4

The simulated galaxy luminosity function: input physics, dust attenuation and galaxy selection

Abstract

We investigate the luminosity function (LF) resulting from cosmological hydrodynamical simulations with varying input physics, with and without an estimate for dust attenuation. We find that in simulations in which the supernova (SN) feedback is inefficient in massive galaxies, due to too low a wind velocity, a 'bump' in the luminosity function appears due to the overproduction of luminous galaxies. Invoking efficient feedback in these massive galaxies (either through the use of a momentum-driven wind prescription in which the energy in the winds increases with galaxy mass, a top-heavy IMF for star formation at high pressure or AGN feedback) results in a monotonically decreasing LF. Dust attenuation, implemented by assuming that the optical depth scales with the metallicity-weighted column density, is more efficient in galaxies with less efficient feedback, as there is more (high metallicity) gas available in such galaxies. With efficient feedback, little gas is left in the galaxies, reducing the effect of attenuation to close to zero. In low luminosity galaxies the column densities and optical depth are in general lower. From virtual observations we find that the LF as obtained using techniques used for observations results in LFs very similar to those obtained directly from halo catalogues. Nevertheless, for large PSFs (corresponding to typical ground-based seeing conditions) very deep observations may result in shallower faint-end slopes of the LF, due to the preferential removal of low-surface brightness galaxies.

4.1 Introduction

The luminosity function (LF) has proven to be a very powerful tool in studies of galaxy formation and evolution. Observationally, the luminosities of galaxies in some wavelength band are relatively straightforward to obtain provided reasonable estimates of the galaxies' distances can be made (usually from spectroscopic or photometric redshifts) and ignoring dust attenuation. Galaxy formation models, combined with stellar population synthesis models, can predict the luminosities of galaxies. Here, several uncertainties come in to play. For example, the star formation histories and metallicities of model galaxies may not be representative of real galaxies. Population synthesis models also do not come without uncertainty, because the contribution of exotic kinds of stars (e.g. thermally pulsing asymptotic giant branch stars, TP-AGBs) may dominate the spectral energy distribution (SED) of a simple stellar population (in this example, the near-infrared emission) and the inclusion of these exotic stars is non-trivial (e.g. Maraston, 2005; Tonini et al., 2010). The initial mass function of stars is also crucial to determine the SED of galaxies and is not very well constrained, especially not at high redshift. Further issues, specific to simulations are how to model dust extinction without the presence of a multi-phase interstellar medium (ISM) and the (spatial and mass) resolution of the simulations used (see e.g. Jonsson, 2006; Wuyts et al., 2009).

Going the other way, i.e. converting the observed luminosities of galaxies into physical properties through SED modelling, comes at least with the same uncertainties. Added to those are the fact that many different combinations of physical properties are degenerate in their contributions to the SED of a galaxy. As an example, a population may become redder due to ageing, having more metals and/or more dust extinction. In order to break the degeneracies between models a very large range of wavelengths is generally necessary, ideally ranging al the way from the ultraviolet to the far-infrared (or even sub-millimetre). Even then, some problems still exist. Noise in the data can allow different solutions to the results of the SED modelling. Extinction will probably not be uniform across a galaxies, whereas most SED modelling attempts take one single value for the attenuation of a galaxy. For a recent review of SED modelling techniques, see Walcher et al. (2010). Very strong extinction will even result in the removal of a galaxy from the sample, an effect for which correction is nearly impossible.

Semi-analytic recipes for the evolution of the baryonic component of galaxies on top of some underlying dark matter halo merger history, are usually tuned to reproduce the redshift zero LF in one or more broadband filters (e.g. Cole et al., 2000; Croton et al., 2006; De Lucia & Blaizot, 2007; Monaco et al., 2007; Bower et al., 2008). Predictions for higher redshift LFs, or other (physical or observable) properties of the galaxy population, can then be made and tested in order to vali-

date the models. Hydrodynamical cosmological simulations, on the other hand, are usually not tuned to reproduce any LF. The LF is therefore a prediction of the models, but often already ruled out by observations. It is, nevertheless, still interesting to investigate the LF from hydrodynamical cosmological simulations. Especially when varying input physics of the simulations, as is done in this paper, as this makes it possible to investigate what physical ingredients shape the LF of a galaxy population. Previous studies have already focused on the photometric properties of simulated galaxies in various wavelength bands. For example, Nagamine (2002); Nagamine et al. (2004); Night et al. (2006) focus on the photometry of Lyman Break Galaxies at various redshifts and find that many simulations predict a LF with a steeper faint end slope than observed, for moderate assumptions about the extinction. Using their package 'SUNRISE', Jonsson et al. (2010) investigate the photometric properties of simulated mergers and from mock light cones constructed from the Millennium Simulation (Springel et al., 2005) with semi-analytic models, Kitzbichler & White (2007) and Stringer et al. (2009) compare the observed optical properties of a large range of galaxies at $0 < z < 5$.

In cosmological simulations, much of the physics that is very important for galaxy formation is not resolved, i.e. important processes take place on (mass, length and time) scales below the resolution limit of the simulations. For this reason, these simulations rely on sub-grid models that describe the effects of the small scale processes on the scale of their resolution (currently typically $\sim 10^6\ M_\odot$ and/or $\sim$ 1 kpc). There is considerable freedom in the implementation of these sub-grid models and in the values of their parameters. Many possible choices are well motivated, but result in substantially different galaxy populations.

In the current paper we use a small sub-set of the OverWhelmingly Large Simulations (Schaye et al., 2010), in order to investigate the LFs of simulated galaxies. Many of the physical parameters, including the stellar mass function, have already been discussed in Chapter 2. Here we implement the population synthesis models of Bruzual & Charlot (2003) in order to estimate the luminosities of the galaxies in several filters. As already alluded to, the LFs cannot be expected to give as good a match to observations as achieved by semi-analytic models. A major draw-back is also the cosmology used in these simulations, which match the WMAP 3-year results (Spergel et al., 2007), and have a lower amplitude of fluctuations than the currently favoured WMAP-7 year results (Komatsu et al., 2010). A more fundamental prediction of the simulations than the LF is the distribution of luminosities of galaxies as a function halo mass. This relation can then be convolved with a halo mass function from the favoured cosmology in order to predict the LF. We postpone such an analysis to future work and will focus on the LF as directly obtained from the simulations' halo catalogues.

For these reasons, the goal of this paper is not to match the observed galaxy LF

as well as we can, but rather to

1. investigate how the input physics of the *OWLS* simulations affects the shape of the LF.

2. investigate how we can model dust attenuation in simulations with a resolution typical of current hydrodynamical cosmological simulations.

3. compare the LFs constructed from virtual observations with those from (sub-) halo catalogues.

Observers identify galaxies by selecting regions of pixels, whose surface brightness exceeds the background by some specified threshold. Simulators, on the other hand, usually identify gravitationally bound groups of particles, and call the centres of these structures (containing the stars and cold gas) galaxies. These methods are so fundamentally different, that it is not at all guaranteed that they would yield the same LF, even if they would both have the exact same population of galaxies. We will therefore project our star particles onto images, convolve them with a reasonable point spread function, add noise, and consequently measure the LF with the tools observers would use for this (in this Chapter we use SExtractor, Bertin & Arnouts, 1996). We will show that these two methods yield very similar LFs, which is encouraging for studies comparing modelled and observed galaxies.

The structure of this chapter is laid out as follows. In Section 4.2 we describe the simulations used in this chapter and the physics therein. Section 4.3 describes how we deal with the population synthesis and we show the convergence and evolution of the LFs, followed by a description of how different input physics influence the shape of the LF. Dust attenuation and its effect on the LF are discussed in Section 4.4. The creation of virtual observations and the LFs resulting from the procedure observers would follow are described in Section 4.5 and we conclude in Section 4.6.

4.2 Simulations

For a detailed discussion of the full set of *OWLS* runs we refer the reader to Schaye et al. (2010). Here we will briefly summarize the set of simulations, their relevant numerical properties and the sub-grid models under consideration.

4.2.1 OverWhelmingly Large Simulations

The simulations are performed with an extended version of the N-Body Tree/smooth particle hydrodynamics (SPH) code GADGET3 (an improved version

of GADGET2, last described in Springel, 2005) in periodic boxes of 25 and 100 comoving Mpc h^{-1}. There are 512^3 dark matter and equally many baryonic particles (which can be either collisionless 'stars' or collisional 'gas' particles). The particle mass of the highest resolution simulation under consideration (25 h^{-1}Mpc box size, 2×512^3 particles) is $8.68 \times 10^6\ M_\odot$ for dark matter and initially $1.85 \times 10^6\ M_\odot$ for baryons (the baryonic particle masses change during the course of the simulation due to mass transfer from star to gas particles).

Initial conditions are generated with CMBFAST (Seljak & Zaldarriaga, 1996) and evolved forward in time from an initial glass-like state using the Zel'Dovich (1970) approximation to $z = 127$, where the simulation is started. The cosmology assumed is specified by $\Omega_\mathrm{m} = 0.238$, $\Omega_\mathrm{b} = 0.0418$, $\Omega_\Lambda = 0.762$, $\sigma_8 = 0.74$, $n = 0.951$ and $h = H_0$ / (100 km s^{-1}/Mpc) = 0.73. These values were deduced from the WMAP 3-year results (Spergel et al., 2007) and are largely consistent with the more recent WMAP7 results (Komatsu et al., 2010). The most significant discrepancy is in σ_8, which is 8% lower in WMAP3 than in WMAP7 (resulting in slightly delayed structure formation in the WMAP3 cosmology).

The names of the simulations are as follows: '*NAME_LxxxNyyy*', in which '*NAME*' is a very short description about which parameters are changed (always specified in the text), '*Lxxx*' is the box size, in which $L = \{100, 050, 025\}$, corresponding to 100, 50 and 25 comoving h^{-1}Mpc and '*N*' denotes the number of particles, such that $N = \{512, 256, 128\}$ corresponds to 2 times 512^3, 256^3 and 128^3 particles, respectively. As an example, the reference model in a 25 h^{-1}Mpc box with 2 times 512^3 particles will be denoted by '*REF_L025N512*'.

4.2.2 Subgrid physics in the reference model

Radiative cooling and heating are treated by explicitly following 11 elements in photo-ionization equilibrium with the CMB and a Haardt & Madau (2001) model for the UV/X-ray background radiation from quasars and galaxies, as described in Wiersma et al. (2009a). At some density, deep inside haloes, we know that the gas is composed of several phases, ranging from hot/warm tenuous gas to cold, dense molecular clouds. This high density, multi-phase ISM is not resolved and particles with proper physical hydrogen number densities $n_\mathrm{H} > 10^{-1}$ cm^{-3} and temperatures $T < 10^5$K are put on a polytropic effective equation of state (EoS), in which the pressure $P \propto \rho^{\gamma_\mathrm{eff}}$, where $\gamma_\mathrm{eff} = 4/3$ is the polytropic index (this value is chosen, such that both the Jeans mass and the ratio of the Jeans length and the SPH kernel are independent of the density, thus preventing spurious fragmentation due to a lack of numerical resolution, see Schaye & Dalla Vecchia 2008), and ρ is the mass density of the gas. The normalization of the polytropic equation of state is such that the energy per unit mass corresponds to 10^4 K at a mean molecular

weight of 1.2. Star formation is followed stochastically, with a pressure dependent star formation rate, obtained from the observed Kennicutt-Schmidt law (Kennicutt, 1998a) and local hydrostatic equilibrium, as discussed in Schaye & Dalla Vecchia (2008). Gas particles are only allowed to form stars when they are on the EoS, so there is a threshold density for star formation of $n_{\rm H} > 10^{-1}$ cm^{-3}.

The star particles are assumed to be simple stellar populations (SSPs) with a Chabrier (2003) IMF. Stellar feedback from massive stars and supernovae is implemented kinetically, which means that we launch a wind with wind velocity v_w = 600 km s^{-1}, in which the mass loading is such that the energy in the wind corresponds to about 40% of the energy available from supernovae of type II (including Ib,c), which for our IMF means that the mass loading in the wind $\eta = 2\times$SFR. For details on the kinetic wind implementation, see Dalla Vecchia & Schaye (2008). The enrichment of the gas by AGB stars, Type Ia and Type II (including Type Ib,c) supernovae is followed explicitly for the 11 elements needed for the cooling as described in Wiersma et al. (2009b).

4.2.3 Variation of subgrid models

In this chapter, we will make use of three variations to the reference model described above. One, which we will call 'No SN/No Z cooling' includes neither supernova feedback nor metal-line cooling. By comparing this model to the reference simulation we can investigate the influence of SN feedback on the LF, with and without dust attenuation. We turned off metal-line cooling because the metallicity of dense gas becomes unrealistically high in the absence of SN feedback. Note that there is still gas cooling through hydrogen and helium. The OWLS name of the simulation is 'NOSN_NOZCOOL'.

The SN feedback as implemented in the reference model becomes inefficient for high-mass galaxies (Chapter 2), because the pressure in these galaxies is sufficiently high to prevent the wind from escaping. The newly enriched gas surrounding young stellar particles stays where it is and the high metal content and high densities result in effective cooling and star formation. In order to have a simulation with effective feedback from star formation for a wider range of galaxy masses, we implement a model that has a top-heavy IMF if the star particle forms at a pressure $P/k > 2.0 \times 10^6$ cm^{-3}K (evaluated at the resolution limit of the simulation). The IMF used is a top heavy IMF with $\mathrm{d}N/\mathrm{d}M \propto M^{-1}$ (in these units, Salpeter would have an index of -2.35). The excess energy corresponding to the higher fraction of high mass stars per unit stellar mass formed is used to increase the initial wind velocity from 600 km s^{-1} to 1618 km s^{-1}. This model is called 'Top-heavy IMF', while its OWLS name is 'DBLIMFCONTSFV1618'.

In order to investigate the effect of a varying wind mass loading (and to get a

higher mass loading in low mass galaxies) we use a wind model in which the mass loading and the wind velocity depend on the circular velocity ($v_c = \sqrt{GM_{\rm vir}/R_{\rm vir}}$) of the halo (determined using an on-the-fly group finder during the simulation) they were launched from, as follows: $v_w = 5v_c/\sqrt{2}$ and $\eta = \frac{1}{\sqrt{2}} \times (v_c/150{\rm km\ s^{-1}})^{-1}$. Note that the energy in these wind models is not constant and increases with galaxy mass (and exceed the total available energy from SNe for the most massive galaxies). In low mass galaxies the mass loading is higher than in high mass galaxies, but the velocity is lower. These models are motivated by a wind driving mechanism in which the winds are accelerated by radiation pressure from the stellar population on dust grains (Murray et al., 2005). We will here call this model 'Momentum driven winds' and its OWLS name is 'WVCIRC'. This model is very similar to the model of Oppenheimer & Davé (2006, 2008)

The last simulation we will discuss here includes AGN feedback. The prescription for the growth of the black holes and the corresponding feedback are described in Booth & Schaye (2009). It is the OWLS simulation 'AGN'. This model produces black holes that follow the observed scaling relations (Booth & Schaye, 2009) and effectively suppress star formation in massive haloes (Chapter 2).

4.2.4 Halo identification

Haloes are identified using a Friends-of-Friends (FoF) algorithm, linking together all dark matter particles which are closer to each other than the linking parameter ($b = 0.2$ times the mean inter particle distance). Baryonic particles are linked to their nearest dark matter particle and belong to the same group, if any. FoF identifies iso-overdensity contours of $\delta \simeq 3/(2\pi b^3) \simeq 60$ (Lacey & Cole, 1994).

Within these haloes, gravitationally bound substructures are identified using the SubFind algorithm (Springel et al., 2001). SubFind starts with the output of FoF and removes any unbound particles. Bound substructures are separated from the main halo and classified as subhaloes. The separation of the subhalo and the main halo occurs at saddle points in the density distribution. All particles inside the subhalo are removed from the main halo, so the mass of the main halo decreases whenever a subhalo is identified. Each subhalo (above some resolution limit, see Sect. 4.3.1) is considered to be a galaxy.

4.3 Population synthesis

We obtain observables (such as magnitudes and colours) for our galaxies using the technique commonly known as population synthesis. The idea behind this technique is to obtain a spectrum of the galaxy by summing up all the spectra of its

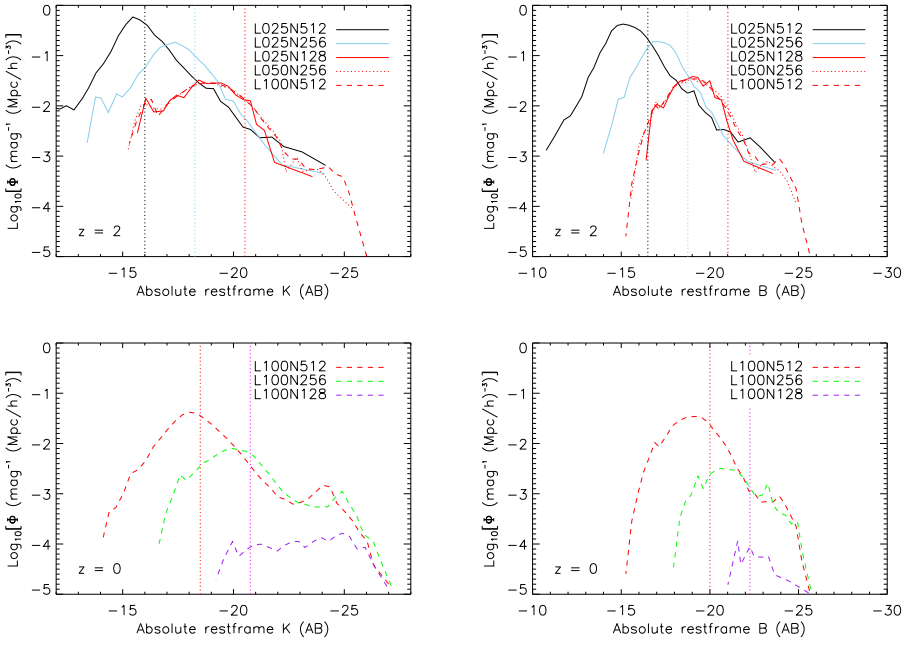

Figure 4.1: In the left-hand panels we show the rest-frame *K*-band LF of all subhaloes in the reference *OWLS* runs at redshift 2 (upper panel) and 0 (lower panel). Lines with the same line style have the same box size (and different resolutions), while lines with the same colour, but different style have different box sizes at the same resolution. In the right-hand panels we show the same, but now for the *B*-band. The vertical dotted lines show the adopted resolution limits in the simulations with the resolution corresponding to the same colour. We conclude that the LF is well converged with respect to resolution and that the box size is only important for sampling the high luminosity end of the LF.

stars. To this end, we assume every star particle to be an SSP, i.e. a stellar population with a single age and composition. The spectrum of this SSP then depends on its age, metallicity, mass and the IMF. The spectrum is convolved with a filter profile to obtain (broad-band) magnitude in any waveband desired. As input SSP spectra we use the package of Bruzual & Charlot (2003, BC03 from now on), with a Chabrier (2003) IMF, consistent with the IMF used in the simulations. The mass and metallicity are simply the initial mass and metallicity of the star particle (inherited from its parent gas particle) and the age the time since its creation.

We interpolated the BC03 spectra to a regular grid of 1000 ages and 20 metallicities, bracketed by the lowest and highest age and metallicity available in the BC03 package (so extrapolation was necessary). This step makes the assignment of spectra to star particles computationally more efficient. For every wavelength bin we used a 2-dimensional cubic spline interpolation to interpolate in $\log_{10}$(age) and $\log_{10}(Z)$, where Z is the metallicity. To obtain magnitudes we use the BC03 filter integration algorithms. This procedure gives a very smooth interpolation between the SSP magnitudes given by the standard BC03 software.

We assign to every star particle a magnitude (or spectrum) from this 20x1000 table by taking the nearest age and metallicity combination available scaling it to the appropriate mass (the initial mass of the star particle). We do not use any other spectra than SSP spectra, as the formation of a star particle is an event that is a delta function in time. In marginally resolved galaxies this means that, because of the stochastic nature of the star formation in these simulations, the age distribution of the stellar content is very spiky. In higher mass galaxies, where there are thousands ($M_{\rm star} \sim 10^9\ M_\odot$) to millions of star particles this stochasticity is washed out.

We will here first show LFs without correcting for attenuation. We will focus on rest-frame K-band absolute luminosities, because attenuation should be relatively unimportant in the $K-$band.

4.3.1 Convergence of the LF

The upper left panel of Fig. 4.1 shows the LF at redshift two in the $K-$band for 5 different simulations, all with the same physics, but using different box sizes (at fixed resolution, all red lines) and different resolutions (at fixed box size, all solid lines). By comparing the solid lines to one another one can see that with respect to numerical resolution, our LFs are reasonably well converged over a large range of luminosities. The size of the box is only important for the high-luminosity end: we sample the LF to higher luminosities in larger boxes, as expected. At the low-luminosity end there is a down-turn of the LF, which is expected to be due to a lack of resolution in the lowest mass systems. The vertical dotted lines show the resolution limits we adopt for simulations with a resolution of the corresponding colour.

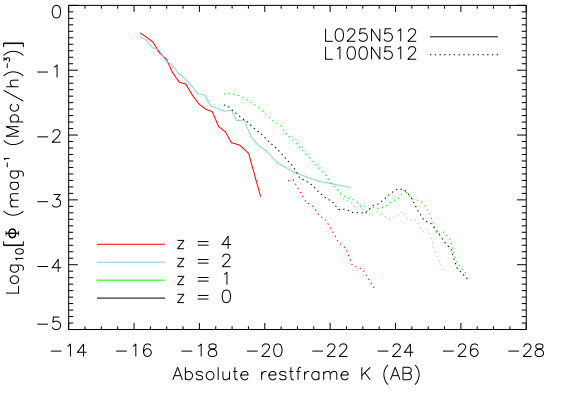

Figure 4.2: The evolution of the LF as a function of redshift for high resolution (solid lines, the *L025N512* simulation) and lower resolution (dotted lines, *L100N512* simulation, which has 64 times lower mass resolution). Different colours indicate different redshifts. The LFs are only shown for the magnitude range we consider converged.

The difference between the vertical lines is $\Delta K = -2.5 \log_{10} 8$, as expected if the resolution limit in luminosity equals the ratio in mass resolution of the simulations. The adopted resolution limit for simulations in a 25 h^{-1}Mpc box with 2×512^3 is $K = -16$.

For simulations with different input physics we verified that the same resolution limit holds at all redshifts we present in this paper. In the upper right panel we show the same, but for the *B*-band. In the *B*-band the adopted resolution limit is $B = -16.5$ for the same resolution simulations.

At redshift zero the resolution can be slightly relaxed (i.e. at the same resolution, the LF is converged down to fainter luminosities), as shown in the lower panels of Fig. 4.1. The adopted resolution limits in the at $z = 0$ are $K = -18.5$ and $B = -20$, respectively, for the 100 h^{-1}Mpc boxes with 2×512^3 particles. The very lowest resolution shown is not well converged at any luminosity. From now on we will only show LFs on the converged luminosity range.

4.3.2 Evolution of the luminosity function

For the reference simulation we show in Fig. 4.2 the evolution of the LF from $z = 4 - 0$, for both the low and the high resolution simulations (the high resolution only for $z = 4 - 2$ as the simulation stops at $z = 2$).

The number densities of galaxies of all $K-$band luminosities grow with time,

and the growth is stronger for higher luminosities. At redshifts lower than two we observe some peculiar behaviour in the low-resolution simulation. Towards $z = 0$ a bump appears in the LF. This feature has also been observed by Oppenheimer et al. (2010) in the stellar mass function at comparable resolution for constant wind models like the simulations presented in Fig. 4.2.

The appearance of the bump is closely related to the feedback prescription used. As discussed in Chapter 2, winds may or may not escape the galaxy and/or host halo depending on the prescription used and the mass of the halo. The winds need to be launched with a sufficiently high velocity in order for the gas to overcome hydrodynamical drag forces in the ISM. The required velocity increases with the pressure of the ISM and thus with the mass of the galaxy (see Dalla Vecchia & Schaye, 2008). If the wind velocities are too low, metal-enriched gas piles up, the star formation rate increases and too many stars are formed. This is visible in Fig. 4.3 as a relatively sharp upturn in the number density of galaxies For $K < -23$.

4.3.3 The effect of feedback from star formation and AGN

Simulations that use wind prescriptions which are efficient in removing gas at all masses do not predict a bump in the LF, as shown in Fig. 4.3. The bump is absent in both the model with a top-heavy IMF for starbursts and for the model that includes AGN feedback, which confirms that it was indeed the result of inefficient feedback in massive galaxies. AGN feedback is most efficient in suppressing star formation in massive galaxies, as was also shown in Chapter 2. At the low-luminosity end the LFs predicted by the the models that include SN feedback, but not the momentum driven winds, converge. This is because in lower mass galaxies the pressure in the ISM is low enough to form the majority of the stars with the default IMF, such that the SN feedback in all these simulations is the same, and because AGN are not active. The momentum driven winds result in a considerably shallower faint-end of the LF. Because the galaxies are of low mass, all wind velocities used are sufficient for the winds to escape the galaxies. In the lower mass galaxies, the wind mass loadings are higher, removing a larger part of the ISM in these galaxies.

So far, we have only looked at the stellar luminosities, which can be compared to attenuation insensitive bands, like the $K-$band, or to extinction corrected data. Because extinction corrections come with large uncertainties, and does not help for galaxies that were removed from the sample by extinction, it is also useful to try and go the other way around: estimate the extinction that would arise from the distribution of gas (and especially metals) in the simulation and compare directly to observed LFs. We will do this in the next section.

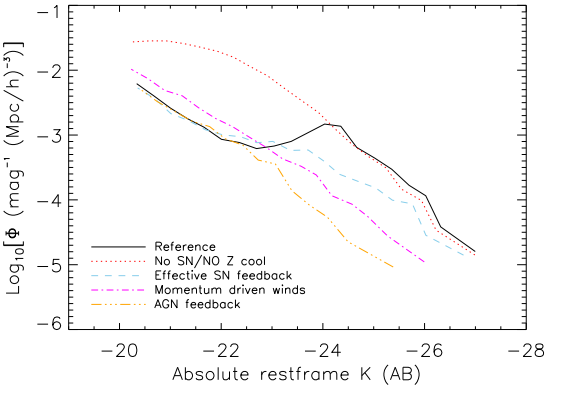

Figure 4.3: The K-band luminosity function at redshift zero for the reference (black solid line), no SN feedback and no metal-line cooling, top-heavy IMF for starbursts (red dotted line), momentum driven wind (magenta dot-dashed line) and AGN feedback (blue dashed line) models. The bump, which is very obvious in the reference simulation, is hardly present in the effective feedback models and absent in the AGN feedback run.

4.4 Dust attenuation

On its way outward through the ISM of a galaxy, star light encounters gas and dust which can absorb or scatter photons. Dependent on the wavelength under consideration, an estimate for the amount of dust attenuation is crucial to get the luminosity (or colour, as dust changes the colour by the wavelength dependency of the scattering and absorption) of a galaxy is crucial.

The clouds in the ISM in the solar neighbourhood that are responsible for the bulk of the attenuation are not resolved by our simulations. We will therefore have to rely on some parametrization (or sub-grid model) for the dust attenuation. Our approach is to model the dust attenuation on a star particle by star particle basis. This allows for gradients of attenuation throughout a galaxy, due to varying gas densities and metallicities. In Appendix 4.6 we explain the procedure of obtaining column densities from the simulations and we show the correlation between hydrogen column densities and several properties of the haloes. The hydrogen column densities towards star particles correlate strongly with the total mass, stellar mass and star formation rate of their host haloes. Correcting for these correlations leaves the spread of about two orders of magnitude in the column densities (for given halo mass, stellar mass or star formation rate) in tact, but there is no correlation left with

either of the other parameters we tried.

Here we will use the metal column densities from star particles integrated outward along a line of sight (L.O.S.) for 100 kpc. We tested the distance up to which the L.O.S. integration should be carried out in the following way. Increasing the distance up to which gas particles are still counted increases the column density up to a few to several tens of kpc, after which it remains constant (there is no EoS gas in between haloes). We therefore chose to use 100 proper h^{-1}kpc for the maximum distance at which particles can still contribute to the L.O.S. towards a star particle. Absorption by the intervening intergalactic medium can be taken into account separately, but we will not consider it here.

4.4.1 Method

In Appendix 4.6 we we explained how we compute hydrogen column densities to individual star particles. Because the optical depth due to dust extinction is expected to scale with the metal column rather than the hydrogen column density (dust after all is made up of metals), we estimate dust attenuation for an individual star particle from the metallicity-weighted hydrogen column density in front of the particle and the observed relation between the gas column density and the optical depth in some band for solar metallicity. We then use an extinction law in order to get optical depths for any desired effective wavelength.

The observed relation is based on observations in the solar neighbourhood (at solar metallicity) and relates the optical depth in the *B*-band to the hydrogen column density: $\tau_B = 5.3{\times}10^{-22} N_{\rm H}$ as derived by Xu et al. (1997), which is consistent with the results of Larson et al. (2000). With a solar metallicity of 0.012 and the assumption that the optical depth in the *B*-band due to dust attenuation scales linearly with metallicity, we become

$$\tau_B = 5.3 \times 10^{-22} \int n_{\rm H} \left(\frac{Z}{Z_\odot}\right) {\rm d}L \qquad (4.1)$$

where $n_{\rm H}$ is the hydrogen number density and Z is the metallicity. In order to get the attenuation in other wavebands an extinction law can be used. Because the present approach is already full of uncertainties itself, we decide to use a simple power law approximation of the extinction law, given by

$$\frac{\tau_\lambda}{\tau_B} = \left(\frac{\lambda_{\rm eff}}{\lambda_{B,\,\rm eff}}\right)^{-n} \qquad (4.2)$$

in which τ_λ is the optical depth at the effective wavelength $\lambda_{\rm eff}$, τ_B the optical depth in *B* from Eq. 4.1 and $\lambda_{B,\,\rm eff} = 4391$Å. The value for the extinction law index *n* varies among different authors, ranging from 0.7 in the model of Charlot & Fall

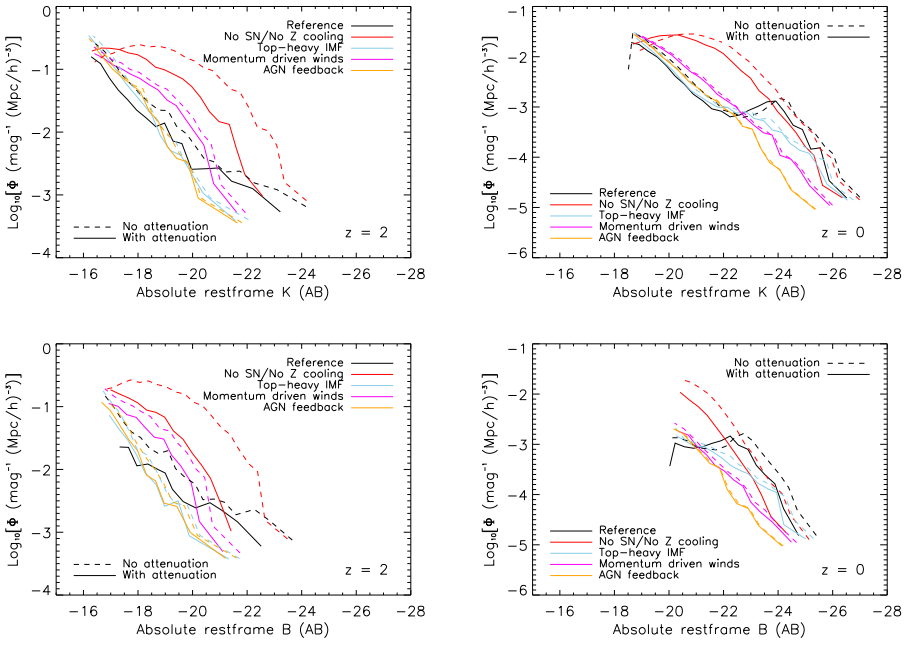

Figure 4.4: The LFs in K (upper panels) and B (lower panels) at redshift 2 (left panels) and 0 (right panels), both including dust attenuation (solid lines) and excluding attenuation(dotted curves, which are also shown in Fig. 4.3). The K-band LF is less influenced by dust attenuation than the B-band LF. Strong SN feedback and AGN feedback lower the high density gas content of galaxies and therefore result in a smaller difference between the attenuated and unattenuated LF.

(2000) to 1.5 for the SMC (Weingartner & Draine, 2001). We use 0.96, after Shao et al. (2007). Some studies use extinction laws which are not pure power-laws, but have features, like the often used Calzetti et al. (1994, 2000) extinction law.

4.4.2 Luminosity functions with dust extinction

The inclusion of dust extinction can potentially change the sign of the effect of feedback on the LF. In simulations with more effective feedback, haloes contain less gas and form less stars. However, the dust attenuation may be much lower in such simulations due to the much lower gas densities.

In Fig. 4.4 we show the rest-frame $B-$ and $K-$band LFs with and without dust for the same simulations and redshifts as in Fig. 4.3. The effective wavelength of the K-band is almost 5 times longer than that of the B-band, so the optical depth

in *K* is roughly 5 times smaller than that in *B*, which corresponds to about 1.7 magnitudes. The difference between the LFs (compare adjacent panels) seems, however, to be somewhat smaller. This arises because the star particles which are visible are the ones with a low extinction, whereas the ones that are heavily obscured are invisible in both filters (even in the *K*-band the optical depth is high for the highest column densities).

Fig. 4.4 shows clearly that attenuation is less important in simulations with more effective feedback: the simulation with AGN feedback hardly shows any attenuation at all, even in the *B*-band, because of the efficient removal of ISM gas, especially in the high mass galaxies.

4.4.3 The shortcomings of dust attenuation estimates in cosmological SPH simulations

The ideal situation for simulators would be to rely on dust attenuation estimates observers obtain from their data, and compare their attenuation-corrected luminosities directly to the unattenuated luminosities of simulated galaxies. Unfortunately, the observationally inferred extinction corrections are highly uncertain and are only available for galaxies in which the attenuation was sufficiently modest for the galaxy to be detectable. Moreover, there is not one single number for attenuation of a galaxy. Different stellar populations are covered by different dust clouds, but the only quantity that can usually be inferred from the observations is the effective attenuation of the integrated light of the galaxy. Although these issues make the inclusion of dust attenuation in simulations worthwhile, this procedure is also highly uncertain.

We have only showed one particular estimate of the dust attenuation, while much more are possible. Some authors have used even simpler methods than the one presented here, and others have used much more complicated, (approximate) radiative transfer schemes in order to estimate the simulated galaxy SEDs (e.g. Jonsson, 2006). In real galaxies, small, cold and dense clouds are responsible for the strongest extinctions, but cosmological simulations such as ours do not yet include this cold, interstellar gas phase.

If more ‘realistic’ scenarios for extinction are desired, knowledge about the number, covering factor and column densities of absorbing clouds is required and would thus necessitate some ad hoc assumptions. Even if we were to use a multiphase model for the gas on the equation of state (e.g. Jonsson et al., 2010; Scannapieco et al., 2010) we would still need to assume some distribution of cloud numbers and sizes. As all this information is not predicted by the simulation itself, the model for the extinction would become disconnected from the simulation. This is the main reason why we instead chose to use an observed relation between metal

column density and the optical depth, which itself is the average of a sample of sight lines that pass through some number of absorbing clouds.

Another method for estimating the attenuation, based on an inversion of the Kennicutt-Schmidt law (in which the column density of gas is estimated from the star formation rate of a galaxy), tends to strongly overestimate the attenuation, because the total attenuation of a galaxy is weighted by the visible luminosity of star particles, and therefore biased towards the least attenuated parts of a galaxy.

4.5 Mock images and galaxy selection

Observers identify galaxies on an image by grouping regions of pixels that have a certain minimum surface brightness. Using packages like SExtractor (Bertin & Arnouts, 1996) this processed can be automized in a clear and reproducible way. The methods for galaxy identification used for simulations and observations are therefore very different. We have used SubFind to identify gravitationally self-bound structures, in the centres of which there usually are stars. A galaxy's luminosity is then just the sum of the luminosities of the bound star particles. It is therefore not at all clear that the observers' way of identifying galaxies will give the same brightness for the galaxies as the methods commonly applied to simulations, even if the underlying distributions of galaxies are identical. For example, there might be projection effects in which smaller (satellite) galaxies cannot be separated from the bright galaxy in front or behind them and extended haloes of stellar light may dissolve into the background.

By creating virtual observations, we can perform a galaxy selection procedure very similar to the one used by observers. In this section we will first describe the creation of the images, followed by a description of the SExtractor settings we use to identify galaxies in the images. We will investigate the influence that SExtractor parameters and the parameters for the creation of the mock images (e.g. the size of the point spread function (PSF), the noise properties and the pixel size) have on the obtained LFs. We will use parameter values that are roughly equal to those in a selected set of observations . We will only use the reference simulation at redshift 2, in a 25 h^{-1}Mpc volume, with $2{\times}512^3$ particles, except in the resolution study, where we vary the particle number. We will not use the $z = 0$ resolution, because our simulations that go down to $z = 0$ do not have sufficient resolution to match observed data sets.

4.5.1 Creating mock images

We will make images with properties that roughly match those of real observations. We focus on three different surveys. For the *Hubble* Ultra Deep Field (UDF,

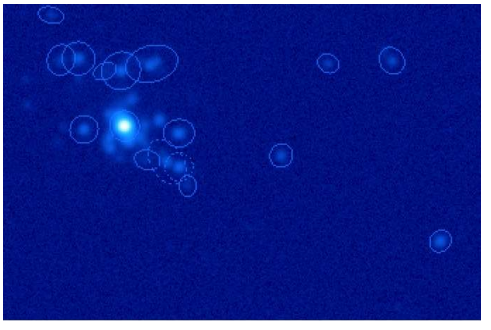

Figure 4.5: An example of the images produced by the described procedure. These come from the reference simulation at redshift two, in the observed MEGACAM i-band. The image shows 340×230 pixels, which corresponds to 17×11.5 arcsec and 580×390 comoving kpc. The parameters are those used to mimic the Hubble UDF, so the PSF is a Gaussian with a FWHM of 0.12 arcsec and the noise is Gaussian with a FWHM of 28 mag arcsec^{-2}. The overdrawn apertures are the apertures defined by SExtractor (check image type 'APERTURES') on which the photometry is performed.

Beckwith et al., 2006) the pixel size is 0.05 arcsec, the point spread function can be approximated by a Gaussian with a full width at half maximum (FWHM) of 0.12 arcsec, and the noise on the image (after subtracting the mean) is well approximated by a Gaussian with a standard deviation of 28 mag arcsec^{-2}. For the VIMOS VLT Deep Survey (VVDS, Le Fèvre et al., 2004) the pixel size and PSF (Gaussian FWHM) are 0.205 arcsec and 0.8 arcsec, respectively. The background was different for different nights. For simplicity we assume the noise level to be the same as the one for the UDF: 28 mag arcsec^{-2} (which is unrealistic, but as we will show below does not influence the results). The last survey we investigate is the Canada France Hawaii Telescope Legacy Survey (CFHTLS, Cabanac et al., 2007). In order to compare to the 'wide' survey of CFHTLS, we assume typical exposure times of one hour, which with their standard read-out noise and (AB) magnitude zero-points results in a background subtracted noise described by a Gaussian with FWHM 26.9 mag arcsec^{-2} on pixels of 0.186 arcsec with a PSF (average seeing conditions) of 0.7 arcsec.

The observational surveys used slightly different filter sets, but in order to facilitate direct comparisons we always use the i-band filter of MEGACAM, as used in the CFHTLS. This is their reddest filter with a reasonable depth. It has an effective wavelength of 770 nm, which at redshift 2 corresponds to an effective rest-frame wavelength of 257 nm. The noise levels for the other data sets are taken from

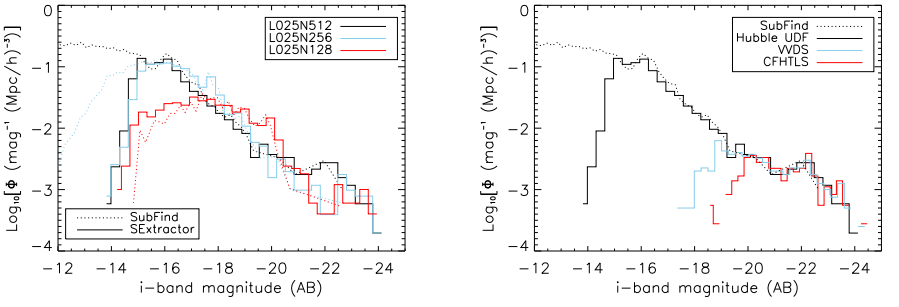

Figure 4.6: *Left panel:* the LF from both SubFind (dotted lines) and SExtractor (solid lines), for three simulations with different numerical resolution, all with the same parameters for the creation of the images and for the detection of galaxies. The differences between the LFs obtained by SExtractor for the different resolutions are entirely due to the underlying distribution of sources and the creation of images and measuring the luminosity functions using SExtractor are insensitive to numerical resolution. *Right panel:* The LF as obtained from SubFind, i.e. directly from the simulations (dotted line), compared to the luminosity function obtained using mock images, mimicking three different observational data sets: Hubble UDF, VVDS and CFHTLS, respectively. The parameters of those data sets can be found in the text.

nearby filters (*F814* and *I* for UDF and VVDS, respectively).

For computational efficiency we cut the simulation box in 10×10 regions (of $2.5 \times 2.5 \times 25$ comoving h^{-1}Mpc) and project the stars along the long axis onto pixels of the desired size.

In Fig. 4.5 we show an example of (part of) an image created from the reference simulation at redshift two, with the SExtractor apertures overdrawn. The image shows 340×230 pixels, which corresponds to 17×11.5 arcsec and 580×390 comoving kpc. The parameters are those used to mimic the Hubble UDF, so the PSF is a Gaussian with a FWHM of 0.12 arcsec and the noise is Gaussian with a FWHM of 28 mag arcsec^{-2}. In low density regions it is clear that most if not all sources are well defined. In higher density regions, like the area left of the centre of the image, some emission can be missed. We will use the 'mag_auto' magnitudes of SExtractor in the remainder of this chapter.

4.5.2 Comparing the LFs from halo catalogues and mock observations

Fig. 4.6 compares the LFs obtained from virtual observations with those obtained from halo catalogues. The mock observations used parameters for the images and for SExtractor as described above for the Hubble UDF). The left panel compares three simulations with different numerical resolution. In each case the LFs from the mock observations (solid histograms) match those from the halo catalogues (dotted lines) well down to about $i = -15$ ($i = -17$ for the lowest resolution shown). The difference in the LFs from SExtractor for different resolutions are therefore the result of the different star formation histories of the haloes in the simulations, and not of the mock image procedure.

In the right panel of Fig. 4.6 we show the LF from SubFind together with three LFs from mock images, all for the reference simulation at the highest resolution at $z = 2$. These three closely mimic the UDF, the VVDS and the CFHTLS. Down to the luminosities that define the depths of the surveys all three observational data sets follow the LF obtained directly from halo catalogues very well. Thus, if our simulated galaxy sample would be observed by either of the three telescope/instrument combinations described, then the LF obtained would be nearly exactly the same as the true LF of the galaxies in the simulation, down to some limiting magnitude. It is, nevertheless, not straightforward to compare the three LFs, as they differ simultaneously in terms of the background noise level, pixel size and PSF size (the SExtractor settings for the three LFs are, however, identical). In the following sections we will look at the effect of varying the parameters for the mock image creation and for the detection and photometry of the galaxies. We will vary the parameters one-by-one using the parameter values corresponding to the HUDF as our baseline.

4.5.3 Mock image parameters

In this section we will investigate the dependence of the galaxy selection technique using mock images and SExtractor on the parameters used for creating the virtual observations. Specifically, we will vary the noise level, the size of the pixels and the size of the Gaussian PSF. These are all varied away from the HUDF values by factors of 5 and 10, roughly bracketing the other observational data sets.

Increasing the noise level by factors of 5 or 10 results in the LFs shown in the upper left panel of Fig. 4.7. As expected, the minimum observed galaxy luminosity increases with the noise level. The LF is already complete for galaxies that are only 1 magnitude brighter than the faintest detected objects at a given noise level (the underlying galaxy population is the same for the three LFs and shown by the dotted

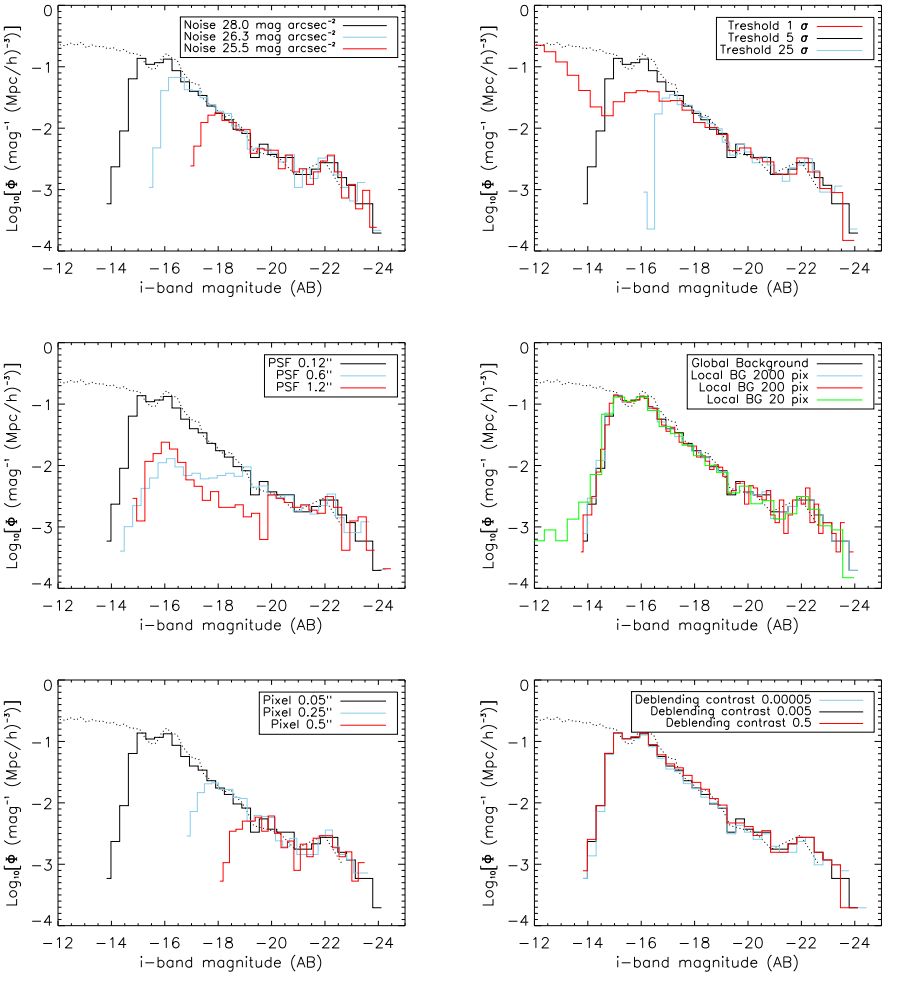

Figure 4.7: The influence of several image creation and SExtractor parameters on the obtained LF. In the *left column* we vary the parameters used when creating the image, from top to bottom: the size of pixels (and thus the resolution) of the images, the size of the PSF (varied over a factor 5 and 10) and the amount of noise (the HUDF value and that increased by a factor of 5 and 10). In the *right column* we vary three important SExtractor parameters, from top to bottom: the minimum deviation of the pixels above the background, the deblending contrast and the background (global as used in the other plots versus a local background determined on a small and on a large area). The solid black line is the default model and is the same in every panel. The dotted black line is the LF of the underlying galaxy population directly from the simulations.

line).

Changing the size of the PSF (from the default, a Gaussian with a standard deviation of 0.12") strongly affects the recovered galaxy population. If we increase the PSF (at fixed pixel size) by a factor of 5 or 10, so to a Gaussian with standard deviations of 0.6" and 1.2", we detect less galaxies as we smear the galaxies out more and we start to loose the low surface brightness objects. Increasing the PSF therefore results in a much flatter LF over a large range of galaxy luminosities, as shown in the middle left panel of Fig. 4.7. Only the very brightest galaxies follow the same LF as the one obtained from images with a smaller PSF. While the SExtractor LF agrees with the intrinsic (i.e SubFind) LF for $i < -15$ for a 0.12" PSF, the two only agree for $i < -19$ and $i < -20$ for PSFs of 0.6" and 1.2", respectively. Note that a PSF with a standard deviation of 1.2" is not extremely large. This may indicate that at high redshift the LF, if observed with spatial resolutions typical for ground based surveys, may be strongly flattened due to the selection technique used to identify the galaxies. Hence, if the ground based surveys shown in Fig. 4.6 had integrated longer and therefore produced observations with have a higher signal-to-noise ratio, then the low luminosity end of the measured LF would most likely be shallower than that of the underlying population of galaxies.

The size of the pixels (for a fixed PSF size) is also important for the selection of galaxies, as shown in the lower left panel of Fig. 4.7. Increasing the pixel size results in a higher minimum detected luminosity. Note, however, that in the images with larger pixel sizes, the FWHM of PSF is smaller than a pixel, which is unrealistic.

4.5.4 SExtractor configuration file parameters

The detection of galaxies with SExtractor depends on many parameters, some of which are more crucial than others. In this section we will compare the LFs obtained after changing some of the parameters in the configuration file within acceptable bounds.

The first parameter we vary is the detection threshold. On the image, the software determines a background (for variations of the background estimate, see below) and specifies it by a mean and a standard deviation σ. Sources are then defined to be objects if at least some number of adjacent pixels (5, here) stand out above the background by at least some number of standard deviations. By default, we have set this threshold to 5σ. Decreasing this parameter from 5 to 1σ results in the detection of fainter sources (i.e. sources with a lower surface brightness) as we can see in the upper right panel of Fig. 4.7. A deviation of only 1σ is not rare, so sometimes a collection of noise pixels will be mistaken for a low brightness galaxy. The blending of sources will be more important in highly populated regions, as it

is easier to connect blobs of emission with pixels exceeding the background by 1σ than with pixels exceeding the background by 5σ. The threshold does not influence the shape of the LF at higher luminosities, as the few pixels with noise-like surface brightness do not add significantly to the luminosity of a bright galaxy. Increasing this threshold by a factor of 5 (i.e. demanding a source to stand out 25σ above the background) results in much less faint galaxies, but high luminosity end is again unaffected.

The estimate of the background on the image is also subject to a few SExtractor settings. In the first place, one can choose between a global or a local background. Global means that it is the same over the whole image, while local refers to a locally determined background.We have so far chosen to use a global background, because that is what our mock images contain. Here we investigate whether or not using a local background influences the selection of galaxies (we leave the background the same when creating the images). For example, in highly populated regions of the universe the local background might be higher, blending low luminosity objects. In the middle right panel of Fig. 4.7, we compare the results of using a global background to results using a local background, for which we set the area used to determine the background to three different (and extreme) values: 2000 pixels (almost as big as the image), 200 pixels and 20 pixels (only slightly larger than the objects themselves). Although the last choice extends the LF to slightly fainter levels, the differences are marginal. We conclude that for our method of adding noise, the selection of galaxies with SExtractor is insensitive to the details of the background estimation (but this would probably change if we were to make the background change gradually across the image).

The last parameter we will investigate is the minimum contrast for deblending. Whenever there is a saddle point in the surface brightness distribution, SExtractor has to decide whether there are two slightly overlapping sources, or whether it is one, with substructure. The parameter that influences this decision is the minimum deblending contrast, which is the ratio between the integrated intensity of both sub-peaks. A very high minimum contrast means that fluctuating fields are more likely to be regarded as one source. For more details we refer to Bertin & Arnouts (1996). The number of levels in surface brightness is left at the default setting, 32. In the lower right panel of Fig. 4.7 we compare several minimum deblending contrasts. Varying the parameter over 4 orders of magnitude (from 0.5 to 5×10^{-5}) does not make an appreciable difference for the recovered LF. We therefore conclude that this parameter is not crucial for our purposes. Another conclusion is that blending of sources is not an important effect for the mock images. If it were, a very low minimum contrast would have been able to detect satellite galaxies as seen in projection before or behind a more luminous galaxy. Note, however, that we may underestimate the number of superpositions of physically unrelated objects

due to the small size of our volume. The spatial resolution of the simulations and images (of order 1 kpc) is also large enough to ensure that all galaxies have fairly smooth surface brightness profiles.

We want to note here that at lower redshift (e.g. $z = 0.1$, the median redshift of SDSS) the differences between the LFs from halo catalogues and from virtual observations could well be larger. At low redshift, the massive haloes contain a large halo of 'intra-cluster light', stars that are far away from the central galaxy in a large diffuse halo (up to $\sim$ 30% of the luminosity of what would be called the central galaxy). It is well possible that such haloes are not picked up by SExtractor as being part of the galaxy. The spatial resolution of the simulations that go down to redshift zero are comparable to the SDSS angular resolution, so resolution issues are likely to play a big role, and for this reason we postpone such an analysis to future work.

4.6 Conclusions

We have investigated the LF of galaxies in several simulations, differing in the input physics. We focused on $z = 2$ and $z = 0$ and investigated an implementation of dust attenuation and its effect on the luminosity function in both the K- and the B-band. Finally, we created virtual observations and obtained the LF using the tools observers would use for the same purpose.

Whenever LFs are obtained from observations with inefficient SN feedback (and no AGN feedback) in high mass galaxies, an over-abundance of high luminosity galaxies appears. This shows as a 'bump' in the LF. The high luminosity ends of such simulations correspond to the high luminosity end of the LF of a simulation without any SN feedback, indicating that the feedback in these high luminosity galaxies is indeed very inefficient. In simulations with a top-heavy IMF for star formation at high pressures, for momentum-driven wind models (which have more energy in the winds than available from SN, especially in high mass galaxies) and for simulations with AGN feedback (which are mainly effective in shutting off star formation in high mass haloes) the 'bump' disappears and the LF goes steeply down. The low mass end slope of the luminosity function mainly depends on the mass loading in the winds: higher mass loading results in a flatter slope.

The inclusion of dust attenuation can potentially change the sign of the effect of feedback on shape of the LF, because besides lowering the star formation in massive galaxies it also lowers the gas content and possibly the dust content. We estimated the dust attenuation from the metallicity-weighted gas column density towards individual star particles. By doing so, we make sure that the dust atten-

uation may vary with location in the galaxy and that the attenuation scales with the metallicity and gas column towards star particles. The effect of dust attenuation implemented as such is that the attenuation is strongest in the most luminous galaxies and stronger for simulations with weaker feedback (at given intrinsic luminosity). The difference in attenuation in the *K* and *B*-bands is less than the factor of 5 difference in the optical depth for given column density for these two effective wavelengths, because the total apparent luminosity is dominated by the least attenuated stellar emission.

It needs to be noted, that due to the low spatial (and mass) resolution of cosmological simulations, dust attenuation estimated may not be very reliable. As the bulk of the attenuation in the solar neighbourhood happens due to small high opacity clouds (and there is no reason to suspect this is different in other galaxies), dust extinction happens on scales much smaller than the resolution limit of current simulations.

We made mock observations, mimicking three different observational data sets (the Hubble Ultra Deep Field, the VIMOS VLT Deep Survey and the CFHT Legacy Survey) at redshift 2 in a rest-frame UV band. We conclude that down to the flux limits of the surveys the LFs obtained from the virtual observations agree very well with those obtained directly from the halo catalogues. Interestingly though, the size of the point spread function may become very important for very deep, ground-based surveys. For large PSFs ($\sim$ 1 arcsec), but deep images and low noise levels, objects of low surface brightness are preferentially removed from the sample. As the fraction of low surface brightness galaxies is higher for lower luminosities, this may substantially flatten the low-luminosity end of the observed LF, compared to the underlying intrinsic LF.

LFs should not be used as discriminators between different cosmological hydrodynamical simulations (as is often done for semi-analytic models of galaxy formation). Even if LFs are obtained by convolving halo mass functions (obtained with the favoured cosmology) with the galaxy luminosity as a function of halo mass, plenty of difficulties still remain. The star formation histories of galaxies in models invoking different sub-grid physics vary strongly and a distinction solely based on the shape of the LF will probably prove degenerate. Also, especially for blue rest-frame wavelengths, dust attenuation will be important, and how to estimate dust attenuation for a galaxy is far from trivial.

Acknowledgements

It is a pleasure to thank Hendrik Hildebrandt and Daniel Szomoru for discussions about the observational sides of this study. The simulations presented here were

run on Stella, the LOFAR BlueGene/L system in Groningen and on the Cosmology Machine at the Institute for Computational Cosmology in Durham as part of the Virgo Consortium research programme.

Appendix A. Column densities in SPH simulations

In order to obtain galaxy magnitudes which include the effect of dust attenuation, we calculate column densities through the simulation box. First, we will summarize how densities are defined in SPH and how we calculate column densities along lines of sight through the simulation box. We will then show how hydrogen column densities depend on galaxy properties. Finally, we will use the column densities of metals in order to estimate the attenuation of stellar light on a star particle by star particle basis.

SPH interpolation and densities

In SPH simulations the density field is discretized by a set of particles whose smoothing kernels determine their contribution to a given physical quantity at a given point in space. Although in principle a kernel can have a plethora of shapes, they are usually chosen to be similar to Gaussians, although they go to zero at small distance from the particle. The most commonly used kernel is the one originally proposed by Monaghan (1992), which consist of two parts, both described by a polynomial. This form goes to zero for a value equal to two times the kernel length. Springel (2005) introduced a kernel, which we use, whose shape is exactly the same, but goes to zero at one kernel length and changes prescription half-way:

$$W(r,h) = \frac{8}{\pi h^3} \begin{cases} 1 - 6(\frac{r}{h})^2 + 6(\frac{r}{h})^3 & 0 \leq \frac{r}{h} \leq 0.5 \\ 2(1 - \frac{r}{h})^3 & 0.5 < \frac{r}{h} \leq 1 \\ 0 & \frac{r}{h} > 1 \end{cases} \quad (4.3)$$

Both transitions from one to the other prescription are continuous up to the second derivative. The value of the smoothing length h is chosen such that for every density there are about $N_{\rm SPH}$ particles within one kernel (in our simulation $N_{\rm SPH} = 48$). The value of a given physical quantity at the location of particle i (ρ_i), in this example the density, is now given by a summation of all gas particles that fall within the kernel of point p, weighted by that kernel:

$$\rho_i = \sum_{j=1}^{N} m_j W(|\mathbf{r}_{ij}|, h_i), \quad (4.4)$$

where $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$. The value of the kernel in the point of interest is determined by increasing a sphere until it contains a constant mass, and therefore about $N_{\rm SPH} \simeq 48$ particles. Note that the value of the kernel depends on the kernel at the location the density is desired of, but not on the kernels of the other particles. This is called the 'gather approach', contrary to the 'scatter approach', which we will describe below.

Column densities

Surface mass densities are obtained by integrating the density along a line of sight (L.O.S.):

$$\Sigma = \int_{\rm LOS} \rho(\mathbf{x})\, {\rm d}L \qquad (4.5)$$

where L is a spatial coordinate along the L.O.S.. At every point in space, the physical density is determined by a kernel-weighted sum of all particles (of which ~ 48 should contribute). The method we use to obtain column densities changes the order of the summation over the particles and the integration along the L.O.S.. For every particle we obtain its contribution to the L.O.S. integral, as a function of its impact parameter b. We tabulate the column densities as a function of b in a look-up table. We distinguish between particles whose kernels are crossed completely (from one edge to the other) and particles for which the L.O.S. either starts or ends within its kernel.

For computational efficiency we use the kernel of the particle under consideration, rather than the kernels of all points along the L.O.S.. This is slightly inconsistent with the actual SPH formulation used in GADGET, as explained in Section 4.6 (we use the 'scatter' rather than the 'gather' approach). However, within one kernel length the density will not vary much, leaving the kernels very similar and the errors introduced by this procedure are minor ($\lesssim 1\%$).

For particles with kernels that are completely crossed by the L.O.S.the look-up table is one dimensional, having the impact parameter as the only parameter. We tabulate a value for the column density, normalized to mass ('column density per unit mass') and we use spatial coordinates normalized to the kernel (i.e. b/h). The integral is evaluated numerically using Romberg integration. The impact parameter is taken to be between 0 and 1 in 10.000 steps. This reduces the relative error in the kernels of the particles along the L.O.S. to be $O(10^{-4})$.

For particles whose kernel overlaps with the start or the end of the L.O.S., the estimate described above would be too high. For these particles we use a two-dimensional table of (mass- and kernel-normalized) column densities, with the impact parameter and a 'distance parameter' (the distance through the kernel, perpendicular to the impact parameter). Due to memory issues we use a lower precision

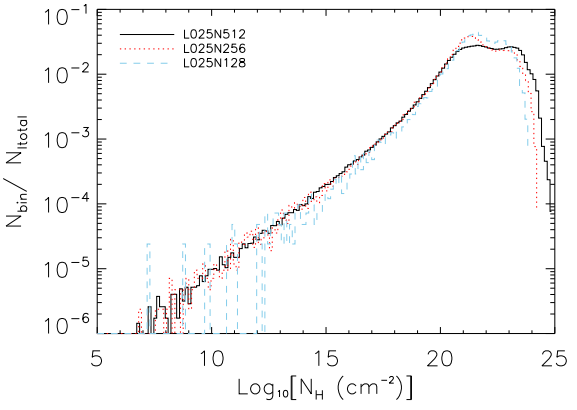

Figure 4.8: The distribution function of column densities in the 'REF' simulation at three different resolutions: 512^3, 256^3 and 128^3 particles. The number of lines of sight in a bin is divided by the total number in the box. Each L.O.S. targets a different star particle.

for this table than for the one-dimensional one. As this table will only be used for a small fraction of the particles (for L.O.S. much longer than a typical particle kernel), the number of steps for both parameters between one and zero is chosen to be 1000. This is a minor limitation, as there are many more particles crossed completely than partially, for L.O.S. longer than a typical particle kernel ($\sim$ kpc). In this work we will mostly consider L.O.S. of 100 kpc, whereas the kernels deep inside the haloes are $\sim$ 1 kpc in the high-resolution simulations.

Column densities towards star particles in cosmological SPH simulations

Selection of gas

In this paper we are interested in the attenuation towards star particles by the gas in a galaxy. Because we are interested in the gas of a galaxy that resembles the ISM, we will only take gas into account that has a density exceeding the star formation threshold. This gas is expected to be responsible for the biggest part of the attenuation and is probably closely related to the gas measured in 21 cm observations and the gas probed by molecular indicators (although this would be a fraction of all EoS gas (in volume even a small fraction), the other part of the EoS gas being the warm and ionized ISM). The L.O.S. are taken along the main axes of the simulation box, and therefore the galaxies are oriented randomly relative to the L.O.S..

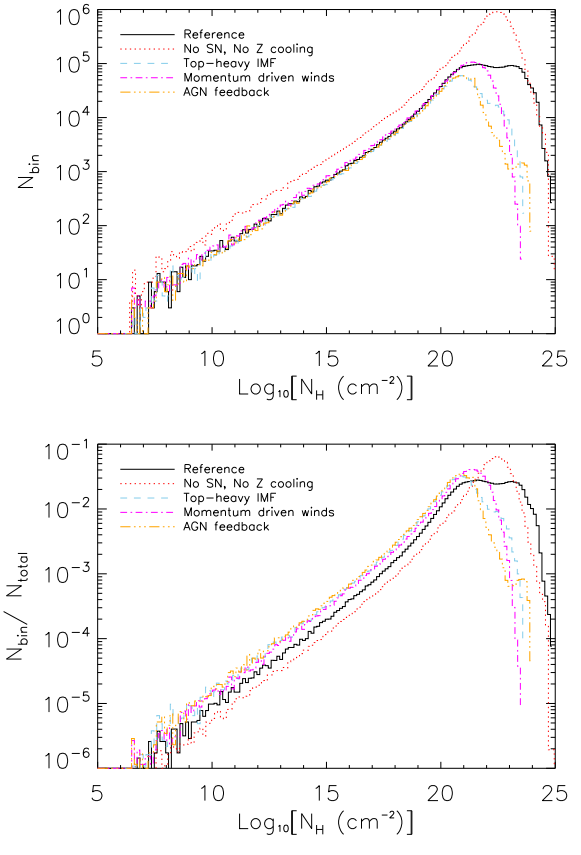

Figure 4.9: The distribution functions of column densities towards star particles in the reference simulation, the simulation without metal-line cooling and SN feedback the simulation in which star formation at high densities occurs with a top-heavy IMF (and the extra available energy is used to increase the wind velocity), the simulation with momentum-driven winds and the simulation including AGN feedback. The low column density ends of the distribution are exactly the same for the simulations with SN feedback The second, higher peak in the reference simulation is absent in the simulation with effective feedback in high mass haloes.

We tested the distance up to which the L.O.S. integration should be carried out in the following way. Increasing the distance up to which gas particles are still counted increases the column density up to a few to several tens of kpc, after which it remains constant (there is no EoS gas in between haloes). We therefore chose to use 100 proper kpc/h for the maximum distance at which particles can still contribute to the L.O.S. towards a star particle. When obtaining observables, absorption by the IGM could be added separately, but we will not do so here.

Convergence of the L.O.S. column densities

To be sure that the column densities obtained are not strongly dependent on the numerical resolution, we plot the normalized distributions in Figure 4.8 for three different particle numbers in the 'REF' simulation in a 25 Mpc box. The value of every bin is the number of lines of sight in that bin divided by the total number of lines of sight in the box (i.e. the number of star particles in the box). The high column density end of the distribution is slightly dependent on resolution. This was expected as the effectiveness of feedback is also somewhat resolution dependent (Chapter 2). Below $N_H = 10^{21}$ cm^{-2} the distribution functions are, however, very similar. The high column density cut-off is higher for higher resolution simulations, because the highest volume density that can be reached depends on the resolution (it scales roughly with the particle mass divided by the softening length squared, and softening and particle mass depend on resolution). In the next section we will show that the differences between different physical models are far larger than the resolution effects.

Distributions of hydrogen column densities for different input physics

The distribution of hydrogen column densities ($N_H = X_H \Sigma / m_{\rm H}$) towards all star particles in a box of (25 Mpc/h)^3, with the reference parameter set and the three different simulations used in this paper are shown in Figure 4.9. The double peak is a feature arising from the ineffective feedback in high mass haloes, where the winds are not able to escape (Chapter 2). To illustrate this, Figure 4.9 also shows the distribution for a simulation in which star formation at densities above some threshold pressure happens with a top-heavy IMF. The extra available energy per unit stellar mass formed is used to increase the velocity of the wind from 600 to 1618 km s^{-1}. The second, high column density peak is absent, while the rest of the distribution is unaffected. In the model without any SN feedback and without metal-line cooling the very highest column densities reached are as high as the highest in the reference simulation. This (and the fact that the highest column densities in the simulation with a top heavy IMF at high pressure are lower) indicates that the highest (res-

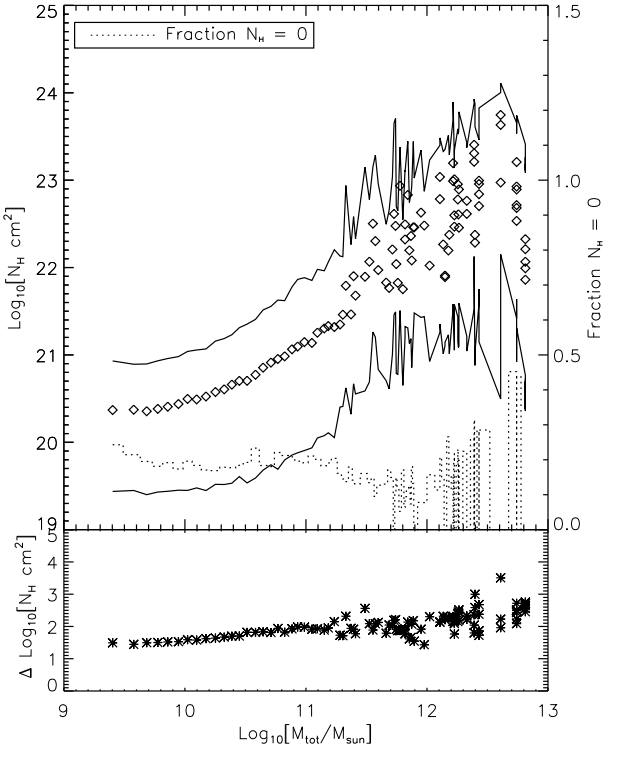

Figure 4.10: The relation between the hydrogen column densities along lines of sight to all star particles, binned according to the total mass of the halo they are in. In the upper panel, the diamonds are the medians in the bins on both axes (all bins contain equally many L.O.S.). The thin solid lines are the 16 and 84 percentile lines of the distribution in the bin. The dotted line (and right vertical axis) indicates the fraction of L.O.S. with zero column density. The lower panel indicates the range between the percentile lines in the upper plot.

olution dependent) column densities that can be reached are only reached when feedback is inefficient. The overall normalization of the model without SN feedback and without metal-line cooling is much higher, because many more stars have formed in that simulation. In the simulation with momentum driven winds and in the simulation with AGN feedback the distribution is very similar to the one in the simulation with the double IMF, with some small differences at the high column density end. The low column density end in all simulations with SN feedback is very similar, which is an illustration of the self-regulation of star formation by SN feedback.

Correlations between hydrogen column densities and halo properties

In order to investigate the dependence of the column densities along the line of sight to star particles on properties of the galaxies they are in, we use Friends-of-Friends haloes. For the star particles in haloes, which we will loosely call galaxies, we will correlate the median column densities in bins of physical properties like stellar mass, star formation rate or ISM mass of the same halo.

We bin the L.O.S. such that in every bin are equally many L.O.S., according to halo mass and star formation rate surface density ($\Sigma_{\rm SFR} = {\rm SFR}/R_{vir}^2$, where the total FoF group mass is used as the virial mass) for the reference simulation in Figures 4.10 and 4.11, respectively. The symbols plotted are the medians in both horizontal and vertical quantities. The thin lines are the 25th and 75th percentiles, while the thick straight solid line in Fig. 4.11 is a power-law fit to the medians. The dotted line indicates the fraction of L.O.S. that have zero column densities, as indicated on the right vertical axis (bins for which this fraction exceeds 0.5 are not used for the power law fits we discuss later on). The lower panel shows the difference between the two percentile lines as a function of the same halo quantity. In every case, the spread is about 2 orders of magnitude, with only a weak dependence on halo properties.

As can be seen from these plots, there is a tight correlation between the hydrogen column density and halo mass or star formation rate surface density. This also holds for the correlation between the hydrogen column density and stellar mass and between hydrogen column density and star formation rate (not shown), The spread in the data is about two orders of magnitude and increases slightly with halo mass, star formation rate and/or star formation rate surface density. The fraction of star particles that have zero column density is roughly constant at $\sim$ 15%. In high mass haloes this starts fluctuating more, because the fraction of star particles far away from the centre is higher, and also satellite galaxies contribute (centrals and satellites will be in the same FoF halo).

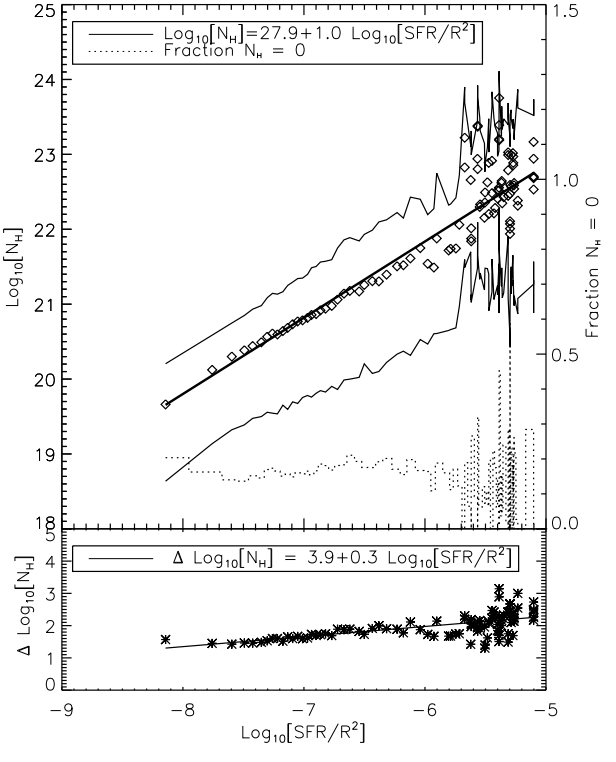

Figure 4.11: The relation between the hydrogen column densities along lines of sight to all star particles, binned according to the star formation rate surface density of the halo they are in for the reference simulation at redshift 2. Symbols are explained in Figure 4.10. The solid lines are power law fits to the medians, the relations for which are given in the legend.

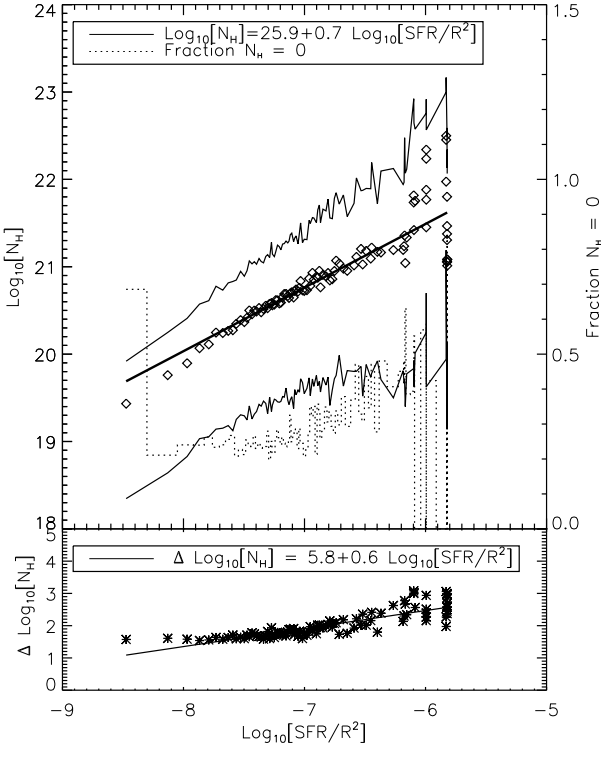

Figure 4.12: Same as Fig. 4.11, but now for the simulation with a top heavy IMF for star formation above a density threshold. The relation between median column density and star formation rate surface density is much flatter than in the reference simulation. Note that the extent of the vertical axis is different than in Fig. 4.11.

Effective feedback

In the DBLIMF, in which stellar feedback is efficient at all halo masses, winds escape more easily from high mass haloes. In these haloes, the gas density is expected to be lower, and the star formation rate also is. Therefore, this different implementation of star formation and feedback might result in different correlations between halo properties and column densities. We saw already before that the very high column density end was truncated for this simulation (c.f. Figure 4.9). In Figure 4.12 we show the correlation between the hydrogen column density and the star formation rate surface density as defined before. Indeed, the slope of the correlation is strongly different (~ 0.7 instead of ~ 1).

Table 4.1: Correlations (Spearman rank correlation ρ) and power law exponents (n) from fits between the hydrogen column density and several halo properties in two of our simulations. The two not available power law indices represent simulations for which the correlation between that halo property and halo mass cannot be well described by a power law, as shown in Fig. 4.10 for the reference simulation.

	Reference		Top-heavy IMF	
Halo property	ρ	n	ρ	n
$M_{\rm halo}$	0.54	n.a.	0.36	n.a.
M_*	0.55	0.76	0.37	0.36
SFR	0.55	0.60	0.41	0.33
SFR/$R^2_{\rm vir}$	0.54	1.02	0.41	0.73

Correlation strengths

To discriminate between the different correlations found in the previous section, we calculate a Spearman rank correlation coefficient. The value of this coefficient will always be between -1 (perfect anti-correlation) and 1 (perfectly correlated), while a value close to 0 indicates a lack of correlation.

In Table 4.1 we compile the correlation coefficients ρ for correlations between the hydrogen column density and different halo properties. For the simulations for which the relation between the halo property and the column densities are well represented by a power-law, we also give the result for a fit of the form

$$N_H(X) \propto X^n \tag{4.6}$$

in which X are the halo properties, and n is the power law index quoted in the table. These are the fits shown in the figures of the correlations as well. We only fit on bins in column densities with non-zero medians, as otherwise the fit result depends on the value we assign to those. For the rank correlation coefficients we do include star particles with zero column density.

Correlation of residuals

If we correct all values for the column density for the value expected from one of the halo quantities (i.e. using the fit results quoted in Table 4.1, with the proper normalization), we remove the dependency on one of the relevant parameters. We can now investigate whether or not residuals correlate with other halo properties. So, we now do the same analysis, but between a corrected hydrogen column density ($\Delta \log(N_H) = \log(N_H) - \log(N_H)(X)$) and the other halo properties.

When doing so for every combination of parameters reported in Table 4.1, we find that all other correlations are insignificant and power law fits give slopes very close to zero (with normalization also close to zero), with the same spread of two orders of magnitude in the corrected hydrogen column density. The example of the REF simulation, with the L.O.S. corrected for the median value as a function of star formation rate surface density plotted against the stellar mass of their halo, is shown in Figure 4.13. There is no correlation left, the distribution is still two orders of magnitude wide and the median is very close to zero with a very weak dependence on the stellar mass. Without showing the results for other combinations of halo properties, we emphasize that this is true for all combinations of halo properties that correlate well with the halo mass (like, e.g. stellar mass, star formation rate and star formation rate surface density).

This means that the the median of column densities in a halo is uniquely determined by either one of the parameters, and that, after correcting for the correlation with that property, there are no correlations left in the data set between these halo parameters and residual column density. We are not able to determine the source of the scatter and do not identify what causes its extent.

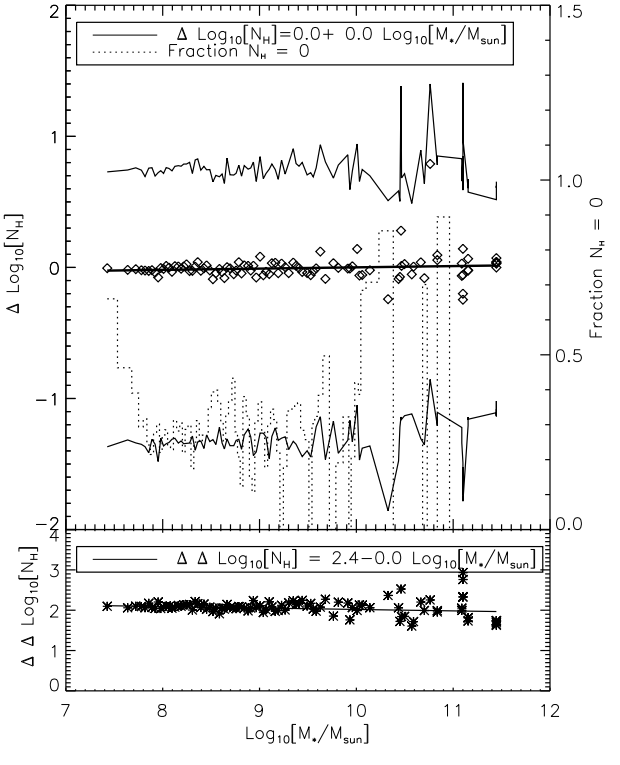

Figure 4.13: The relation between the column densities, corrected for the trend found with star formation rate surface density, and the stellar mass of the halo the corresponding star is in for the reference simulation at redshift 2. Symbols are explained in Figure 4.10. The lower panel shows the width of the distribution of residuals, similar to the lower panels of Figs. 4.10, 4.11 and 4.12. There is no relation between the residuals and any other halo property that correlates with halo mass.