that is heated far above 10^3 K, but may reform more rapidly in the shocked gas as it cools, as the density is increased near the shock front (creating the 'bubbles' visible in the density map). Qualitatively, the differences between the RNG runs here are small compared to the differences seen between simulations with different physics prescriptions (shown in Figures 4.1 to 4.4). Nonetheless, this demonstrates the effect that noise, particularly that which affects star formation (and thus supernovae), can have on the structure of simulated galaxies.

Run	z_{PopIII}	$\log_{10} M^{\star\ \mathrm{init}}_{\mathrm{PopIII}}$	$\log_{10} M^{\star\ \mathrm{init}}_{\mathrm{PopII}}$	$\log_{10} M^{\star}$	Pop. III τ VMR	Pop. II τ VMR
Ord1	17.3202	3.7603	4.6690	4.5664	10.3432	24.7685
Ord2	17.3413	3.8129	4.6156	4.5131	9.5057	18.8592
RNG1	17.3447	3.8808	4.9888	4.8862	12.0182	34.9066
RNG2	17.3297	3.8343	5.5356	5.4330	6.4048	15.8103
RNG3*	17.3400	3.8391	4.7990	4.6965	10.1690	26.3698

Table 4.6: Properties of the star formation histories of the 5 simulations of Halo 1 with Prescription ii. *RNG3 is the label given to the same simulation of Halo 1 with Prescription ii introduced earlier in this chapter. τ VMR is the variance to mean ratio of inter-formation-event times (in Myr) for the stated population.

Figure 4.20 shows the star formation rate over time in each of the additional simulations. The Ord runs, as expected, start identically, but begin to noticeably diverge at $z \approx 14$ as the accumulated rounding errors become macroscopic. Conversely, the RNG runs immediately diverge, forming their first star at similar times but with different masses, and continuing to produce different (though qualitatively similar) star formation histories. Table 4.6 shows some quantitative properties of the star formation histories of the Ord and RNG runs. As also seen in Figure 4.20, all runs form their first Pop. III star at a similar redshift. However, larger differences are introduced as the simulations progress, especially in the mass of Pop. II stars formed, and the burstiness of formation of both populations.

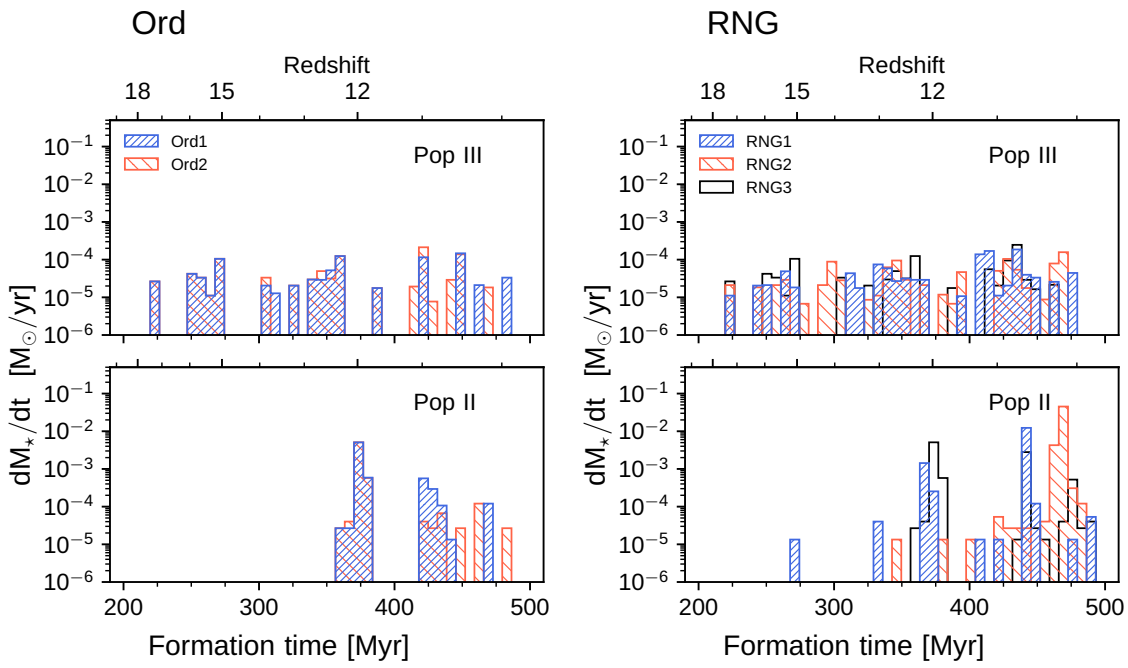


Figure 4.20: Mass of stars formed per year in multiple simulations of Halo 1 with Prescription II. **Left:** simulations with identically-seeded PRNGs, with differences occurring only due to accumulated non-deterministic rounding-errors in parallel reduction operations. **Right:** simulations with differently-seeded PRNGs.

The formation of Pop. III stars is less bursty than Pop. II stars in all of the runs. Burstiness may usually be attributed to feedback: the formation of new stars ionises, heats, and evacuates gas from nearby star-forming clouds, suppressing star formation immediately afterwards. Pop. III stars form in pristine environments by definition, too far away from other stars (if any have yet formed) for any metal enrichment to have reached them. It stands to reason that they would then be less strongly influenced by the radiative feedback from star formation elsewhere in the host halo. Being far from other stars, the regions they form in also contain a relatively high fraction of neutral (and molecular) gas that is opaque to ionising radiation. It is thus not surprising that Pop. III star formation is less bursty than Pop. II.

The VMRs of the comoving separation of star pairs in the ORD and RNG runs are shown in Table 4.7, along with the total mass of metals produced by these stars in the zoom region. RNG2, since it forms the greatest mass of stars, also produces the greatest mass of metals. Pop. III stars are consistently less clustered than Pop. II stars across the runs. The VMR

Run	$\log M_Z^{\rm zoom}$	Pop. III pair-separation VMR	Pop. II pair-separation VMR
Ord1	2.9932	78.81	291.77
Ord2	3.0098	74.90	323.95
RNG1	3.2013	76.76	330.14
RNG2	3.4886	99.34	353.39
RNG3*	3.0816	75.38	202.68

Table 4.7: Zoom region metal masses, and variance to mean ratios of the comoving separation of star pairs (in kpc) the 5 simulations of Halo 1 with Prescription ii. *RNG3 is the label given to the same simulation of Halo 1 with Prescription ii introduced earlier in this chapter.

of the separation of Pop. II stars in RNG3, the original simulation with Prescription ii of Halo 1, is an outlier in this set of runs, demonstrating the danger in interpreting samples of noise-sensitive quantities from a single simulation without knowledge of the distribution they are produced from.

4.3.3.1 Comparing Simulations Under Different Physics Prescriptions

The Ord and RNG runs may be considered samples from the distribution of possible runs, which span a non-zero configuration-space volume due to the accumulation of non-deterministic errors. Thus, the variances in each of the results derived from simulations shown earlier in the chapter represent the level of noise that a 'signal' must stand out from to say with confidence that a simulation with Prescription i (or any of the others) is different from one with Prescription ii *due to only the differences in the physics treatment.*

If the underlying distribution is known, the probability that a given result has been sampled from a distribution other than the 'reference' (or 'null') distribution (under the 'null hypothesis') can be found with a *hypothesis test.* In, for example, a comparison of a result from a simulation with Prescription i with those from the set of runs with Prescription ii, it is possible to calculate the probability that a sample from the *empirical distribution* of results from ii lies in a given interval, but such samples could not extend beyond the limits of the data, so one would infer that the probability of sampling a value that was above the maximum or below the minimum limits would be zero. For such a test to be convincing, a high degree of belief in the empirical distribution would be required, i.e. one must believe that the simulations performed here would never produce results

beyond the observed limits. This belief is not justified by the very small data set from the two Ord runs and three RNG runs. Instead, to perform a quantitative comparison between Prescription II and another, a distribution may be assumed and fit to the data. As the quantities measured from the simulations in this section may be considered the result of combining the contributions of many uncorrelated random or chaotic processes, a normal distribution is a natural choice for these (loosely applying the *Central Limit Theorem*). This is a crude argument, and the choice of distribution may introduce a heavy bias, but this is a bias that will conservatively reduce the strength of the results (by extending the tail compared to the empirical distribution), in line with the belief that the tail of any given distribution of simulation results is as yet unexplored.

In the following analysis, I combine the Ord and RNG samples of any given quantity, assuming that they are drawn from the same underlying distribution. This is another crude simplification, justified by the fact that the Ord runs, despite starting very similarly, are heavily diverged by the final common redshift at which most results are considered (except for the properties of the star formation history). The parameters of a normal distribution are estimated from these 5 samples for each quantity of interest derived from a simulation, and a two-tailed test of a sample from the simulation with Prescription I is performed. This test returns a *p-value* (otherwise known as the *significance*), which is the probability of drawing a sample from the reference distribution with a value equal to or further away from the median (deeper into the tail) than the sample under test.

I am interested in the comparison between Prescriptions I and II, as II is the most advanced other than my own, and the differences between them are subtle, and most relevant to the formation of stars in gas where H_2 and HD cooling is dominant (rather than metal rich gas). These simulations are the most similar of the suite, and thus in the most need of a quantitative treatment to interpret the differences in the context of noise. Table 4.8 shows the fitted parameters to the data in the Ord and RNG runs, the sampled value from the run with Prescription I, and the p-value returned by the two-tailed test, for a selected collection of quantities (samples of which have already been presented in this section for Prescription II in Halo 1). For all quantities except the Population II pair-separation VMR, the hypothesis that the values from the simulation with Prescription I were drawn from

the inferred distribution can be rejected with very high likelihood. However, it must be noted that these values are likely to be very sensitive to the 'fatness' of the tail of the assumed distribution. Nonetheless, in lieu of data that probes the tails better (in either prescription), this analysis is sufficient to show that the differences between Prescription I and II lead to statistically significant differences between the predicted stellar populations.

Quantity	μ in P. II	σ in P. II	Value from P. I	Significance
z_{PopIII}	17.34	0.01	16.11	0*
$\log_{10} M^{\star\ \mathrm{init}}_{\mathrm{PopIII}}$	3.83	0.04	3.41	4.25×10^{-26}
$\log_{10} M^{\star\ \mathrm{init}}_{\mathrm{PopII}}$	4.92	0.33	5.83	6.35×10^{-3}
$\log_{10} M^{\star}$	4.82	0.33	5.73	6.20×10^{-3}
$\log M_Z^{\mathrm{zoom}}$	3.15	0.18	3.75	1.09×10^{-3}
Pop. III τ VMR	9.69	1.84	34.36	0*
Pop. II τ VMR	24.14	6.61	44.39	2.20×10^{-3}
Pop. III pair-separation VMR	81.04	9.25	107.03	4.96×10^{-3}
Pop. II pair-separation VMR	300.39	52.67	224.00	1.47×10^{-1}

Table 4.8: Estimated normal distribution parameters to sampled results from simulations of Halo 1 with Prescription II, and the corresponding values of the results from the simulation of the same ICs with Prescription I. The rightmost column is the p-value from a two-tailed test of the Prescription I value under the null hypothesis that it was drawn from the inferred distribution, i.e. the probability of drawing a sample this far away from the mean in either tail. *A listed significance of 0 indicates that the p-value is so small that it cannot be resolved by a double precision calculation.

4.3.4 Non-Equilibrium H_2 Chemistry

A primary goal of this work is to study the importance of using a fully non-equilibrium treatment of the formation of H_2 in simulations of galaxies. If star-forming gas is significantly out of chemical equilibrium (either in its abundance of H_2, or its reactants), then a non-equilibrium treatment is required to properly resolve its behaviour under radiative cooling.

The concentration of a chemical species at equilibrium depends, in general, on the density, temperature, concentrations of all species in the mixture, and the incident radiation field. However, information about the radiation field is often discarded in such calculations. For example, Kimm et al. (2017) approximate the H_2 formation rate based on the

concentration of H^- in *collisional equilibrium*; this may be problematic[2] in the modelling of gas in the neighbourhood of stars, as electrons are detached from H^- (forming $H + e^-$) by radiation with $h\nu > 0.75$ eV. To investigate the effect of radiation on the position of chemical equilibrium, I compare the concentrations of H_2, and the primary reactants involved in its formation H^- and H_2^+, in runs with Prescription 1 (which solves a complete non-equilibrium network) to the collisional equilibrium concentrations, and then equilibria calculated taking into account the local non-equilibrium electron density (which is dominated by photoionisation after many stars have formed).

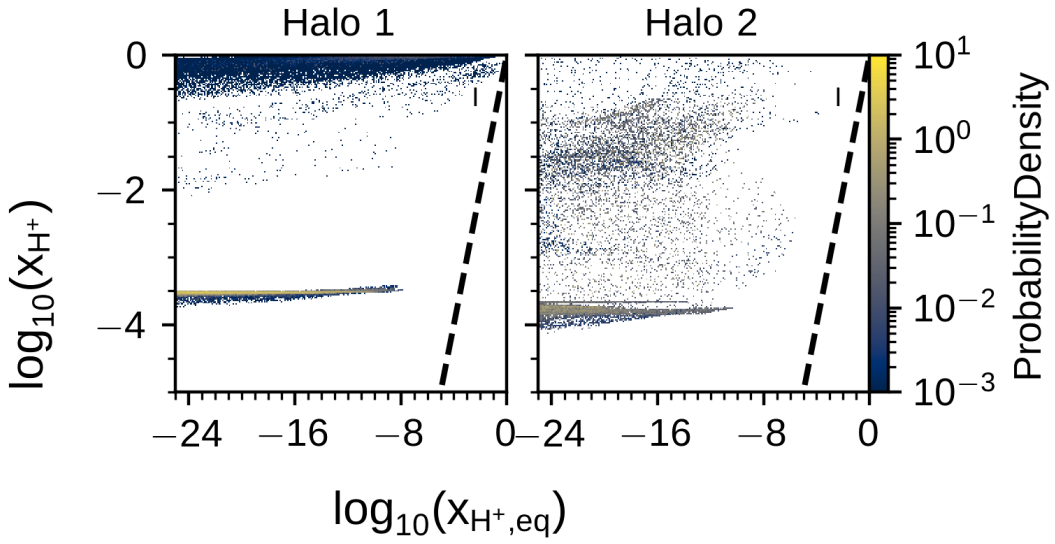


Figure 4.21: Histogram of hydrogen ionisation fraction of gas in cells in subhaloes of simulations with Prescription 1 just after the formation of the first star, with the measured value on the left axis, and the collisional equilibrium value on the bottom axis. The dashed line passes through the points where the measured values would be equal to the collisional equilibrium values.

Figure 4.21 compares the ionisation fraction of gas measured in subhaloes of simulations with Prescription 1 just after the formation of the first star in the simulation (at $z = 16.11$ for Halo 1, and 19.94 for Halo 2) to the ionisation expected in collisional equilibrium. Immediately after the formation of the first star, most gas is much more ionised than it would be in purely collisional equilibrium. This attributable to the ionising effect of radiation from the first star. The dichotomy seen in Figure 4.21 is between highly ionised

[2]However, Kimm et al. (2017) rightly point out that the cooling due to metals will eventually dominate after the first stars form, so this is not always important.

gas around the first star, and a separate phase of mostly neutral gas (but still with ionisation many orders of magnitude more than the collisional ionisation equilibrium) in haloes that are further away and have not yet formed stars (but are still illuminated by the star in the other halo).

As free electrons are needed in order to form H^-, and these are far more abundant than in collisional equilibrium, it would then be expected that the concentrations of H^-, and thus H_2^+ and H_2, should also be far from collisional equilibrium. These chemicals may also be dissociated or ionised in photoreactions with UV photons around Pop. III stars. Indeed, as seen in Figure 4.22, the concentrations of H^-, H_2^+ and H_2 are far greater than the collisional equilibrium values in almost all subhalo gas cells. The collisional equilibrium concentrations depend only on the relative concentrations of reactants, and the temperature. The *independence* of (i.e. low correlation between) the simulated concentrations from the equilibrium concentrations in the majority of cells (manifesting as horizontal structures on the histograms) indicates that the instantaneous reaction rates are not the major factor in determining the current chemical state; this could rather be due to a strong incident radiation field dominating the destruction reactions, or historical formation during a time when the position of equilibrium was different, leaving a higher-than-equilibrium final concentration.

Calculating the collisional equilibrium concentrations of H^-, H_2^+ and H_2, but this time substituting the simulated value of the free electron number density n_e, shows a markedly different result in the comparison of the H^- concentration with equilibrium (shown in Figure 4.23). Now, the measured values of n_{H^-} are almost all significantly below the equilibrium value; this may be explained by the photodetachment of electrons from H^-: it is being destroyed more quickly than expected from the collisional rate alone. The same is not true of H_2^+, since this is not only formed from the combination of simpler atomic/ionic reactants (mainly by $H + H^+$), but also by the ionisation of pre-existing H_2 (which still exists in greater concentrations than expected from this equilibrium calculation). The equilibrium H_2 concentration, like that of H_2^+, is also mostly unchanged by substituting the real electron density, as it is dominated by the concentration of other reactants.

Finally, substituting the measured concentrations of H^- and H_2^+ (in addition to the

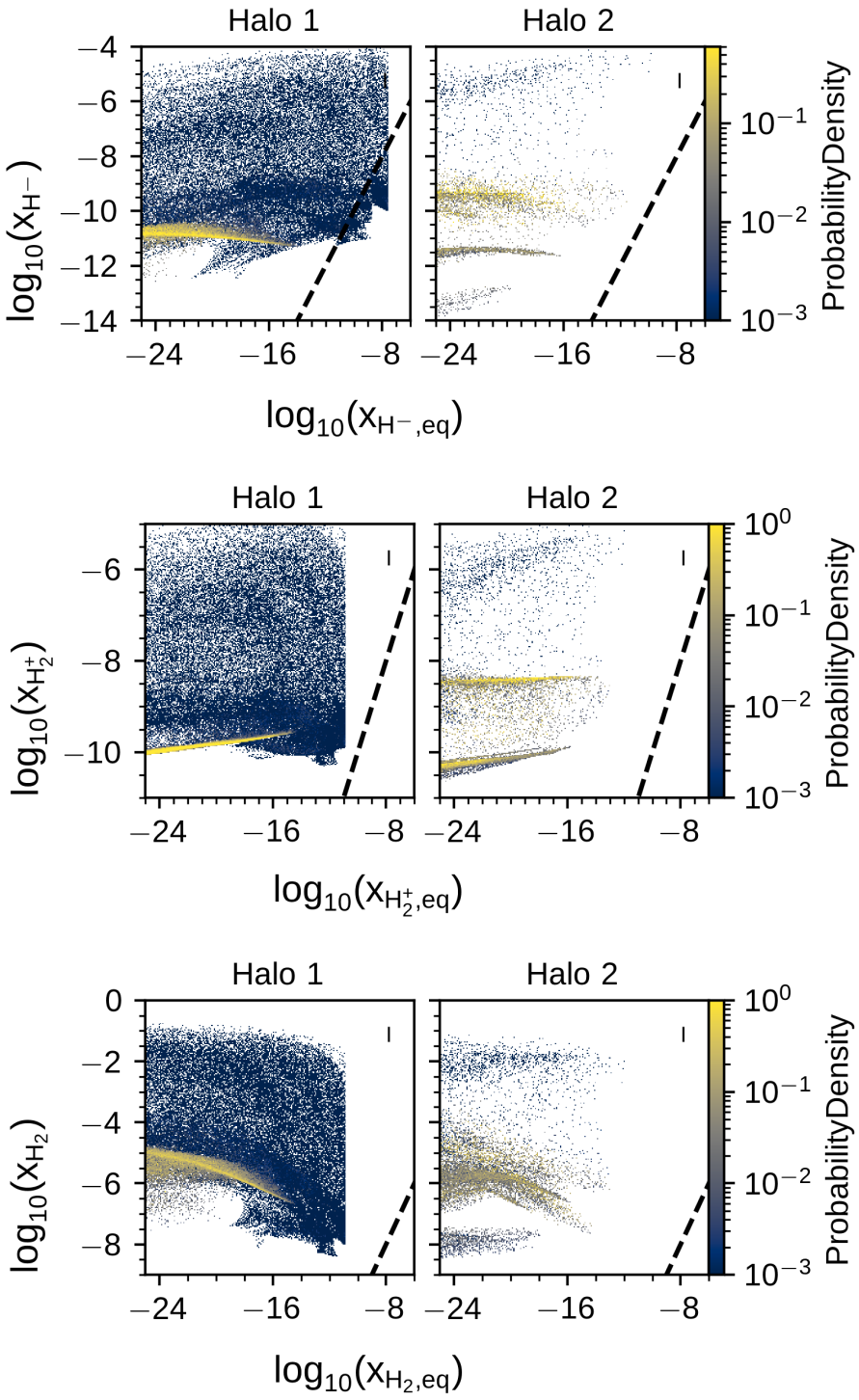


Figure 4.22: Histogram of the mass fraction of H^- (**top**), H_2^+ (**middle**) and H_2 (**bottom**) of gas in cells in subhaloes of simulations with Prescription 1 just after the formation of the first star, with the measured value on the left axis, and the collisional equilibrium value on the bottom axis. The dashed line passes through the points where the measured values would be equal to the collisional equilibrium values.

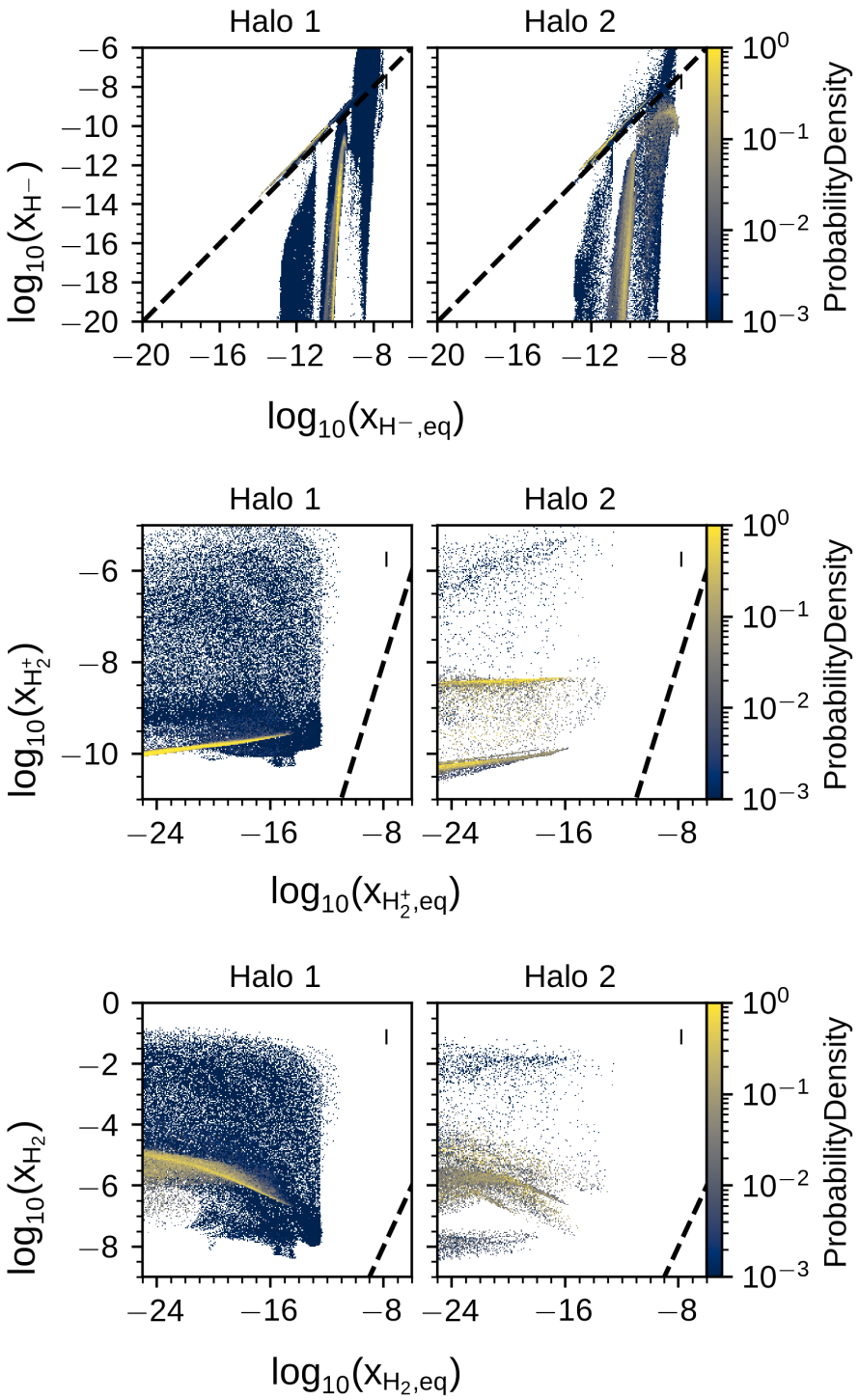


Figure 4.23: Histogram of the mass fraction of H^- (**top**), H_2^+ (**middle**), and H_2 (**bottom**) of gas in cells in subhaloes of simulations with Prescription 1, with the measured value on the left axis, and the collisional equilibrium value, with the measured value of the free electron number density substituted into all other reaction rates, on the bottom axis. The dashed line passes through the points where the measured values would be equal to the collisional equilibrium values.

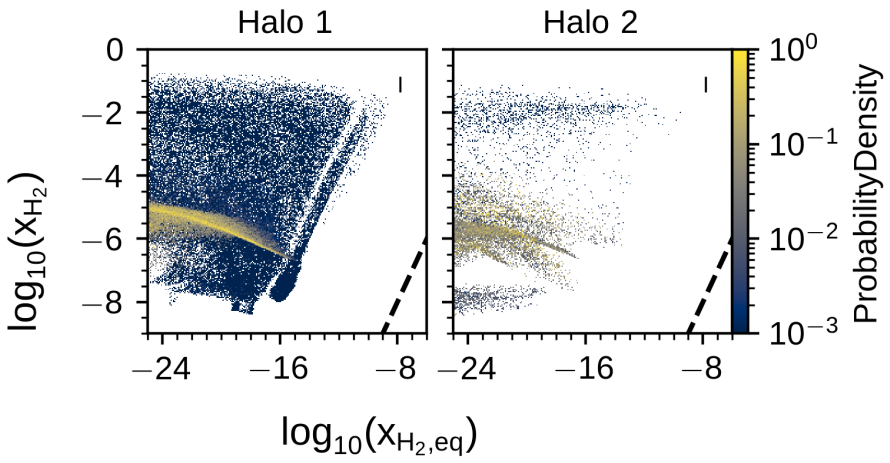


Figure 4.24: Histogram of the mass fraction of H_2 of gas in cells in subhaloes of simulations with Prescription 1, with the measured value on the left axis, and the collisional equilibrium value, with the measured value of the concentration of H^-, H_2^+ and free electrons substituted into the H_2 formation and destruction rates, on the bottom axis.The dashed line passes through the points where the measured values would be equal to the collisional equilibrium values.

free electron number density) into the H_2 formation and destructions, an approximate comparison of the measured H_2 concentration with the equilibrium value — taking into account the actual concentration of all reactants, rather than their collisional equilibrium values — can be obtained (plotted in Figure 4.24). This indirectly considers the effect of the radiation field, via the non-equilibrium concentrations of the reactants, but neglects the effect of photodissociation and photoionisation of H_2. This does not affect my conclusions, as all cells already have a H_2 mass fraction above the computed equilibrium value, despite the fact that these neglected processes would decrease this equilibrium concentration. Despite accounting for non-equilibrium concentrations of reactants, the measured H_2 concentrations are far above the expected values in equilibrium.

4.4 Discussion

4.4.1 Comparison of Simulated Galaxies Under Different Physics Prescriptions

My results show that the choice of numerical treatment of radiation and chemistry can lead to significantly different outcomes in simulated dwarf galaxies. Neglecting radiation transport from stars in the galaxy, and relying on a metallicity floor, rather than self-consistently formed molecular gas, to cool the medium (as in Prescription v, used in NewHorizon; Dubois et al. 2021) leads to highly-clustered, relatively late star formation. By the final commonly-reached redshift in either set of initial conditions, all of the stars were concentrated in one subhalo, with the onset of formation at $z \approx 11$ in Halo 1, and $z \approx 14$ in Halo 2. Reionisation is 'forced' upon these runs by the empirical UV background of (Faucher-Giguère et al., 2009), rather than it being a prediction of the simulation itself. The lack of radiation transport means that radiative *feedback* is not accounted for, allowing stars to accumulate in the same subhalo (the most massive in its host halo, in both sets of ICs) in a large quantity without evacuating the cool, dense, star-forming gas. That is, the supernova feedback alone is insufficient to suppress this clustering. Prescription v is not designed to simulate primordial objects in detail, but rather to produce a set of objects that are statistically-consistent with more local (later time) observations, where molecular gas cooling is much less relevant, and Pop. III star formation has long since ceased. Thus, it is not surprising to see that it is so discrepant with more complete prescriptions that are deliberately focused on the very early universe.

Simulations with Prescription iv, based on that of Kimm et al. (2017) but without radiation transport, demonstrate the drastic effect of ignoring radiative feedback. Unchecked by photodissociation, H_2 forms in high concentrations across the simulation, efficiently cooling the medium and leading to widespread and continuous star formation in many of the subhaloes, in the largest quantity of all prescriptions (and enriching the medium with metals much more rapidly). The supernova feedback is, again, insufficient to regulate the star formation, before the UV background becomes more intense during reionisation. This

Prescription is not at all realistic (problematically ignoring dissociation of H_2 due to the UV background, and local starlight), and the results here demonstrate the importance of radiative feedback to predicting primordial star formation (and the subsequent transition to Pop. II star formation) in haloes of this size.

Prescriptions I (my code) and II (that of Kimm et al. 2017) are broadly similar. They both include radiation transport, with energy bins aligned to the important absorption processes in the chemical species they track, and a non-equilibrium treatment of hydrogen and helium ionisation, and molecular hydrogen formation. They only differ in their H_2 self-shielding model (which is irrelevant to the formation of the very first star), and the extent to which their molecular hydrogen formation model is 'non-equilibrium' (with Kimm et al. (2017) assuming that H_2 formation proceeds only via the H^- channel, and that H^- is at its collisional equilibrium concentration), whereas I use a complete network of the primordial chemistry involved in both H_2 and HD formation, enabling non-equilibrium concentrations of intermediate reactants (also taking into account photodissociation/ionisation reactions that change the position of equilibrium from the purely collisional value). I find that Prescription I forms more stars, in more concentrated clusters (and fewer subhaloes), than II, but does not ionise its medium as quickly. These stars also form in a more 'bursty' manner in Prescription I. Though the H_2 mass fractions of star-forming subhaloes are not obviously different between the two Prescriptions, eventually there are many more cells with a density of $\rho > 10^2 m_p$ cm^{-3} with a H_2 mass fraction of $x_{H_2} > 10^{-4}$ in these subhaloes in Prescription I, which cool to near 10 K, whereas cells at these densities in II remain closer to 100 K (which may be mostly due to the effect of HD cooling in I, which radiates much more efficiently than H_2 below $\sim$ 100 K, as it has a permanent dipole moment that makes the first rotational transition much easier to excite than the quadrupole of H_2; see Figure 1.4). Conversely, at lower densities, Prescription II has more cells with a similarly high H_2 mass fraction. This may explain why Prescription I forms more stars (as there are more cool, dense cells with a high x_{H_2}), and also why Prescription II is able to form stars in a less-clustered distribution spread out over more subhaloes (it forms cells with high x_{H_2} at lower densities, which exist in smaller clouds, despite the total quantity of H_2 in the simulation being similar).

4.4.2 Comparing Simulations of a Chaotic System

Five initially-identical simulations (including one from the original set) of Halo 1 with Prescription II were run in order to understand the impact of non-deterministic 'noise' (accumulated floating point errors, and differences in the mass of sampled stars) on comparisons of simulations with the same initial conditions. It was found that the three of these runs that had differently-seeded PRNGs have significantly different star-formation histories, and even the other two runs with the same initial seeds their histories diverge early on, as the accumulated rounding errors built up.

In each run, the distribution of stars, though covering the same subhaloes, is differently weighted between them, with supernovae then occurring at different times and places. Despite the quantitative differences, the maps (Figure 4.19) of these simulations at the final time appear qualitatively similar, and are recognisably distinct from those of other prescriptions (Figures 4.1 to 4.4).

Indeed, using hypothesis tests to qualitatively compare the simulation of Halo 1 with Prescription I with the five simulations with Prescription II, it is found that the properties of the stellar populations (both Pop. III and II) and their formation histories are significantly different. These differences can thus be entirely attributed to the non-equilibrium H_2 and HD formation treatment in my code, rather than the inherent noise.

4.4.3 Importance of Non-Equilibrium, Radiation-Coupled Primordial Chemistry

The most important difference between Prescription I and II is in their treatment of molecular chemistry. II (Kimm et al., 2017) assumes a collisional equilibrium concentration of H^- in its calculation of the H_2 formation rate (and only considers the H^- formation channel), whereas Prescription I solves the entire system of relevant reactions self-consistently, properly capturing non-equilibrium behaviour. The simulations with Prescription I can be used to judge how far from equilibrium both H_2 and its reactants are in subhalo gas. I find that the hydrogen ionisation fraction is dominated by photoionisation, rather than collisions, leading to a H^+ concentration far above that of collisional equilibrium. With

electrons being necessary for the formation of H^-, this also causes the H^- concentration to be far above the collisional equilibrium value, violating the assumption of Kimm et al. (2017). The concentration of H_2^+, a reactant of secondary importance, is also significantly out of equilibrium. Thus, in simulations of Halo 1 and 2, a fully non-equilibrium and radiation-coupled treatment of H_2 formation is necessary.

Chapter 5

Conclusions

The IMF of Population III stars is poorly constrained by observations, with theoretical predictions spanning a wide range of values (Greif et al., 2011; Hirano et al., 2015; Stacy et al., 2016; Hosokawa et al., 2016; Hirano & Bromm, 2017; Sharda et al., 2019). The formation of Pop. III stars, and their subsequent effect on their host galaxies is thus not fully understood. Even before the formation of the first stars (and the enrichment of the medium with metals), a rich chemistry exists in the primordial medium (Abel et al., 1997; Galli & Palla, 1998). The formation rates of the primary low-temperature coolants H_2 and HD are dictated by this complex system of reactions, which then control the formation of stars in the early universe (since gas cannot radiatively cool enough without these).

Once the first stars have formed, but before metals come to dominate the cooling, stellar radiation begins to play a role in this primordial chemistry, ionising hydrogen and helium, and dissociating molecules. This introduces the need for a model of radiation transport in cosmological simulations, which must also be coupled to the chemical network via photoreactions. While previous work has self-consistently included complete models of radiation-coupled non-equilibrium primordial chemistry (e.g. Ricotti et al. 2008; Wise et al. 2012; Pallottini et al. 2019), the importance of using such a model, as opposed to e.g a computationally cheaper and simpler semi-equilibrium model of H_2 formation (Kimm et al., 2017; Katz et al., 2017), or even neglecting radiation transport from local stars completely (Lenoble et al., 2024), has not been quantitatively justified in simulations of galaxies in the early universe.

In order to study the formation of Pop. III stars, and quantify the performance of approximate chemical models against a full treatment, I have modified the radiation hydrodynamics code Ramses-rt (Teyssier, 2002; Rosdahl et al., 2013) to solve a comprehensive primordial chemical system, self-consistently coupled to the hydrodynamics and multi-group radiation transport. The workings of this code (including the already existing numerical schemes in Ramses-rt) are described in Chapter 2. The new additions

are thoroughly validated in Chapter 3, comparing the results in idealised photoionisation/dissociation problems (inspired by the *Strömgren sphere*; Strömgren 1939), against analytic solutions and the public version of RAMSES-RT; I also compare the self-shielding prescriptions of Gnedin et al. (2009) and Nickerson et al. (2018), displaying their key features (such as the artificially increased penetration of dissociating radiation in that of Gnedin et al. 2009) in a test problem. The code is further verified in collapsing cloud tests, where I compare a 3D simulation of a cloud collapsing under its own gravity as it cools (as star forming clouds do) to a one-zone model with the same chemical treatment, inspired by Omukai (2000); near-perfect agreement between the one-zone and 3D models is found, once the density evolution is forced to be consistent between them.

I use my new code to run a suite of simulations at a physical resolution of 0.7 pc, in which I compare my code to 2 other radiative/chemical treatments from literature (Kimm et al., 2017; Dubois et al., 2021), and that of Kimm et al. (2017) with radiation transport and Pop. III star formation disabled. I find that including a model of radiation transport drastically changes how stars form. Radiative feedback acts to spread star formation out between the most massive subhaloes (also slowing down star formation in subhaloes as the number of stars in them increases), but also suppress it in the smaller ones. Accounting for the formation of H_2, but not radiative feedback, leads to the formation of large quantities of molecular gas, extremely strong cooling, and widespread star formation in almost all subhaloes.

The differences between my code and that of Kimm et al. (2017) are more subtle, but still significant. Considering the non-equilibrium concentration of H_2 and HD and their reactants leads to more clustered, bursty star-formation, and decreased total Pop. III formation. The total mass of metals and H_2 formed is similar between the two prescriptions, but the greatest difference is found in the densest, coldest gas, where my code predicts that significant H_2 fractions are only found at larger densities, explaining why some of the smaller subhaloes that hosted stars in the simulations with the physics prescription of Kimm et al. (2017) were not able to form stars with mine.

Due to the chaotic nature of the system, and the stochastic model of star-formation used in these simulations, two runs with the same set of initial conditions do not give the same result, with floating point rounding errors accumulating non-deterministically over time, and star formation only happening identically when the PRNG is seeded similarly across all processes between runs. To attempt to separate the 'signal' due to the use of a different physics prescription from this 'noise', I ran 5 more simulations of Halo 1 with the prescription of Kimm et al. (2017) (which is much less computationally expensive than mine). Though these simulations are clearly chaotic, with each of these runs noticeably

different from the others (with very different star-formation histories), they share similar 'macroscopic' characteristics, and have star formation histories with mostly similar statistical properties (burstiness, total mass of stars formed) and onset times. Hypothesis tests show that the characteristic quantities of the star formation history of the simulation with my code and chemical treatment are significantly different from those of the simulations with the treatment of Kimm et al. (2017), showing that the differences in the results from these two treatments cannot be attributed to noise alone.

Finally, I investigated the deviations from equilibrium in the chemical state of subhaloes in simulations run with my code. The semi-equilibrium treatment of H_2 formation in Kimm et al. (2017) assumes that H_2 only forms via the H^- channel, and that the concentration of H^- is as it would be in collisional equilibrium. Thus, I compared the concentration of H^- in cells in simulations with my code to the collisional equilibrium values, finding that, due to photoionisation providing abundant free electrons, there is much more H^- than expected in most cells, contradicting the assumption of Kimm et al. (2017). This leads to more rapid H_2 formation than in collisional equilibrium, and a resultant H_2 concentration far out of collisional equilibrium in most cells, despite it being destroyed by Lyman-Werner radiation from nearby stars. I may therefore conclude that a fully non-equilibrium, radiation-coupled treatment of H_2 formation is necessary to properly simulate star formation in the primordial universe.

Bibliography

Abbott T. M. C., et al., 2022a, Phys. Rev. D, 105, 023520

Abbott T. M. C., et al., 2022b, Phys. Rev. D, 105, 043512

Abel T., Anninos P., Zhang Y., Norman M. L., 1997, New A, 2, 181

Abgrall H., Le Bourlot J., Pineau Des Forets G., Roueff E., Flower D. R., Heck L., 1992, A&A, 253, 525

Agertz O., et al., 2007, Mon. Not. Roy. Astron. Soc., 380, 963

Albrecht A., Steinhardt P. J., 1982, Phys. Rev. Lett., 48, 1220

Arnett D., 1996, Supernovae and Nucleosynthesis: An Investigation of the History of Matter from the Big Bang to the Present

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

Baczynski C., Glover S. C. O., Klessen R. S., 2015, MNRAS, 454, 380

Barnes J., Hut P., 1986, Nature, 324, 446

Barry N., Bernardi G., Greig B., Kern N., Mertens F., 2022, Journal of Astronomical Telescopes, Instruments, and Systems, 8, 011007

Baur J., Palanque-Delabrouille N., Yèche C., Magneville C., Viel M., 2016, J. Cosmology Astropart. Phys., 2016, 012

Becker R. H., et al., 2001, AJ, 122, 2850

Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143

Bethe H. A., Fuchs K., Hirschfelder J. O., Magee J. L., Neumann R. v., 1958, Technical report, Blast Wave. LOS ALAMOS NATIONAL LAB NM

Bialy S., Bihr S., Beuther H., Henning T., Sternberg A., 2017, ApJ, 835, 126

Black J. H., 1981, MNRAS, 197, 553

Black J. H., Dalgarno A., 1977, ApJS, 34, 405

Blondin J. M., Wright E. B., Borkowski K. J., Reynolds S. P., 1998, ApJ, 500, 342

Böhringer H., Chon G., Fukugita M., 2017, A&A, 608, A65

Bouwens R. J., et al., 2012, ApJ, 752, L5

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Bunker A. J., et al., 2010, MNRAS, 409, 855

Bunker A. J., et al., 2023, A&A, 677, A88

Burgess A., Seaton M. J., 1960, MNRAS, 120, 121

Capitelli M., Coppola C. M., Diomede P., Longo S., 2007, A&A, 470, 811

Carniani S., et al., 2024, Nature, 633, 318

Carr B., Kohri K., Sendouda Y., Yokoyama J., 2021, Reports on Progress in Physics, 84, 116902

Cen R., 1992, ApJS, 78, 341

Chabrier G., 2005, in Corbelli E., Palla F., Zinnecker H., eds, Astrophysics and Space Science Library Vol. 327, The Initial Mass Function 50 Years Later. p. 41 (arXiv:astro-ph/0409465), doi:10.1007/978-1-4020-3407-7_5

Chandrasekhar S., 1943, Reviews of Modern Physics, 15, 1

Chen K.-J., Heger A., Woosley S., Almgren A., Whalen D. J., Johnson J. L., 2014, ApJ, 790, 162

Chevalier R. A., 1974, ApJ, 188, 501

Clowe D., Gonzalez A., Markevitch M., 2004, ApJ, 604, 596

Colella P., 1990, Journal of Computational Physics, 87, 171

Coppola C. M., Longo S., Capitelli M., Palla F., Galli D., 2011, ApJS, 193, 7

Cowie L. L., Barger A. J., Trouille L., 2009, ApJ, 692, 1476

Dalgarno A., Lepp S., 1987, in Vardya M. S., Tarafdar S. P., eds, Vol. 120, Astrochemistry. pp 109–118

Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827

Dodelson S., 2003, Modern Cosmology. Academic Press, Amsterdam

Draine B. T., Bertoldi F., 1996, ApJ, 468, 269

Driver S. P., et al., 2022, MNRAS,

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

Dubois Y., et al., 2021, A&A, 651, A109

Einstein A., 1905, Annalen der Physik, 322, 549

Eisenstein D. J., Hut P., 1998, ApJ, 498, 137

Eisenstein D. J., et al., 2023, arXiv e-prints, p. arXiv:2306.02465

Evans C. R., Hawley J. F., 1988, ApJ, 332, 659

Fan X., Narayanan V. K., Strauss M. A., White R. L., Becker R. H., Pentericci L., Rix H.-W., 2002, AJ, 123, 1247

Fan X., et al., 2006, AJ, 132, 117

Faucher-Giguère C.-A., Lidz A., Zaldarriaga M., Hernquist L., 2009, ApJ, 703, 1416

Federrath C., Klessen R. S., 2012, ApJ, 761, 156

Federrath C., Klessen R. S., Schmidt W., 2008, ApJ, 688, L79

Feldmann R., et al., 2022, arXiv e-prints, p. arXiv:2205.15325

Ferland G. J., Peterson B. M., Horne K., Welsh W. F., Nahar S. N., 1992, ApJ, 387, 95

Ferland G. J., et al., 2013, Rev. Mexicana Astron. Astrofis., 49, 137

Ferland G. J., et al., 2017, Rev. Mexicana Astron. Astrofis., 53, 385

Fontanot F., Cristiani S., Pfrommer C., Cupani G., Vanzella E., 2014, MNRAS, 438, 2097

Forrey R. C., 2013, ApJ, 773, L25

Fowler W. A., 1966, Transactions of the International Astronomical Union, Series B, 12B, 581

Fromang S., Hennebelle P., Teyssier R., 2006, A&A, 457, 371

Galli D., Palla F., 1998, A&A, 335, 403

Galli D., Palla F., 2002, Planet. Space Sci., 50, 1197

Geen S., Rosdahl J., Blaizot J., Devriendt J., Slyz A., 2015, MNRAS, 448, 3248

Genel S., et al., 2019, ApJ, 871, 21

Gingold R. A., Monaghan J. J., 1977, MNRAS, 181, 375

Glover S. C. O., Abel T., 2008, MNRAS, 388, 1627

Glover S. C. O., Federrath C., Mac Low M. M., Klessen R. S., 2010, MNRAS, 404, 2

Gnedin N. Y., Abel T., 2001, New A, 6, 437

Gnedin N. Y., Tassis K., Kravtsov A. V., 2009, ApJ, 697, 55

Godunov S., Bohachevsky I., 1959, Matematičeskij sbornik, 47, 271

Grassi T., Bovino S., Schleicher D. R. G., Prieto J., Seifried D., Simoncini E., Gianturco F. A., 2014, MNRAS, 439, 2386

Greif T. H., Springel V., White S. D. M., Glover S. C. O., Clark P. C., Smith R. J., Klessen R. S., Bromm V., 2011, ApJ, 737, 75

Grisdale K., Thatte N., Devriendt J., Pereira-Santaella M., Slyz A., Kimm T., Dubois Y., Yi S. K., 2021, MNRAS, 501, 5517

Guillet T., Teyssier R., 2011, Journal of Computational Physics, 230, 4756

Gunn J. E., Gott J. Richard I., 1972, ApJ, 176, 1

Gunn J. E., Peterson B. A., 1965, ApJ, 142, 1633

Guth A. H., 1981, Phys. Rev. D, 23, 347

Haehnelt M. G., Madau P., Kudritzki R., Haardt F., 2001, ApJ, 549, L151

Hahn O., Abel T., 2013, MUSIC: MUlti-Scale Initial Conditions, Astrophysics Source Code Library, record ascl:1311.011

Hahn O., Teyssier R., Carollo C. M., 2010, MNRAS, 405, 274

Harten A., Lax P. D., Leer B. V., 1983, SIAM Review, 25, 35

Heays, A. N. Bosman, A. D. van Dishoeck, E. F. 2017, A&A, 602, A105

Heger A., Fryer C. L., Woosley S. E., Langer N., Hartmann D. H., 2003, ApJ, 591, 288

Hernquist L., Katz N., 1989, ApJS, 70, 419

Hestenes M. R., Stiefel E., et al., 1952, Journal of research of the National Bureau of Standards, 49, 409

Heymans C., et al., 2021, A&A, 646, A140

Hindmarsh A. C., 1983, Scientific Computing, pp 55–64

Hinshaw G., et al., 2013, ApJS, 208, 19

Hirano S., Bromm V., 2017, MNRAS, 470, 898

Hirano S., Hosokawa T., Yoshida N., Omukai K., Yorke H. W., 2015, MNRAS, 448, 568

Hirt C. W., Amsden A. A., Cook J., 1974, Journal of computational physics, 14, 227

Hoag A., et al., 2019, ApJ, 878, 12

Hockney R. W., Goel S. P., Eastwood J. W., 1973, Chemical Physics Letters, 21, 589

Hollenbach D. J., 1990, in Blitz L., ed., Astronomical Society of the Pacific Conference Series Vol. 12, The Evolution of the Interstellar Medium. pp 167–181

Hollenbach D., McKee C. F., 1979, ApJS, 41, 555

Hopkins P. F., 2015, MNRAS, 450, 53

Hopkins P. F., et al., 2018, MNRAS, 480, 800

Hosokawa T., Hirano S., Kuiper R., Yorke H. W., Omukai K., Yoshida N., 2016, ApJ, 824, 119

Hubble E., 1929a, Proceedings of the National Academy of Science, 15, 168

Hubble E. P., 1929b, ApJ, 69, 103

Huebner W. F., Mukherjee J., 2015, Planetary and Space Science, 106, 11

Hui L., Gnedin N. Y., 1997, MNRAS, 292, 27

Iliev I. T., et al., 2006, MNRAS, 371, 1057

Iliev I. T., et al., 2009, MNRAS, 400, 1283

Janev R. K., Langer W. D., Evans K., 1987, Elementary processes in Hydrogen-Helium plasmas - Cross sections and reaction rate coefficients

Jebsen J. T., 1921, Arkiv for Matematik, Astronomi och Fysik, 15, 18

Jiang L., et al., 2022, arXiv e-prints, p. arXiv:2206.07825

Jones E., Oliphant T., Peterson P., et al., 2001, SciPy: Open source scientific tools for Python, `http://www.scipy.org/`

Kahn F. D., 1954, Bull. Astron. Inst. Netherlands, 12, 187

Karpas Z., Anicich V., Huntress W. T., 1979, J. Chem. Phys., 70, 2877

Katz N., 1992, ApJ, 391, 502

Katz N., Weinberg D. H., Hernquist L., 1996, ApJS, 105, 19

Katz H., Kimm T., Sijacki D., Haehnelt M. G., 2017, MNRAS, 468, 4831

Katz H., et al., 2022, MNRAS, 510, 5603

Katz H., Kimm T., Ellis R. S., Devriendt J., Slyz A., 2023, Monthly Notices of the Royal Astronomical Society, 524, 351

Keller B. W., Wadsley J. W., Wang L., Kruijssen J. M. D., 2019, MNRAS, 482, 2244

Kim C.-G., Ostriker E. C., 2015, ApJ, 802, 99

Kimm T., Cen R., 2014, ApJ, 788, 121

Kimm T., Katz H., Haehnelt M., Rosdahl J., Devriendt J., Slyz A., 2017, MNRAS, 466, 4826

Komiya Y., Suda T., Fujimoto M. Y., 2016, ApJ, 820, 59

Koopmans L., et al., 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 1 (arXiv:1505.07568), doi:10.22323/1.215.0001

Kreckel H., Bruhns H., Čížek M., Glover S. C. O., Miller K. A., Urbain X., Savin D. W., 2010, Science, 329, 69

Kroupa P., 2001, MNRAS, 322, 231

Kutta M., 1901, Zeitschrift für Mathematik und Physik, 46, 435

Lacey C., Cole S., 1993, MNRAS, 262, 627

Latif M. A., Whalen D. J., Khochfar S., Herrington N. P., Woods T. E., 2022, Nature, 607, 48

Lazar A., Bromm V., 2022, MNRAS, 511, 2505

Leitherer C., et al., 1999, ApJS, 123, 3

Lenoble R., Commerçon B., Rosdahl J., 2024, A&A, 685, A7

Levermore C. D., 1984, J. Quant. Spec. Radiat. Transf., 31, 149

Linde A. D., 1982, Physics Letters B, 108, 389

Lipovka A., Núñez-López R., Avila-Reese V., 2005, MNRAS, 361, 850

Lloyd S., 1982, IEEE Transactions on Information Theory, 28, 129

Lotz W., 1967, ApJS, 14, 207

Lynden-Bell D., 1967, MNRAS, 136, 101

Markevitch M., Gonzalez A. H., David L., Vikhlinin A., Murray S., Forman W., Jones C., Tucker W., 2002, ApJ, 567, L27

Markevitch M., Gonzalez A. H., Clowe D., Vikhlinin A., Forman W., Jones C., Murray S., Tucker W., 2004, ApJ, 606, 819

Martel H., Shapiro P. R., 1998, MNRAS, 297, 467

Martin P. G., Schwarz D. H., Mandy M. E., 1996, ApJ, 461, 265

McElroy D., Walsh C., Markwick A. J., Cordiner M. A., Smith K., Millar T. J., 2013, A&A, 550, A36

Mihalas D., Weibel-Mihalas B., 1984, Foundations of radiation hydrodynamics

Milgrom M., 1983, ApJ, 270, 365

Misner C. W., Thorne K. S., Wheeler J. A., 1973, Gravitation. W. H. Freeman, San Francisco

Mitchell G. F., Deveau T. J., 1983, ApJ, 266, 646

Mitchell P. D., Blaizot J., Devriendt J., Kimm T., Michel-Dansac L., Rosdahl J., Slyz A., 2018, MNRAS, 474, 4279

Monaghan J. J., Gingold R. A., 1983, Journal of Computational Physics, 52, 374

Monsalve R. A., Rogers A. E. E., Bowman J. D., Mozdzen T. J., 2017, ApJ, 847, 64

Mori M., Yoshii Y., Tsujimoto T., Nomoto K., 1997, ApJ, 478, L21

Navarro J. F., White S. D. M., 1993, MNRAS, 265, 271

Nelson D., et al., 2019, Computational Astrophysics and Cosmology, 6, 2

Nickerson S., Teyssier R., Rosdahl J., 2018, MNRAS, 479, 3206

Oesch P. A., et al., 2016, ApJ, 819, 129

Okabe N., Takada M., Umetsu K., Futamase T., Smith G. P., 2010a, PASJ, 62, 811

Okabe N., Okura Y., Futamase T., 2010b, ApJ, 713, 291

Omukai K., 2000, ApJ, 534, 809

Omukai K., Tsuribe T., Schneider R., Ferrara A., 2005, ApJ, 626, 627

Osterbrock D. E., Ferland G. J., 2006, Astrophysics of gaseous nebulae and active galactic nuclei

Ostriker J. P., Peebles P. J. E., 1973, ApJ, 186, 467

Ouchi M., et al., 2010, ApJ, 723, 869

Padoan P., Nordlund Å., 2011, ApJ, 730, 40

Padoan P., Nordlund A., Jones B. J. T., 1997, MNRAS, 288, 145

Pakmor R., Bauer A., Springel V., 2011, MNRAS, 418, 1392

Pallottini A., et al., 2019, MNRAS, 487, 1689

Pawlik A. H., Schaye J., 2008, MNRAS, 389, 651

Planck Collaboration et al., 2016, A&A, 594, A14

Planck Collaboration et al., 2020, A&A, 641, A1

Poulaert G., Brouillard F., Claeys W., McGowan J. W., Van Wassenhove G., 1978, Journal of Physics B Atomic Molecular Physics, 11, L671

Power C., Navarro J. F., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T., 2003, MNRAS, 338, 14

Press W. H., Schechter P., 1974, ApJ, 187, 425

Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical recipes in C. The art of scientific computing

Péroux C., et al., 2019, Monthly Notices of the Royal Astronomical Society, 485, 1595

Raga A. C., Cantó J., Rodríguez L. F., 2012, MNRAS, 419, L39

Richard J., et al., 2010, MNRAS, 404, 325

Richings A. J., Schaye J., 2016, MNRAS, 458, 270

Ricotti M., Gnedin N. Y., Shull J. M., 2008, ApJ, 685, 21

Riemer-Sørensen S., Jenssen E. S., 2017, Universe, 3, 44

Riess A. G., et al., 1998, AJ, 116, 1009

Ripamonti E., Abel T., 2004, MNRAS, 348, 1019

Roe P. L., 1997, Journal of Computational Physics, 135, 250

Rosdahl J., Blaizot J., 2012, MNRAS, 423, 344

Rosdahl J., Blaizot J., Aubert D., Stranex T., Teyssier R., 2013, MNRAS, 436, 2188

Rosdahl J., et al., 2018, MNRAS, 479, 994

Rubin V. C., Ford W. K. J., Thonnard N., 1980, ApJ, 238, 471

Runge C., 1895, Mathematische Annalen, 46, 167

Safranek-Shrader C., Krumholz M. R., Kim C.-G., Ostriker E. C., Klein R. I., Li S., McKee C. F., Stone J. M., 2017, MNRAS, 465, 885

Saltzman J., 1994, Journal of computational physics, 115, 153

Savin D. W., Krstić P. S., Haiman Z., Stancil P. C., 2004, ApJ, 606, L167

Scannapieco C., et al., 2012, MNRAS, 423, 1726

Schaerer D., 2002, A&A, 382, 28

Schaye J., et al., 2015, MNRAS, 446, 521

Schmidt M., 1959, ApJ, 129, 243

Sedov L. I., 1946, Journal of Applied Mathematics and Mechanics, 10, 241

Shapiro P. R., Iliev I. T., Alvarez M. A., Scannapieco E., 2006, ApJ, 648, 922

Sharda P., Krumholz M. R., Federrath C., 2019, MNRAS, 490, 513

Silk J., 1968, ApJ, 151, 459

Singh S., et al., 2022, Nature Astronomy, 6, 607

Spitzer L., 1978, Physical processes in the interstellar medium, doi:10.1002/9783527617722.

Springel V., 2005, MNRAS, 364, 1105

Springel V., 2010, MNRAS, 401, 791

Springel V., Yoshida N., White S. D. M., 2001, New A, 6, 79

Stacy A., Bromm V., Lee A. T., 2016, MNRAS, 462, 1307

Stancil P. C., Lepp S., Dalgarno A., 1998, ApJ, 509, 1

Stecher T. P., Williams D. A., 1967, ApJ, 149, L29

Steinhardt C. L., Jespersen C. K., Linzer N. B., 2021, ApJ, 923, 8

Stenrup M., Larson Å., Elander N., 2009, Phys. Rev. A, 79, 012713

Stinson G., Seth A., Katz N., Wadsley J., Governato F., Quinn T., 2006, MNRAS, 373, 1074

Strömgren B., 1939, ApJ, 89, 526

Sutherland R. S., Dopita M. A., 1993, ApJS, 88, 253

Taylor G., 1950, Proceedings of the Royal Society of London Series A, 201, 159

Teyssier R., 2002, A&A, 385, 337

Teyssier R., Fromang S., Dormy E., 2006, Journal of Computational Physics, 218, 44

Thornton K., Gaudlitz M., Janka H. T., Steinmetz M., 1998, ApJ, 500, 95

Tornatore L., Ferrara A., Schneider R., 2007, MNRAS, 382, 945

Toro E., 2009, Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction. Springer Berlin Heidelberg

Toro E. F., Spruce M., Speares W., 1994, Shock Waves, 4, 25

Tweed D., Devriendt J., Blaizot J., Colombi S., Slyz A., 2009, A&A, 506, 647

Vazza F., Dolag K., Ryu D., Brunetti G., Gheller C., Kang H., Pfrommer C., 2011, MNRAS, 418, 960

Von Neumann J., Richtmyer R. D., 1950, Journal of Applied Physics, 21, 232

Weinberger R., Springel V., Pakmor R., 2020, ApJS, 248, 32

Wikipedia contributors 2023, Octree — Wikipedia, The Free Encyclopedia, `https://en.wikipedia.org/wiki/Octree`

Williams C. C., et al., 2018, ApJS, 236, 33

Wise J. H., Turk M. J., Norman M. L., Abel T., 2012, ApJ, 745, 50

Wise J. H., Demchenko V. G., Halicek M. T., Norman M. L., Turk M. J., Abel T., Smith B. D., 2014, MNRAS, 442, 2560

Wolfire M. G., Tielens A. G. G. M., Hollenbach D., Kaufman M. J., 2008, ApJ, 680, 384

Woods T. E., Willott C. J., Regan J. A., Wise J. H., Downes T. P., Norman M. L., O'Shea B. W., 2021, ApJ, 920, L22

Yee H. C., Warming R. F., Harten A., 1985, NASA STI/Recon Technical Report A, 85, 357

Yoshida N., Omukai K., Hernquist L., Abel T., 2006, ApJ, 652, 6

Zackrisson E., Rydberg C.-E., Schaerer D., Östlin G., Tuli M., 2011, ApJ, 740, 13

Zeldovich Y., Raizer Y. P., 1967, Physics of Shock Waves and High Temperature Hydrodynamic Phenomen. Academic Press, doi:10.1016/B978-0-12-395672-9.X5001-2

Zwicky F., 1933, Helvetica Physica Acta, 6, 110

de Blok W. J. G., Walter F., Brinks E., Trachternach C., Oh S. H., Kennicutt R. C. J., 2008, AJ, 136, 2648

de Souza R. S., Krone-Martins A., Ishida E. E. O., Ciardi B., 2012, A&A, 545, A102

van Leer B., 1979, Journal of Computational Physics, 32, 101

von Neumann J., Richtmyer R., 1947, Los Alamos Rept. Series A, LA-657