

Chapter 2

The Galaxy Formation Model

Right now, Will, you're 'tooling up', you're learning FORTRAN and PYTHON; it's like you're in the bike shop deciding if you want the blue shorts or the red shorts. Here, GALFORM is the bike, and once you get on it, it will take you places.

— Carlos S. Frenk, 2012

2.1 Introduction

The galaxy formation model used throughout this thesis is the Durham semi-analytical model GALFORM. It was first introduced in Cole et al. (2000), and builds on ideas outlined earlier by White and Rees (1978), White and Frenk (1991), Lacey and Silk (1991), Cole (1991) and Cole et al. (1994). There have been subsequent revisions since its initial inception (e.g. Benson et al. 2003; Baugh et al. 2005; Bower et al. 2006; Lagos et al. 2011; Gonzalez-Perez et al. 2014). However, this thesis is mostly concerned with the latest version presented in Lacey et al. (2016). This particular version is used as it can simultaneously reproduce optical luminosity functions at $z = 0$, rest-frame near-IR luminosity functions for $z < 3$, rest-frame UV luminosity functions for $z \sim 3 - 6$ and, importantly for this thesis, the number counts and redshift distributions of galaxies selected at sub-mm wavelengths (see Chap. 3). Some features of this version of GALFORM are discussed in more detail in Sect. 2.10.

In GALFORM, galaxy formation is modelled *ab initio*,¹ beginning with a specified cosmology and a linear power spectrum of density fluctuations, and ending with predicted galaxy properties at different redshifts. The process is treated in two stages. First the gravitational collapse of dark matter into halos is modelled, either through a Monte Carlo technique based on the extended Press-Schechter formalism

¹In the sense that the calculation begins at high redshift ($z \gtrsim 20$).

(e.g. Parkinson et al. 2008), or through N -body simulation (e.g. Springel et al. 2005). Structure formation proceeds *hierarchically*. Smaller halos form earlier and coalesce, or merge, over time to create larger structures. Once the merging history of the halos is specified, baryonic physics is added as a set of equations which are designed to describe the complicated (and in many cases poorly understood) physical processes important for galaxy formation. In GALFORM these equations essentially track the exchange of mass and metals between different baryonic reservoirs: (i) hot halo gas, (ii) cold gas in galaxies, (iii) stars, (iv) ejected gas outside the virial radius and (v) supermassive black holes. The cold gas and stars can either be in a disc component formed through the cooling of hot gas, or in a bulge/spheroid, formed through a dynamical process, either a galaxy merger or a disc instability. Stellar luminosities are computed through combining the star formation and metal enrichment history of a galaxy predicted by the model with a stellar evolutionary population synthesis model (e.g. Tinsley 1972; Bruzual and Charlot 2003; Maraston 2005). The absorption and re-emission of stellar radiation by interstellar dust is computed assuming a simple geometry for the distribution of dust. The equations of radiative transfer are then solved for this geometry and global energy balance arguments are applied for the dust emission, assuming the dust emits as a modified blackbody.

The simplified nature of the equations used to describe the complex physical processes involved introduces a number of free parameters into the semi-analytical approach. These are calibrated by comparing the predictions of the model against a pre-determined set of observational data. This calibration is typically done manually [though see Bower et al. (2010), Henriques et al. (2013) and Rodrigues et al. (2017) for examples of statistical techniques being used to automate this procedure for semi-analytical models] and provides a strong constraint on the available parameter space, as discussed in Lacey et al. (2016). The GALFORM strategy is to calibrate the model against directly observed quantities (e.g. luminosities) rather than physical properties that have been inferred from observations (e.g. stellar mass). Inferring properties such as stellar mass from observed broadband photometry requires making assumptions about the star formation and chemical enrichment history of a galaxy, which may not be consistent with those predicted by GALFORM. In addition, as the model adopts multiple IMFs (see Sects. 2.8.2 and 2.10) the only consistent way to compare with observations is with directly observed properties, as inferred properties such as stellar mass are sensitive to the assumed IMF, which is no longer unique. For more details of the semi-analytical method, see the reviews by Baugh (2006) and Benson (2010).

The semi-analytical approach to galaxy formation on a cosmological scale is complementary to recent gas-dynamical cosmological simulations (e.g. Vogelsberger et al. 2014; Schaye et al. 2015) and the two methods can produce similar results when calibrated in a similar manner (e.g. Guo et al. 2016). Gas-dynamical simulations compute the distribution and dynamics of the gas with fewer approximations and can thus make detailed predictions regarding the internal structure of galaxies and their halos, rather than just global properties. However, they are still limited in terms of the volumes that can currently be simulated in this way. For example the reference EAGLE (Evolution and Assembly of GaLaxies and their Environments) simulation (e.g. Schaye et al. 2015; Crain et al. 2015) took ~ 4.5 million CPU hours

to simulate a $\sim 100^3 \text{ Mpc}^3$ volume to $z = 0$, whereas the GALFORM model used throughout this thesis takes only ~ 140 CPU hours to populate a $\sim 700^3 \text{ Mpc}^3$ volume with galaxies.² Thus, semi-analytical models are more suited to interpreting large observational galaxy surveys, or populations of rare objects, and can be used to explore the parameter space relevant for reproducing various observational datasets. Gas-dynamical simulations are also forced to use ‘sub-grid’ models to describe many of the physical processes important for galaxy formation such as star formation and supernova feedback, that occur on scales currently below the resolution of the simulation. These are analogous to what is done in semi-analytical models, albeit on smaller spatial scales, and thus current cosmological simulations also require that free parameters in the sub-grid model are calibrated against observational data.

Just as advances in our understanding of how physical processes such as supernova and AGN feedback affect galaxy formation and evolution, gleaned from studies of semi-analytical models (e.g. Cole et al. 2000; Benson et al. 2003; Bower et al. 2006; Croton et al. 2006), have informed the current gas-dynamical simulations, it is expected that further study of the predictions that these simulations can make will inform future semi-analytical models.

2.2 Hierarchical Structure Formation

In GALFORM the baryonic processes are implemented within a halo merger tree, the merging history of a dark matter halo from its progenitors down to the redshift of interest. Once this is specified, various halo properties, and their evolution, can be incorporated into the galaxy formation calculation.

These halo merger trees can be constructed in two ways: (i) from a Monte Carlo technique following the extended Press-Schechter formalism (e.g. Bond et al. 1991; Lacey and Cole 1993; Cole et al. 2000; Parkinson et al. 2008) or; (ii) calculated directly from N -body dark matter only simulations (e.g. Helly et al. 2003; Springel et al. 2005; Bower et al. 2006; Guo et al. 2013; Jiang et al. 2014).

The former has the advantages that it is computationally cheap and straightforward to implement, and that it can give the desired halo mass resolution on the fly. However, throughout this thesis we utilise merger trees constructed from N -body simulations, as this allows the spatial distribution of galaxies, and hence their clustering, to be predicted directly (see Chap. 4). It also allows environmental effects to be studied. However, it comes with the drawback that, as the halo mass resolution of the simulation is fixed, a decreasing fraction of the mass in halos is followed with increasing redshift, which may have implications for predictions of integrated properties of the galaxy population e.g. the comoving cosmic star formation rate density,

²We note that the CPU time quoted for GALFORM does not include the time taken to run the initial dark matter only N -body simulation, which is typically much shorter than the time required to run a gas-dynamical simulation, and that both of these quoted times will have some dependence on various details of the codes, the machine used and the mass resolution of the simulations.

$\rho_{\text{SFR}}(z)$. Here we briefly describe the construction of these merger trees, and how the halo properties relevant for galaxy formation are calculated. For more details see Merson et al. (2013), Jiang et al. (2014) and Lacey et al. (2016).

An N -body simulation is performed from high redshift (e.g. $z = 127$) to the present day with particle data being output at pre-specified ‘snapshots’ during the calculation. The initial conditions of the simulation are set by the desired cosmological parameters, halo mass resolution and cosmological volume. Groups of dark matter particles are identified in the simulation snapshots using the *Friends-of-Friends* algorithm (FOF; Davis et al. 1985). The SUBFIND algorithm (Springel et al. 2001) is then used to identify self-bound locally over-dense sub-groups within the FOF groups. Descendant sub-groups are then identified in subsequent snapshots as the sub-groups containing the largest fraction of the most bound particles from a sub-group in the previous snapshot.

It is possible for the SUBFIND algorithm to temporarily ‘lose’ a sub-group. This can happen if a small sub-group temporarily falls below the resolution limit (usually set to 20 dark matter particles) or if it passes close to the centre of its host halo. Whilst this is not necessarily a problem for the construction of halo merger trees, it is problematic for codes such as GALFORM that require the descendent of a sub-group to be present at the next snapshot, and mass conservation. To avoid this, if a sub-group is identified at some snapshot, i , but lost at snapshot $i + 1$, then ‘orphaned’ sub-groups (sub-groups with no immediate progenitors) are investigated at subsequent snapshots ($i + 2$, $i + 3$...). These orphans are then identified as the descendants of the ‘lost’ sub-group at snapshot i if they contain some fraction of its most bound particles and no other progenitors are identified. Once this is done, interpolated sub-groups are inserted at each snapshot for which a sub-group was missing.

The sub-groups are then organised into a hierarchy of halos within their FOF group i.e. halos, sub-halos, sub-sub-halos etc. depending on their mass and if they fall within the half-mass radius of a more massive sub-group. There is also a condition of mass loss such that when a halo falls within another, more massive halo, the two will only be considered to have merged (i.e. the less massive halo becomes a sub-halo) once the smaller halo has lost $\sim 25\%$ of its infall mass. For the GALFORM calculation it is also assumed that once a halo merges within another and becomes a sub-halo, it is subsequently always defined as a sub-halo even if its orbit puts it outside of the virial radius of its host at some later time. In GALFORM it is assumed that once a central galaxy becomes a satellite in a larger host halo, all of its hot gas is stripped away, which can only be done once. These halo fragmentation episodes are identified by searching for satellite sub-groups which split off from their host halo to become an independent halo at a subsequent snapshot. In this case the sub-halo is re-merged. Once these post-processing steps are carried out the merger tree is specified.

When halos form they are assumed to have a virial radius

$$r_{\text{vir}} = [3M_{\text{halo}}/4\pi\Delta_{\text{vir}}\bar{\rho}]^{1/3}, \quad (2.2.1)$$

where M_{halo} is the halo mass (the sum of the mass of all the sub-groups + diffuse mass), $\bar{\rho}$ is the mean cosmological density at that redshift and the overdensity Δ_{vir} is

calculated from the spherical top-hat collapse model (e.g. Eke et al. 1996). The dark matter halos are assumed to have a mass density profile of the Navarro, Frenk and White (NFW) form (Navarro et al. 1997)

$$\rho_{\text{DM}}(r) \propto \frac{1}{(r/r_s)(1 + r/r_s)^2}, \quad (2.2.2)$$

where r_s is the scale radius, related to the virial radius by the concentration parameter $c_{\text{NFW}} = r_{\text{vir}}/r_s$, which is calculated according to the prescription of Navarro et al. (1997). In the simulation halos can grow through merger events and accretion. A halo formation event is defined when a halo has grown by a factor of 2 since the previous such event. At this point, the density profiles are updated with a new value of c_{NFW} , and the virial velocity $V_{\text{vir}} = (G M_{\text{vir}}/r_{\text{vir}})^{1/2}$ is recalculated. These quantities are assumed to remain constant between each halo formation event.

Halos acquire angular momentum from tidal torques during their formation. At each halo formation event a value of the dimensionless spin parameter, λ_{H} , where

$$\lambda_{\text{H}} = \frac{J_{\text{halo}} |E_{\text{halo}}|^{1/2}}{G M_{\text{halo}}^{5/2}}, \quad (2.2.3)$$

is drawn from a lognormal distribution with a median $\lambda_{\text{H,med}} = 0.039$ and dispersion $\sigma_{\lambda_{\text{H}}} = 0.53$ in $\ln \lambda_{\text{H}}$ (e.g. Cole and Lacey 1996). Only the magnitude of the angular momentum is recorded.

2.3 Gas Cooling in Halos

In GALFORM, it is assumed that dark matter halos in the merger trees each accrete their ‘fair’ share of baryons i.e. $(\Omega_{\text{b}}/\Omega_{\text{M}})M_{\text{halo}}$, where M_{halo} is the total (dark matter + baryonic) mass of the halo. Gas falling into halos is shock heated and subsequently must dissipate its energy through radiative cooling in order to condense into a galaxy and form stars. It is assumed that upon infall into the dark matter halo gas is shock heated to the virial temperature

$$T_{\text{vir}} = (\mu m_{\text{H}}/2 k_{\text{B}}) V_{\text{vir}}^2, \quad (2.3.4)$$

where m_{H} is the mass of a hydrogen atom, μ is the mean molecular weight and $V_{\text{vir}} = (G M_{\text{vir}}/r_{\text{vir}})^{1/2}$ is the virial velocity i.e. circular velocity at the virial radius. The gas is then assumed to settle into a spherically symmetric density distribution given by

$$\rho_{\text{hot}}(r) \propto \frac{1}{r_{\text{c}}^2 + r^2}, \quad (2.3.5)$$

with a core radius $r_c = 0.1r_{\text{vir}}$. The thermal energy per unit volume of the gas at radius, r , is given by

$$U_{\text{hot}}(r) = \frac{3}{2} \frac{k_B T_{\text{vir}}}{\mu m_H} \rho_{\text{hot}}(r). \quad (2.3.6)$$

This hot gas then loses its thermal energy through atomic processes, such that its cooling luminosity per unit volume is

$$\mathcal{L}_{\text{cool}}(r) = \rho_{\text{hot}}^2(r) \Lambda(T_{\text{vir}}, Z_{\text{hot}}), \quad (2.3.7)$$

where $\Lambda(T, Z)$ is the temperature and metallicity dependent cooling function tabulated by Sutherland and Dopita (1993). The local cooling rate, $\tau_{\text{cool}}(r)$, is simply the ratio of the two such that

$$\tau_{\text{cool}}(r) = \frac{3}{2} \frac{k_B}{\mu m_H} \frac{T_{\text{vir}}}{\rho_{\text{hot}}(r) \Lambda(T_{\text{vir}}, Z_{\text{hot}})}. \quad (2.3.8)$$

As the cooling time is inversely proportional to the density of the gas, denser gas at smaller radii will cool faster than that at larger radii. It is therefore possible to define a cooling radius r_{cool} as the radius at which the cooling time of the gas is equal to the age of the halo i.e. $\tau_{\text{cool}}(r_{\text{cool}}) = t - t_{\text{form}}$. This will propagate outwards with time and any gas at $r < r_{\text{cool}}$ is assumed to have cooled.

However, this in itself is not sufficient for the formation of a galactic disc as the cooled gas has to have had time to fall to the centre of the dark matter potential well. For a given mass distribution the free-fall time for a test particle to fall to the centre of the potential well, t_{ff} , is given by

$$t_{\text{ff}}(r) = \int_0^r \left[\int_r^{r'} -\frac{G M(r')}{r'^2} dr' \right]^{-1/2} dr''. \quad (2.3.9)$$

This can be inverted to find the free-fall radius, r_{ff} , such that $t_{\text{ff}}(r_{\text{ff}}) = t - t_{\text{form}}$, where a particle at $r < r_{\text{ff}}$ will have had sufficient time to fall to the centre of the potential well.

In GALFORM, the radius relevant for whether gas has become available to be accreted onto a galactic disc is defined by the minimum of these two radii i.e. $r_{\text{acc}} = \min[r_{\text{ff}}, r_{\text{cool}}]$. Gas needs to have had both sufficient time to cool and sufficient time to have fallen to the centre of the halo potential well for it to have been accreted onto the galactic disc. In a given time step in the calculation, r_{acc} is calculated at the beginning and end of the time step i.e. at t_{i-1} and t_i . The amount of gas that is said to have cooled, $\dot{M}_{\text{acc}} \Delta t$, is set to be equal to the mass of gas in the spherical shell defined by $\Delta r_{\text{acc}} = r_{\text{acc}}(t_i) - r_{\text{acc}}(t_{i-1})$. This defines the accretion rate of hot gas onto the cold gas disc, $\dot{M}_{\text{acc}}$ which is used later, and can be written as

$$\dot{M}_{\text{acc}} = 4\pi \rho_{\text{hot}}(r_{\text{acc}}) r_{\text{acc}}^2 \dot{r}_{\text{acc}}. \quad (2.3.10)$$

It is assumed that throughout the cooling process the angular momentum of the gas is conserved, such that it forms a disc. The rotational velocity of the hot gas, V_{rot} , is assumed to be independent of radius and is given by

$$V_{\text{rot}} = A(c_{\text{NFW}})\lambda_{\text{H}} V_{\text{vir}}, \quad (2.3.11)$$

where $A(c_{\text{NFW}})$ is a dimensionless coefficient with a weak dependence on the NFW concentration, determined such that the total specific angular momentum of the gas is in agreement with that of the dark matter, which is determined by the spin parameter λ_{H} (Cole et al. 2000). This results in a change in angular momentum of the cold gas due to accretion from the hot halo gas given by

$$\dot{j}_{\text{cold}} = \pi^2 V_{\text{rot}} \rho_{\text{hot}}(r_{\text{acc}}) r_{\text{acc}}^3 \dot{r}_{\text{acc}}. \quad (2.3.12)$$

It is assumed that the angular momentum of the disc is aligned with that of the halo.

2.4 Star Formation

Once the gas has cooled and been accreted onto the disc it becomes available for star formation. GALFORM has two modes of star formation, one that occurs in the galactic disc and one in a starburst, in a newly formed galactic bulge/spheroid.

2.4.1 Star Formation in the Disc

In the disc the SFR is calculated according to the empirical Blitz and Rosolowsky (2006) relation, based on the fraction of molecular gas in the disc, f_{mol} , which depends on the midplane gas pressure, P , at each radius in the disc

$$R_{\text{mol}} = \frac{\Sigma_{\text{mol}}}{\Sigma_{\text{atom}}} = \left(\frac{P}{P_0} \right)^{\alpha_{\text{P}}}, \quad (2.4.13)$$

where R_{mol} is the ratio of molecular to atomic gas, $\alpha_{\text{P}} = 0.8$ and $P_0/k_{\text{B}} = 1.7 \times 10^4 \text{ cm}^{-3} \text{ K}$ based on the observations of Leroy et al. (2008). The SFR is then assumed to be proportional to the mass of molecular gas, integrated over the whole disc, such that

$$\psi_{\text{disc}} = \nu_{\text{SF, quies}} M_{\text{mol, disc}} = \nu_{\text{SF, quies}} f_{\text{mol}} M_{\text{cold, disc}}, \quad (2.4.14)$$

where $f_{\text{mol}} = R_{\text{mol}}/(1 + R_{\text{mol}})$ and $\nu_{\text{SF, quies}} = 0.74 \text{ Gyr}^{-1}$ (Bigiel et al. 2011). This star formation prescription was first incorporated into GALFORM by Lagos et al. (2011).

2.4.2 Star Formation in Bursts

For star formation in bursts (which occur in a forming bulge/spheroid) it is assumed that $f_{\text{mol}} \approx 1$ and the SFR depends on the dynamical timescale of the bulge

$$\psi_{\text{burst}} = \nu_{\text{SF,burst}} M_{\text{cold,burst}}, \quad (2.4.15)$$

where $\nu_{\text{SF,burst}} = 1/\tau_{\star,\text{burst}}$ and

$$\tau_{\star,\text{burst}} = \max[f_{\text{dyn}} \tau_{\text{dyn,bulge}}, \tau_{\text{burst,min}}]. \quad (2.4.16)$$

Here $\tau_{\text{dyn,bulge}} = r_{\text{bulge}}/V_c(r_{\text{bulge}})$, where r_{bulge} is the half-mass radius of the bulge, and f_{dyn} and $\tau_{\text{burst,min}}$ are model parameters, such that for large dynamical times the SFR timescale scales with the dynamical timescale of the bulge ($f_{\text{dyn}} = 20$), but has a floor value ($\tau_{\text{burst,min}} = 100$ Myr) when the dynamical timescale of the bulge is short. These parameters are primarily constrained by the observed rest-frame far-UV (1500 Å) luminosity function.

2.5 Feedback Mechanisms

In GALFORM there are three channels of feedback, through which star formation is inhibited.

2.5.1 Supernova Feedback

Supernova explosions inject energy into the ISM and in doing so can remove gas from a galaxy. The energy injection is dominated by massive short-lived stars and so is approximately proportional to the SFR

$$\dot{M}_{\text{eject}} = \beta(V_c) \psi. \quad (2.5.17)$$

The ‘mass loading’ factor $\beta(V_c)$ is parametrised as

$$\beta(V_c) = (V_c/V_{\text{SN}})^{-\gamma_{\text{SN}}}, \quad (2.5.18)$$

where V_{SN} and γ_{SN} are adjustable parameters and V_c is the circular velocity of the disc (or of the bulge for starbursts) at the half-mass radius. The rationale behind this parametrisation is that for the same SFR (i.e. energy input) less mass should be ejected from deeper potential wells, for which V_c is a proxy. The removal of gas in this manner is assumed to not affect the specific angular momentum of the disc.

The ejected gas accumulates in the ejected mass reservoir of mass M_{res} beyond the virial radius, from which it is returned to the hot gas reservoir within the virial radius, according to

$$\dot{M}_{\text{return}} = \alpha_{\text{ret}} \frac{M_{\text{res}}}{\tau_{\text{dyn,halo}}}, \quad (2.5.19)$$

where $\tau_{\text{dyn,halo}} = r_{\text{vir}}/V_{\text{vir}}$ is the halo dynamical time, as first implemented into GALFORM in Bower et al. (2006). The parameters in this supernova feedback model are constrained primarily by requirement of the model to reproduce the faint end of the $z = 0$ galaxy luminosity function, values of $\gamma_{\text{SN}} = 3.2$, $V_{\text{SN}} = 320 \text{ km s}^{-1}$ and $\alpha_{\text{ret}} = 0.64$ are adopted in Lacey et al. (2016).

In Chap. 6 we discuss results from a variant supernova feedback model (Hou et al. 2016), in which a break is introduced into the parametrisation of the mass loading factor and its normalisation is allowed to evolve with redshift.

2.5.2 Photoionisation Heating of the IGM

The IGM is reionized and photo-heated by ionising photons produced by stars and AGN. This inhibits star formation in two ways: (i) the increased IGM pressure inhibits the collapse of gas into dark matter halos and (ii) continued photo-heating of gas inside halos by the ionizing UV background inhibits the cooling of gas. This was implemented into GALFORM by to Benson et al. (2003), assuming that reionization occurs instantaneously at z_{reion} . After this no cooling of gas occurs within halos with $V_c < V_{\text{crit}}$. A value of $z_{\text{reion}} = 10$ (e.g. Dunkley et al. 2009) is used and $V_{\text{crit}} = 30 \text{ km s}^{-1}$, based on gas-dynamical simulations (e.g. Okamoto et al. 2008). Despite its simplicity this model has been shown to reproduce more detailed treatments quite well (Font et al. 2011).

2.5.3 Supermassive Black Hole Growth and AGN Feedback

The accretion of material onto supermassive black holes (SMBHs) causes huge amounts of energy to be released from the accreting material. In GALFORM SMBHs can grow via three different channels: (i) accretion of gas during a starburst, the cold gas mass accreted onto the black hole is some fraction f_{BH} of the mass of stars formed during the burst; (ii) accretion of gas from the hot halo; (iii) black hole mergers, in which the BH merger happens at the same time as the galaxy merger and mass loss through radiation of energy via gravitational waves is neglected. Black holes are not seeded in the model and are assumed to form after the first starburst event a galaxy experiences. This is a reasonable approximation as the seed mass of a black hole is expected to be a negligible contributor to its final mass (e.g. Malbon et al. 2007).

In GALFORM it is assumed that AGN feedback occurs in a *hot-halo mode*. The energy released by direct accretion of hot gas from the halo powers relativistic jets which deposit thermal energy into the hot gas. An equilibrium is reached such that this input of thermal energy balances the energy losses due to radiative cooling. This occurs as long as: (1) the gas is quasi-hydrostatic, that is that the cooling time is sufficiently long compared to the free-fall time at the cooling radius

$$\tau_{\text{cool}}(r_{\text{cool}})/\tau_{\text{ff}}(r_{\text{cool}}) > 1/\alpha_{\text{cool}}, \quad (2.5.20)$$

where $\alpha_{\text{cool}} \sim 1$ is an adjustable parameter; and (2) the AGN power required to balance the radiative cooling luminosity is below some fraction f_{Edd} of the Eddington luminosity of the SMBH that has mass M_{BH}

$$L_{\text{cool}} < f_{\text{Edd}} L_{\text{Edd}}(M_{\text{BH}}), \quad (2.5.21)$$

where the total cooling luminosity of the gas, L_{cool} , is given by

$$L_{\text{cool}} = \int_{r_{\text{cool}}}^{r_{\text{vir}}} \rho_{\text{hot}}^2(r) \Lambda(T_{\text{vir}}, Z_{\text{hot}}) 4\pi r^2 dr. \quad (2.5.22)$$

When these criteria are met the black hole is assumed to grow at a rate such that the energy losses of the hot gas through radiative cooling are exactly balanced $\epsilon_{\text{heat}} \dot{M}_{\text{BHC}}^2 = L_{\text{cool}}$, where ϵ_{heat} is an adjustable parameter. This means the cooling rate has effectively been set to zero i.e. $\tau_{\text{cool}} \rightarrow \infty$. This prescription for AGN feedback was first implemented into GALFORM by Bower et al. (2006). Values of $f_{\text{Edd}} = 0.01$ and $\epsilon_{\text{heat}} = 0.02$ are adopted in Lacey et al. (2016), as determined by Fanidakis et al. (in preparation) who used observed X-ray luminosity functions and the black hole mass—bulge mass relation to calibrate these parameters.

2.6 Galaxy Mergers and Disc Instabilities

These dynamical processes are the channels of spheroid growth, and thus morphological transformation, in the model.

2.6.1 Galaxy Mergers

Galaxies are classified as *central* galaxies, which sit at the centre of the dark matter potential and *satellite* galaxies which orbit within the dark matter halo. When dark matter halos merge, the central galaxy of the most massive progenitor halo becomes the new central galaxy, while all other galaxies in the new halo are classed as satellites. Once a galaxy becomes a satellite it is instantaneously stripped of all its hot gas

(which is added to the hot gas reservoir of the central galaxy) through ram-pressure stripping and it is assumed that no further gas can cool onto it. Satellite galaxies are also subject to *dynamical friction* (e.g. Chandrasekhar 1943; Binney and Tremaine 1987), the timescale for which is an expression modified from the Chandrasekhar dynamical friction formula to account for the tidal stripping of dark matter from the satellite's halo, obtained through fitting to the results of N -body simulations (Jiang et al. 2008).

$$\tau_{\text{merge}} = \frac{f(\epsilon)}{2C} \frac{M_{\text{cen}}}{M_{\text{sat}}} \frac{1}{\ln(1 + M_{\text{cen}}/M_{\text{sat}})} \left(\frac{r_{\text{circ}}}{r_{\text{vir}}} \right)^{1/2} \tau_{\text{dyn, halo}}. \quad (2.6.23)$$

Here M_{cen} and M_{sat} are the total dynamical mass, i.e. baryons + dark matter, of the central and satellite galaxies at infall, and ϵ is the circularity of the orbit at infall,³ and r_{circ} is the radius of this equivalent circular orbit. The constant C , and parameters in the function $f(\epsilon) = a\epsilon^b + c$ are fit by Jiang et al. (2008). A random value for ϵ is drawn for each satellite from the probability distribution for infalling satellite halos as measured by Benson (2005) from cosmological N -body simulations.

The fate of the gas and stellar content following a galaxy merger depends on the ratio of the baryonic mass (cold gas + stars), $M_{\text{b, sat}}$ of the satellite to that of the central $M_{\text{b, cen}}$. Mergers with $M_{\text{b, sat}}/M_{\text{b, cen}} > f_{\text{ellip}}$ are classed as major mergers. It is assumed that the stellar discs are destroyed and all of the stars in both galaxies form a new stellar spheroid, whilst all of the cold gas collapses into the spheroid. Other mergers are classed as minor. In minor mergers, stars from the satellite galaxy are added to the spheroid of the central galaxy, but cold gas is added to the gaseous disc of the central, without changing its specific angular momentum.

Additionally, mergers with $M_{\text{b, sat}}/M_{\text{b, cen}} > f_{\text{burst}}$ trigger starbursts, in which all of the cold gas from the merging galaxies is transferred to the spheroid. The parameters f_{ellip} and f_{burst} satisfy $f_{\text{burst}} \leq f_{\text{ellip}} < 1$. Numerical simulations of isolated mergers suggest $f_{\text{ellip}} \approx 0.3$ and $f_{\text{burst}} \approx 0.1$ (e.g. Mihos and Hernquist 1994; Barnes 1998; Hopkins et al. 2009).

2.6.2 Disc Instabilities

If a disc becomes sufficiently massive that its self-gravity is dominant, then it would be unstable to small perturbations by minor satellites or dark matter substructures.

Analytical arguments based on idealized descriptions of rotating discs can show that perturbations to the surface density of the disc can generate a mode that leads

³This is defined as the ratio of the orbital angular momentum to that of a circular orbit of the same energy in the same potential.

to a rotating, elliptical deformation of the whole disc⁴ (e.g. Kalnajs 1972; Binney and Tremaine 1987), referred to as the bar mode due to the similarity between this deformation and bars observed in disc galaxies. This can be thought of generally in terms of the ratio of kinetic energy in random motion (Π , which acts to stabilise the disc) to rotational kinetic energy, K . For example, Ostriker and Peebles (1973) find that their simulated discs were stable to bar formation if $\Pi/K \gtrsim 5$.

Based on the N -body simulations of Efstathiou et al. (1982), in GALFORM it is assumed that discs are dynamically unstable to bar formation (i.e. perturbations would result in the growth of a bar mode) if

$$F_{\text{disc}} \equiv \frac{V_c(r_{\text{disc}})}{(1.68 G M_{\text{disc}}/r_{\text{disc}})^{1/2}} < F_{\text{stab}}, \quad (2.6.24)$$

where M_{disc} is the total disc mass (stars + gas), r_{disc} is the disc half-mass radius and F_{stab} is a parameter approximately equal to unity. Efstathiou et al. derived $F_{\text{stab}} \approx 1.1$ for purely stellar discs embedded within a variety of dark matter halos,⁵ whereas Christodoulou et al. (1995) found that the stability criterion for purely gaseous discs was a bit lower, $F_{\text{stab}} \approx 0.9$. If $F_{\text{disc}} < F_{\text{stab}}$ then it is assumed that the disc forms a bar which subsequently evolves into a spheroid (e.g. Combes 2000). This condition is tested at every timestep in the calculation, as cooled gas is continually being accreted (in the absence of AGN feedback) onto the disc from the hot halo component, and expelled by supernova feedback. It is assumed that the evolution from bar to spheroid occurs instantaneously (though in reality it is likely that this will take at least several disc dynamical times), and that a disc instability triggers a starburst.

2.7 Galaxy Sizes

The size of a disc is basically determined by the angular momentum of the halo gas from which it formed. In GALFORM, the size of the disc is solved for by assuming conservation of angular momentum and centrifugal equilibrium.

As the gas condenses to form a galaxy at the centre of a dark matter halo, it will exert a gravitational force on the dark matter halo, causing dark matter to contract. As long as this process occurs sufficiently slowly the adiabatic invariant, the pseudo-specific angular momentum of a particle in the halo, $r V_H(r)$, will be conserved. Applying this result to spherical shells of matter the final and initial radii of the shell, r , and, r_0 , respectively can be related by

$$r_0 V_{\text{cH}}^0(r_0) = r V_{\text{cH}}(r), \quad (2.7.25)$$

⁴Note that here we are concerned with the global stability of the disc i.e. perturbations with wavelengths comparable to the size of the disc, rather than local (in)stability where the wavelength of perturbation is much smaller than the disc and the Toomre (1964) criterion can be derived.

⁵This is important as the halo helps to stabilise the disc.

where $V_{\text{cH}}(r)$ and $V_{\text{cH}}^0(r_0)$ are the final and initial halo circular velocities. The circular velocity of the halo is then

$$V_{\text{cH}}^2(r) = G[M_{\text{H}}(r) + M_{\text{D}}(r) + M_{\text{B}}(r)]/r, \quad (2.7.26)$$

where $M_{\text{D}}(r)$ and $M_{\text{B}}(r)$ are the masses of the disc and the bulge within radius r respectively. Combining the above equations gives

$$r_0 M_{\text{H}}^0(r_0) = r[f_{\text{H}} M_{\text{H}}^0(r_0) + M_{\text{D}}(r) + M_{\text{B}}(r)], \quad (2.7.27)$$

where $1 - f_{\text{H}}$ is the fraction of the total halo mass that forms the central galaxy. This relates, for the relevant mass profiles, the final radius of a halo shell to its initial radius.

The disc is assumed to have an exponential surface density profile

$$\Sigma_{\text{D}}(r) = \frac{M_{\text{D}}}{2\pi h_{\text{D}}} \exp(-r/h_{\text{D}}), \quad (2.7.28)$$

where h_{D} is the radial scale length of the disc, related to the disc half-mass radius, r_{D} , by

$$2(r_{\text{D}}/h_{\text{D}} + 1) \exp(-r_{\text{D}}/h_{\text{D}}) = 1, \quad (2.7.29)$$

and M_{D} is the mass of the (gas + stars) in the disc. The specific angular momentum of the disc, j_{D} (see Eq. 2.3.12), is given by the integral

$$j_{\text{D}} = \int_0^\infty 2\pi r \Sigma_{\text{D}}(r) r V_{\text{cD}} dr / M_{\text{D}} = k_{\text{D}} r_{\text{D}} V_{\text{cD}}(r_{\text{D}}) \quad (2.7.30)$$

where

$$k_{\text{D}} = \int_0^\infty \frac{r^2}{h_{\text{D}}^2} \exp\left(-\frac{r}{h_{\text{D}}}\right) \frac{V_{\text{cD}}(r)}{1.68 V_{\text{cD}}(r_{\text{D}})} \frac{dr}{h_{\text{D}}}, \quad (2.7.31)$$

and $V_{\text{cD}}(r)$ is the circular velocity of the disc. A flat rotation curve [i.e. $V_{\text{cD}}(r) = V_{\text{cD}}(r_{\text{D}})$] gives a value of $k_{\text{D}} = 1.19$. Using the relation $V_{\text{cD}}^2(r_{\text{D}}) = G M(r_{\text{D}})/r_{\text{D}}$ we can write

$$\begin{aligned} j_{\text{D}}^2 &= k_{\text{D}}^2 r_{\text{D}}^2 V_{\text{cD}}^2(r_{\text{D}}) \\ &= k_{\text{D}}^2 G r_{\text{D}} \left[f_{\text{H}} M_{\text{H}}^0(r_{\text{D},0}) + \frac{1}{2} k_{\text{H}} M_{\text{D}} + M_{\text{B}}(r_{\text{D}}) \right], \end{aligned} \quad (2.7.32)$$

where the relation $M_{\text{H}}(r) = f_{\text{H}} M_{\text{H}}^0(r_0)$ has been used, and by definition $M_{\text{D}}(r_{\text{D}}) = M_{\text{D}}/2$. The constant k_{H} accounts for the fact that the disc is not spherically symmetric. For an exponential disc $k_{\text{H}} = 1.25$.

This can be equated to Eq. (2.7.27) at the half-disc radius to give

$$r_{D,0} M_H^0(r_{D,0}) = \left[f_H M_H^0(r_{D,0}) + \frac{1}{2} M_D + M_B(r_D) \right], \quad (2.7.33)$$

which can be solved to find r_D .

A similar procedure is followed to determine the size of the spheroid. Bulges are assumed to have an $r^{1/4}$ mass density profile in projection

$$\Sigma_B(r) = \Sigma_0 \exp \left[\beta_n \left\{ \left(\frac{r}{r_B} \right)^{1/n} - 1 \right\} \right], \quad (2.7.34)$$

for $n = 4$, where r_B is the half-mass radius of the bulge and β_n is well approximated by $\beta_n = 2n - 0.324$.

2.7.1 Spheroid Formation

GALFORM assumes that a spheroid is formed after a merger or disc instability (see Sect. 2.6). In the case of a merger galaxies are assumed to merge when their separation equals the sum of their half-mass radii. Assuming each galaxy is in equilibrium their gravitational self-binding energy is

$$E_{\text{bind}} = -\frac{c_{\text{bind}}}{2} \frac{G M^2}{r}, \quad (2.7.35)$$

where M is the total (dark matter + stars + gas) mass, r is the half-mass radius and c_{bind} is a constant that depends on the distribution of the mass. For an $r^{1/4}$ -law $c_{\text{bind}} = 0.45$ whereas for an exponential disc $c_{\text{bind}} = 0.49$. For simplicity in GALFORM it is assumed that $c_{\text{bind}} = 0.5$. The orbital energy of two galaxies just prior to a merger is given by

$$E_{\text{orbit}} = -\frac{f_{\text{orbit}}}{2} \frac{G M_1 M_2}{r_1 + r_2}, \quad (2.7.36)$$

where M_1 and M_2 are the total (i.e. dark matter + baryons) masses of the galaxies and r_1 and r_2 are their half-mass radii. The parameter f_{orbit} depends on the orbital parameters of the galaxy pair. A value of $f_{\text{orbit}} = 1$ corresponds to point masses on circular orbits, however in Lacey et al. (2016) a value of $f_{\text{orbit}} = 0$, which corresponds to a marginally (un)bound orbit (or parabolic trajectory), is adopted. This has the effect of making the total binding energy of the new spheroid (see equation below) less negative, resulting in the spheroid sizes being slightly larger. This value for f_{orbit} is chosen to give an improved match to the observed size-luminosity relation for early-type galaxies. Conservation of energy then requires

$$E_{\text{bind,new}} = E_{\text{bind,1}} + E_{\text{bind,2}} + E_{\text{orbit}}, \quad (2.7.37)$$

which can be solved to find the half-mass radius of the new spheroid, r_{new} . This can then be used to compute the angular momentum of the bulge/spheroid which in turn can be used to compute its size r_{B} . For a minor merger M_1 and r_1 are replaced by the mass and half-mass radius of the primary spheroid. A similar expression is used to compute r_{new} in the case of disc instabilities

$$\frac{c_{\text{B}}(M_{\text{disc}} + M_{\text{bulge}})^2}{r_{\text{new}}} = \frac{c_{\text{B}}M_{\text{bulge}}^2}{r_{\text{bulge}}} + \frac{c_{\text{D}}M_{\text{disc}}^2}{r_{\text{disc}}} + f_{\text{int}} \frac{M_{\text{bulge}}M_{\text{disc}}}{r_{\text{bulge}} + r_{\text{disc}}}, \quad (2.7.38)$$

though here the mass refers only to the baryonic (i.e. stars + gas) mass of the galaxy. The final term on the right accounts for the gravitational interaction energy between the disc and the bulge, which is well approximated by $f_{\text{int}} = 2$.

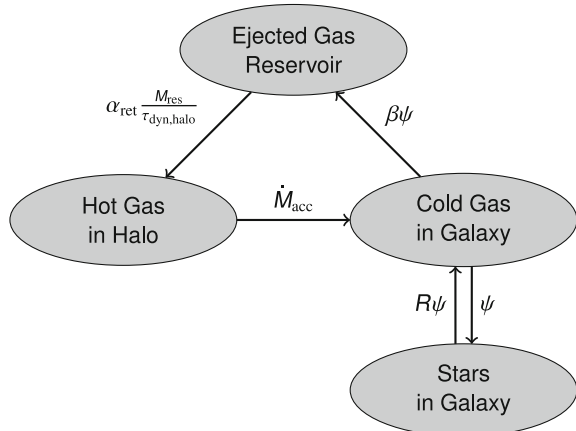
2.8 Chemical Evolution

The interplay of the equations describing gas cooling, star formation and feedback discussed up until results in a set of coupled differential equations which track the exchange of mass and metals between the main baryonic reservoirs of a galaxy: hot halo gas (M_{hot}), cold disc gas (M_{cold}), stars ($M_{\star}$) and ejected gas (M_{res}). This is summarised in Fig. 2.1.

2.8.1 Evolution of Mass and Metals

The equations governing the exchange of mass between a galaxy's baryonic reservoirs are

Fig. 2.1 Schematic of the exchange of mass between a galaxy's baryonic reservoirs in GALFORM



$$\dot{M}_{\text{hot}} = -\dot{M}_{\text{acc}} + \alpha_{\text{ret}} \frac{M_{\text{res}}}{\tau_{\text{dyn, halo}}}, \quad (2.8.39)$$

$$\dot{M}_{\text{cold}} = \dot{M}_{\text{acc}} - (1 - R + \beta)\psi, \quad (2.8.40)$$

$$\dot{M}_{\star} = (1 - R)\psi, \quad (2.8.41)$$

$$\dot{M}_{\text{res}} = \beta\psi - \alpha_{\text{ret}} \frac{M_{\text{res}}}{\tau_{\text{dyn, halo}}}, \quad (2.8.42)$$

and for metals

$$\dot{M}_{\text{hot}}^Z = -Z_{\text{hot}}\dot{M}_{\text{acc}} + \alpha_{\text{ret}} \frac{M_{\text{res}}^Z}{\tau_{\text{dyn, halo}}}, \quad (2.8.43)$$

$$\dot{M}_{\text{cold}}^Z = Z_{\text{hot}}\dot{M}_{\text{acc}} + [p - (1 - R + \beta)Z_{\text{cold}}]\psi, \quad (2.8.44)$$

$$\dot{M}_{\star}^Z = (1 - R)Z_{\text{cold}}\psi, \quad (2.8.45)$$

$$\dot{M}_{\text{res}}^Z = \beta Z_{\text{cold}}\psi - \alpha_{\text{ret}} \frac{M_{\text{res}}^Z}{\tau_{\text{dyn, halo}}}, \quad (2.8.46)$$

where the recycled fraction, R , is the fraction of the initial mass of a stellar population returned to the ISM by mass loss, and the yield, p , is the fraction of the initial mass of a stellar population synthesised into new metals and then ejected into the ISM. These quantities are determined by the choice of IMF, as described below.

2.8.2 The Initial Mass Function

The IMF describes the distribution in mass, m , of stars formed in a simple stellar population, normalized to unit solar mass

$$\int_{m_1}^{m_2} m \Phi(m) \, d \ln m = 1, \quad (2.8.47)$$

such that $\Phi(m) d \ln m$ is the number of stars formed between $[m, m + dm]$ per unit mass of stars formed, where m_1 and m_2 are the mass limits of the IMF. The recycled fraction, R , and yield, p , described above are set by the IMF. The recycled fraction is the fraction of the initial mass of a stellar population that is returned to the ISM by mass loss from dying stars. In the instantaneous recycling approximation⁶ it is given by the integral

$$R = \int_{1M_{\odot}}^{m_2} [m - m_{\text{rem}}(m)] \Phi(m) \, d \ln m, \quad (2.8.48)$$

⁶Instantaneous recycling assumes that all stars with $m > 1 M_{\odot}$ die immediately, whilst all other stars live forever. This is a reasonable approximation as the lifetimes of massive stars are short compared to the timescales over which galaxies evolve. The effect of relaxing this approximation in GALFORM is discussed in e.g. Nagashima et al. (2005).

where $m_{\text{rem}}(m)$ is the mass of the remnant (e.g. white dwarf, neutron star, black hole) left by a star of initial mass m , obtained from stellar evolution calculations (e.g. Marigo et al. 1996; Portinari et al. 1998). The yield, p , is the fraction of initial mass of a stellar population that is synthesised into new metals and then ejected. It is given by

$$p = \int_{1M_{\odot}}^{m_2} p_Z(m) \Phi(m) d \ln m, \quad (2.8.49)$$

where $p_Z(m)$ is the fraction for a single star of initial mass m , also obtained from stellar evolution calculations. It is assumed that metals are instantaneously mixed into the cold gas component.

Due to the assumption of instantaneous recycling, time delays between the star formation event and enrichment of the ISM are ignored. We now briefly outline an example of when such time delays can be useful, in the context of Type I and Type II supernovae (SNe). Type II SNe are the result of the collapse of massive ($\gtrsim 8 M_{\odot}$) stars, and have lifetimes of $\lesssim 10^7$ years, whereas Type I SNe are the result of lower mass ($\sim 3 - 8 M_{\odot}$) stars and have longer lifetimes of 10^8 yr. Type I SNe occur if the star in question has a binary companion. Once the star has evolved into a $\sim 0.6 M_{\odot}$ white dwarf, subsequent accretion of matter from the companion leads to a supernova when the white dwarf reaches the Chandrasekhar mass limit of $\sim 1.4 M_{\odot}$ (such that the electron degeneracy pressure supporting the white dwarf is overcome), it is this constant mass limit that leads to the use of Type I SNe as a ‘standard candle’.

Type I SNe are the main producers of iron (producing $\sim 5 - 10 \times$ more than a Type II), whereas both types can produce α elements (e.g. Ne, Mg, Si; those produced through the ‘alpha’ process which fuses Helium nuclei together). As the different supernovae occur on different timescales (there is a delay of $\sim 10^8$ yr between Type II and Type I) there is a corresponding delay between the enrichment of the ISM with Fe and α elements. Thus, the abundance ratio of $[\alpha/\text{Fe}]$ can be used as a diagnostic of the star formation history of the galaxy, though $[\alpha/\text{Fe}]$ is also affected by the assumed IMF.

In GALFORM, as Type I SNe make only a small contribution to the net production of heavy elements the metal yields are calculated assuming only Type II SNe, though see Nagashima et al. (2005) for an example of Type I SNe being included in GALFORM metal enrichment. The Portinari et al. yields used here incorporate the Type II SNe calculations of Woosley and Weaver (1995).

In GALFORM it is assumed that the IMF is a power law or piecewise power law

$$\Phi(m) = \frac{dN}{d \ln m} \propto m^{-x}, \quad (2.8.50)$$

where, for example, $x = 1.35$ for a Salpeter (1955) IMF. In GALFORM, for quiescent/disc star formation, a Kennicutt (1983) IMF is assumed i.e. $x = 0.4$ for $m < M_{\odot}$ and $x = 1.5$ for $m > M_{\odot}$. For burst star formation x is treated as an adjustable parameter. This is discussed in more detail in Sect. 2.10.

2.9 The Spectral Energy Distribution of Galaxies

The star formation and metal enrichment history of a galaxy, $\Psi(t, Z)$, predicted by GALFORM is combined with models for stellar evolutionary population synthesis (e.g. Tinsley 1972; Bruzual and Charlot 2003; Maraston 2005) and a simple model for the absorption and re-emission of stellar radiation by interstellar dust to produce the spectral energy distributions of model galaxies. Importantly for the dust emission, the temperature of the interstellar dust is not a parameter in the model but is determined from the use of the equations of radiative transfer and global energy balance arguments.

The dust model presented here is motivated by the spectrophotometric code GRASIL (Silva et al. 1998). However, a number of simplifying approximations are made, such as each dust component being described by a single temperature and the dust being optically thin to its own emission. Despite its simplicity this model can reproduce the predictions of GRASIL to within a factor of ~ 2 for $\lambda_{\text{rest}} \gtrsim 70 \mu\text{m}$ (see Appendix C).

2.9.1 Emission of Light

The unattenuated spectral energy distribution (SED) of a galaxy at time t is a convolution

$$L_{\lambda}^{\text{unatten}}(t) = \sum_i \int_0^t dt' \int_0^{\infty} dZ L_{\lambda}^{\text{SSP}}(t', Z; \Phi_i) \Psi_i(t - t', Z), \quad (2.9.51)$$

where $\Psi_i(t, Z; \Phi_i)$ is the mass of stars formed with IMF Φ_i between $(t, t + dt)$ in time and $(Z, Z + dZ)$ in metallicity, summed over all the progenitors of the galaxy in question; and $L_{\lambda}^{\text{SSP}}(t, Z; \Phi_i)$ is the SED of a simple stellar population of unit solar mass, age t and metallicity Z , formed with IMF Φ_i . In turn, $L_{\lambda}^{\text{SSP}}(t, Z; \Phi_i)$ is related to the SED of its constituent stars by (e.g. Bruzual and Charlot 2003)

$$L_{\lambda}^{\text{SSP}}(t, Z; \Phi_i) = \int_{m_1}^{m_2} d \ln m L_{\lambda}^{\text{star}}(t, Z, m) \Phi_i(m). \quad (2.9.52)$$

The dust attenuated SED can be written as

$$L_{\lambda}^{\text{atten}}(t) = \sum_i A_{\lambda}^{\text{diff}} \int_0^t dt' \int_0^{\infty} dZ A_{\lambda}^{\text{MC}}(t') L_{\lambda}^{\text{SSP}}(t', Z; \Phi_i) \Psi_i(t - t', Z), \quad (2.9.53)$$

where $A_{\lambda}^{\text{diff}}$ and $A_{\lambda}^{\text{MC}}(t')$ describe the attenuation at wavelength λ due to the diffuse and molecular cloud dust components respectively. These are described in more detail in the following Section.

2.9.2 Dust Absorption

It is assumed that dust exists in two components: (i) dense molecular clouds of fixed gas surface density in which stars form, escaping on some timescale (t_{esc}) and; (ii) a diffuse ISM which assumes the same radial and vertical scale lengths for the dust and gas as for the stellar disc i.e. $h_{z,\text{dust}}/h_{z,\text{stars}} = h_{R,\text{dust}}/h_{R,\text{stars}} = 1$. The fraction of gas mass in molecular clouds is determined by the parameter f_{cloud} . In GALFORM, the dust mass is calculated assuming a dust-to-gas ratio that scales linearly with cold gas metallicity, Z_{cold} , normalized to the local ISM value (Silva et al. 1998) such that

$$M_{\text{dust}} = \delta_{\text{dust}} M_{\text{cold}} Z_{\text{cold}}, \quad (2.9.54)$$

where $\delta_{\text{dust}} = 0.334$. This value for δ_{dust} is derived assuming a value of 1/110 for the dust-to-hydrogen gas mass ratio at solar metallicity ($Z = 0.02$, e.g. Draine and Lee 1984) and a fixed hydrogen-to-total gas mass ratio of 0.735. However, in practise it is only the quantity $M_{\text{cold}} Z_{\text{cold}}$ that enters into the GALFORM calculation.

The dust is assumed to have the same extinction curve shape, k_{λ} , and albedo, a_{λ} , as for the Solar neighbourhood (Silva et al. 1998). The (extinction) optical depth of dust for light passing through gas with surface density Σ_{gas} is

$$\tau_{\lambda,\text{ext}} = 0.043 \left(\frac{k_{\lambda}}{k_V} \right) \left(\frac{\Sigma_{\text{gas}}}{M_{\odot} \text{ pc}^{-2}} \right) \left(\frac{Z_{\text{cold}}}{0.02} \right), \quad (2.9.55)$$

where the normalisation is based on the local ratio of V-band extinction, A_V , to hydrogen column density, N_{H} , $A_V/N_{\text{H}} = 3.3 \times 10^{-22} \text{ mag cm}^2$ measured by Savage and Mathis (1979). For discs $\Sigma_{\text{gas}} = (1 - f_{\text{cloud}}) M_{\text{cold,disc}} / 2\pi h_R^2$, where h_R is the radial scale length of the disc, for clouds $\Sigma_{\text{gas}} = 3m_{\text{cloud}} / 4\pi r_{\text{cloud}}^2$ where $m_{\text{cloud}} = 10^6 M_{\odot}$ and $r_{\text{cloud}} = 16 \text{ pc}$ are the assumed mass and radius of the clouds respectively. These values are motivated by those found by Silva et al. (1998, see their Table 2). These parameters enter into the calculation in the combination $m_{\text{cloud}}/r_{\text{cloud}}^2$ and only have a significant effect on results in the mid-IR (Vega et al. 2005), provided the clouds are still in the optically thick regime for UV/optical wavelengths.

In a cloud the effective absorption optical depth is approximated as

$$\tau_{\lambda,\text{eff}} = (1 - a_{\lambda})^{1/2} \tau_{\lambda,\text{ext}}, \quad (2.9.56)$$

where a_{λ} is the albedo. The dust attenuation factor in a single cloud is $e^{-\tau_{\lambda,\text{eff}}}$ while the mean attenuation for stars of age t due to clouds is given by

$$A_{\lambda}^{\text{MC}}(t) = 1 - \eta(t)(e^{-\tau_{\lambda,\text{eff}}} - 1), \quad (2.9.57)$$

where $\eta(t)$ is the fraction of stars of age t still in their birth cloud. This is parametrised as

$$\eta(t) = \begin{cases} 1 & t < t_{\text{esc}} \\ 2 - t/t_{\text{esc}} & t_{\text{esc}} < t < 2t_{\text{esc}} \\ 0 & t > 2t_{\text{esc}} \end{cases} \quad (2.9.58)$$

such that stars begin to migrate out of their birth cloud at time t_{esc} and have completely left by time $2t_{\text{esc}}$. This parametrisation is meant to describe the destruction of a giant molecular cloud due to powerful stellar winds and outflows produced by massive stars over their lifetime. In Lacey et al. (2016) values of $t_{\text{esc}} = 1$ Myr and $f_{\text{cloud}} = 0.5$ are adopted, constrained primarily by requiring the model to reproduce the observed high-redshift ($z \gtrsim 3$) rest-frame far-UV (1500 Å) luminosity function. This is discussed in more detail in Sect. 5.3.4.

The dust-attenuated SED (see Eq. 2.9.53) can then be rewritten as

$$L_{\lambda}^{\text{atten}}(t) = A_{\lambda}^{\text{diff}} \langle A_{\lambda}^{\text{MC}}(t) \rangle L_{\lambda}^{\text{unatten}}(t), \quad (2.9.59)$$

where $\langle A_{\lambda}^{\text{MC}}(t) \rangle$ is the dust attenuation by clouds averaged over stellar age given by

$$\langle A_{\lambda}^{\text{MC}}(t) \rangle = 1 - \langle \eta(t) \rangle (e^{-\tau_{\lambda, \text{eff}}} - 1), \quad (2.9.60)$$

where $\langle \eta(t) \rangle$ is given by a luminosity-weighted average in each band

$$\langle \eta(t) \rangle = \frac{\sum_i \int_0^t dt' \int_0^{\infty} dZ \eta(t') L_{\lambda}^{\text{SSP}}(t', Z; \Phi_i) \Psi_i(t - t', Z)}{\sum_i \int_0^t dt' \int_0^{\infty} dZ L_{\lambda}^{\text{SSP}}(t', Z; \Phi_i) \Psi_i(t - t', Z)}. \quad (2.9.61)$$

This is calculated separately for quiescent and burst star formation. For burst star formation an exponentially decaying star formation history is assumed. The energy absorbed by the molecular cloud component is thus given by

$$L_{\text{abs}}^{\text{MC}} = \int_0^{\infty} (1 - \langle A_{\lambda}^{\text{MC}} \rangle) L_{\lambda}^{\text{unatten}} d\lambda. \quad (2.9.62)$$

For the diffuse component $A_{\lambda}^{\text{diff}}$ is calculated from the tabulated radiative transfer results of Ferrara et al. (1999), which are derived from the code described in Bianchi et al. (1996). The central optical depth for the model galaxies are calculated according to Eq. (2.9.55) and the Ferrara et al. tables are interpolated to get the total attenuation as a function of wavelength. The energy absorbed by the diffuse dust medium is given by

$$L_{\text{abs}}^{\text{diff}} = \int_0^{\infty} (1 - A_{\lambda}^{\text{diff}}) \langle A_{\lambda}^{\text{MC}} \rangle L_{\lambda}^{\text{unatten}} d\lambda. \quad (2.9.63)$$

2.9.3 Dust Emission

The dust is assumed to be in thermal equilibrium and optically thin at the wavelengths at which the dust emits. We can therefore write the dust emission as

$$L_{\lambda}^{\text{dust}} = 4\pi\kappa_{\text{d}}(\lambda)B_{\lambda}(T_{\text{dust}})M_{\text{gas}}Z_{\text{gas}}. \quad (2.9.64)$$

Here, T_{d} is the dust temperature (a major simplifying assumption made in this model is to assume that each dust component can be described by a single temperature), $B_{\lambda}(T_{\text{d}})$ is the Planck blackbody function and $\kappa_{\text{d}}(\lambda)$ is the dust opacity per unit mass of metals in the gaseous phase. This is parametrised as

$$\kappa_{\text{d}}(\lambda) = \begin{cases} \kappa_1 \left(\frac{\lambda}{\lambda_1}\right)^{-2} & \lambda < \lambda_{\text{b}} \\ \kappa_1 \left(\frac{\lambda_{\text{b}}}{\lambda_1}\right)^{-2} \left(\frac{\lambda}{\lambda_{\text{b}}}\right)^{-\beta_{\text{b}}} & \lambda > \lambda_{\text{b}}, \end{cases} \quad (2.9.65)$$

with $\kappa_1 = 140 \text{ cm}^2 \text{ g}^{-1}$ at a reference wavelength of $\lambda_1 = 30 \text{ } \mu\text{m}$ (Draine and Lee 1984). Blackbody spectra arise when a medium is in thermal equilibrium with its surroundings and are described by the Planck function

$$B_{\lambda}(T) = \frac{2 h_{\text{p}} c}{\lambda^5} \frac{1}{\exp(h_{\text{p}} c / \lambda k_{\text{B}} T) - 1}, \quad (2.9.66)$$

where h_{p} is the Planck constant, c is the speed of light and k_{B} is the Boltzmann constant (see, for example, the excellent Rybicki and Lightman 1979 text for a derivation). It can also be written equivalently in terms of frequency rather than wavelength,

$$B_{\nu}(T) = \frac{2 h_{\text{p}} \nu^3}{c^2} \frac{1}{\exp(h_{\text{p}} \nu / k_{\text{B}} T) - 1}, \quad (2.9.67)$$

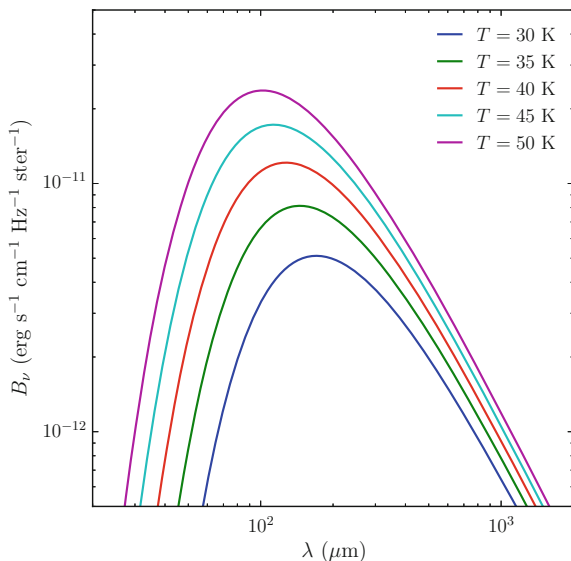
note that $B_{\nu} \neq B_{\lambda}$, but $B_{\nu} d\nu = B_{\lambda} d\lambda$.

Blackbody spectra have a number of interesting properties which we will now discuss briefly. In the limits that $h_{\text{p}} \nu \ll k_{\text{B}} T$ and $h_{\text{p}} \nu \gg k_{\text{B}} T$ we regain respectively the Rayleigh-Jeans law,⁷ $I_{\nu}^{\text{RJ}}(T) = 2(\nu^2/c^2)k_{\text{B}}T$, and the Wien law, $I_{\nu}^{\text{W}}(T) = 2(h_{\text{p}}\nu^3/c^2)\exp(-h\nu/k_{\text{B}}T)$. Throughout this thesis we are mostly concerned with the portion of the spectrum long-wards (in terms of wavelength) of the peak, which is often referred to as the Rayleigh-Jeans tail.

Another property of a blackbody is ‘monotonicity of temperature’, given two blackbody curves the one with the higher temperature will lie entirely above the other. This can be shown by the fact that the derivative with respect to temperature,

⁷Note that the integral of this law over all frequencies diverges, leading to the so-called *ultraviolet catastrophe*.

Fig. 2.2 Examples of blackbody spectra at different temperatures, as indicated in the legend



$$\frac{\partial B_\lambda(T)}{\partial T} = \frac{2 h_p^2 c^3}{\lambda^6 k_B T^2} \frac{\exp(h_p c / \lambda k_B T)}{[\exp(h_p c / \lambda k_B T) - 1]^2}, \quad (2.9.68)$$

is always positive. We give a visual example of this in Fig. 2.2. What is also evident from this Figure is that as the temperature increases so too does its integral (it can be shown that the bolometric luminosity $L = \int B_\lambda d\lambda \propto T^4$) and the wavelength at which the function peaks shifts to shorter wavelengths. This latter point is referred to as the *Wein displacement law*, and it can be shown by solving $\partial B_\lambda / \partial \lambda|_{\lambda=\lambda_{\max}} = 0$, that $\lambda_{\max} T = \text{const}$. Thus, as the temperature of the blackbody increases the wavelength at which its emission peaks must shift to shorter wavelengths to compensate. In the case of the 30 K spectrum it can be seen in Fig. 2.2 that the radiation peaks at $\lambda \sim 10^2 \mu\text{m}$, which is roughly the wavelength at which the FIR SED of many dusty galaxies is observed to peak. This indicates that this is a good ‘rule of thumb’ for the dust temperature of such galaxies.

In practise, real-world materials rarely emit as a perfect blackbody and so they are often described by a ‘modified’ blackbody, the ‘modification’ in this case being described by the κ_d term in Eq. (2.9.64).

The assumption of thermal equilibrium with Eq. (2.9.64) is applied separately to dust in molecular clouds (Eq. 2.9.62) and the diffuse ISM (Eq. 2.9.63).

For galaxies undergoing a starburst a value of $\beta_b = 1.5$ is used whereas for quiescently star forming galaxies an unbroken power law is assumed i.e. $\lambda_b \rightarrow \infty$. This choice was originally motivated by Silva et al. (1998), who found that a lower value of β_b was necessary to reproduce the SED of Arp 220, a local starburst galaxy, and laboratory measurements of silicates suggest that values in the range $\beta_b = 1.5 - 2$ are acceptable (e.g. Agladze et al. 1996). There is also some observational evidence that

suggests an anti-correlation between β and temperature, with hotter dust exhibiting lower values (e.g. Boudet et al. 2005). Note that in GALFORM, starburst galaxies have generally hotter dust temperatures than quiescently star-forming galaxies (Chap. 5).

2.9.4 Absorption by the IGM

Neutral hydrogen in the IGM along the line of sight can absorb ionizing photons. We compute this where relevant for the model galaxies according to the prescription of Meiksin (2005).

2.10 The Lacey et al. (2016) GALFORM Model

In this Section we briefly describe some features of the Lacey et al. (2016) version of GALFORM, the fiducial model throughout this thesis, that are pertinent to the work that follows.

This model incorporates a number of important physical processes from earlier iterations of GALFORM, and is thus thought of as a ‘unifying’ model for various ideas that have been developed throughout the GALFORM literature. Additionally, the cosmological parameters used are updated to those consistent with Wilkinson Microwave Anisotropy Probe 7 year (*WMAP7*) data (Komatsu et al. 2011).⁸ The physical processes incorporated from earlier versions of GALFORM include the star formation treatment described in Sect. 2.4.1, which is based on an empirical relation between star formation rate and molecular gas (Blitz and Rosolowsky 2006) and was first implemented into GALFORM by Lagos et al. (2011), and a *hot-halo mode* AGN feedback prescription (Sect. 2.5.3) in which quasi-hydrostatic hot halo gas is prevented from cooling by energy input from relativistic jets (Bower et al. 2006).

Importantly for this thesis, the Lacey et al. (2016) model can simultaneously reproduce the optical and near-IR galaxy luminosity functions at $z = 0$ (see Fig. 2.3) and the observed 850 μm number counts and redshift distribution. As mentioned earlier, in order to do so it treats the slope of the IMF, x (Sect. 2.8.2), as a free parameter in starbursts, as was found to be necessary to reproduce the 850 μm number counts in the ΛCDM paradigm by Baugh et al. (2005). Following an extensive parameter search using GALFORM, Baugh et al. advocated a slope of $x = 0$ making the IMF ‘top-heavy’ relative to a solar neighbourhood IMF (e.g. a Salpeter 1955 IMF is described by a slope of $x = 1.35$). An earlier GALFORM model, Granato et al. (2000), was able to reproduce the b_J - and K -band luminosity functions at $z = 0$ but underpredicted the sub-mm number counts by a factor of ~ 30 . Attempts to increase

⁸ $\Omega_0 = 0.272$, $\Lambda_0 = 0.728$, $h = 0.704$, $\Omega_b = 0.0455$, $\sigma_8 = 0.81$, $n_s = 0.967$. Though note that in Chap. 6 we use cosmological parameters consistent with more recent *Planck* data (Planck Collaboration et al. 2016).

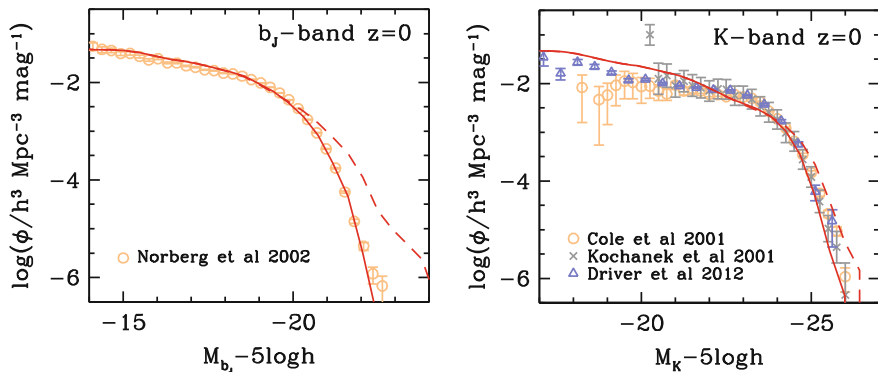


Fig. 2.3 Calibrated predictions of the fiducial Lacey et al. (2016) model for the b_J - and K -band luminosity functions (*left* and *right* panels respectively) at $z = 0$. The *solid* and *dashed* lines show the model predictions with and without dust extinction respectively. Observational data are from Norberg et al. (2002), Cole et al. (2001), Kochanek et al. (2001) and Driver et al. (2012). Figure taken from Lacey et al. (2016)

the star formation rate in the Granato et al. model (e.g. through reducing feedback strength) alleviated the discrepancy between the predicted and observed sub-mm number counts, but meant that galaxies were too massive by the present day and thus reduced the level of agreement between the predicted and observed $z = 0$ luminosity functions. A top-heavy IMF works here by boosting the sub-mm flux per unit star formation in a galaxy. This is through increasing the UV luminosity, as a greater number of massive stars are produced in a top-heavy IMF, and an increase in the dust mass of the galaxy through a larger yield as these stars go supernova and eject their metals into the ISM. This latter point allowed the dust to remain cool enough ($T_d \sim 30$ K) to re-emit the increased absorbed UV radiation effectively in sub-mm bands. It also has little impact on the $z = 0$ stellar mass of a galaxy, as most bursts occur at $z \gtrsim 2$ in the model, and the massive stars that form under the top-heavy IMF do not live long enough to persist to $z = 0$.

The assumption of a varying IMF is a somewhat contentious one. In a review of observational studies Bastian et al. (2010) argued against significant variations from a solar neighbourhood IMF in the nearby Universe. However, studies based on gravity-sensitive absorption features have found evidence for a bottom-heavy IMF in early-type galaxies (e.g. Conroy and van Dokkum 2012) whereas Gunawardhana et al. (2011) found a trend of flattening IMF slope (i.e. becoming increasingly ‘top-heavy’) with increasing SFR from studying integrated $H\alpha$ emission in the Sloan Digital Sky Survey (SDSS). There are also some theoretical arguments for a varying IMF. Larson (2005) argues that the characteristic mass in the ISM should scale with the Jeans mass of a star-forming cloud, and that this should be larger in actively star-forming galaxies due to radiation heating from massive stars, plausibly leading to a more top-heavy IMF than found locally. Thus it seems that IMF variation is still a somewhat open issue and is therefore worth investigating in models such as

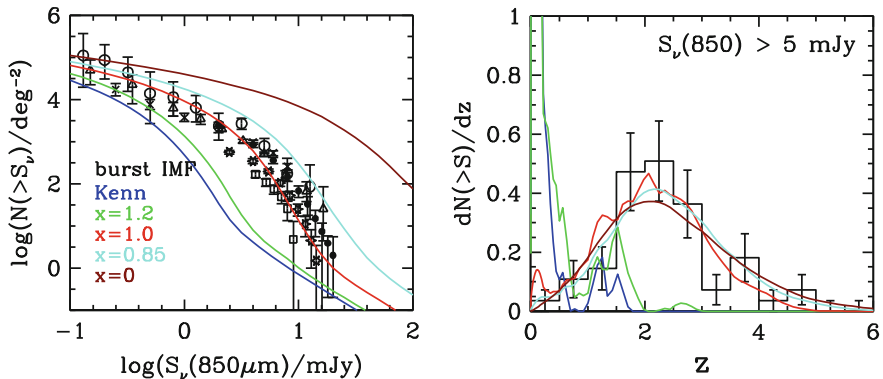


Fig. 2.4 Dependence of cumulative 850 μm number counts (*left panel*) and redshift distribution of galaxies with $S_{850\mu\text{m}} > 5$ mJy (*right panel*) on starburst IMF slope, x , in the Lacey et al. (2016) model. The different colour lines indicate the predictions from models with different values of x , as indicated in the legend. Observational number count data in the *left panel* are from Coppin et al. (2006, filled circles), Knudsen et al. (2008, open circles), Zemcov et al. (2010, crosses), Karim et al. (2013, open squares) and Chen et al. (2013, open triangles). The *right panel* shows the redshift probability distributions for galaxies with $S_{850\mu\text{m}} > 5$ mJy. Observational data is from Wardlow et al. (2011). Figure taken from Lacey et al. (2016)

GALFORM. Lacey et al. found a slope of $x = 1$ was better able to reproduce the observed sub-mm number counts, rather than the $x = 0$ used by Baugh et al. This is shown in Fig. 2.4.

The reason behind a steeper IMF slope than used by Baugh et al. is related to the inclusion of disc instabilities (Sect. 2.6.2) and AGN feedback (Sect. 2.5.3), which were not considered by Baugh et al. An additional consequence of this is that whereas in Baugh et al. sub-mm galaxies are undergoing merger-induced starbursts, in the Lacey et al. model they are predominantly disc instability-induced starbursts. Disc instabilities allow for more starbursts at early times so that the observed abundance of SMGs at $z \gtrsim 3$ can be reproduced, whilst also growing a galaxy’s SMBH through accretion during a starburst (see Sect. 2.5.3) so that AGN feedback is efficient at later times ($z \lesssim 2$) resulting in further star formation being inhibited such that the model does not overproduce SMGs. This is summarised in Fig. 2.5. Turning off merger-induced starbursts has a negligible impact on the predicted sub-mm number counts. Indeed, we discuss that galaxy mergers are rather inefficient at boosting star formation in the model in Chap. 5, as they are responsible for a population of starburst galaxies that actually tend to lie below the main locus of star-forming quiescent galaxies on the specific star formation rate—stellar mass plane.

We end this Section by summarising the parameters in the Lacey et al. (2016) model in Table 2.1. It is worth noting that adjustable parameters form only a subset of those listed in Table 2.1. For example cosmological parameters are set by the adopted cosmology, similarly, the yield, p , and recycled fraction, R , are set by the choice of IMF, which is only treated as a parameter for burst star formation. Additionally,

Fig. 2.5 Dependence of the model redshift distribution for $S_{850\ \mu\text{m}} > 1\ \text{mJy}$ galaxies on AGN feedback and disc instabilities in the Lacey et al. (2016) model. The red, blue and green lines are, respectively, predictions from the fiducial model, a model in which AGN feedback has been turned off, and a model in which there are no disc instabilities

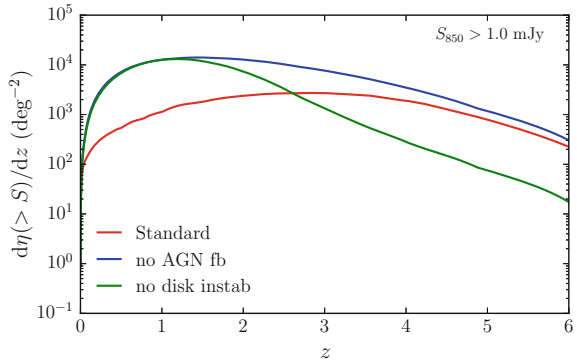


Table 2.1 Values of input parameters for the fiducial Lacey et al. (2016) GALFORM model. Adapted from Lacey et al. (2016)

Parameter	Value	Description	Reference
Cosmological parameters			Komatsu et al. (2011)
Ω_m	0.272	Matter density	
Ω_b	0.0455	Baryon density	
h	0.704	Hubble parameter	
σ_8	0.81	Fluctuation amplitude	
n_s	0.965	Scalar spectral index	
Stellar population			Maraston (2005)
IMF: Quiescent (Kennicutt 1983)			
p	0.021	Yield	(2.8.49)
R	0.44	Recycled fraction	(2.8.48)
IMF: Burst			
x	1	IMF slope	(2.8.50)
p	0.048	Yield	(2.8.49)
R	0.54	Recycled fraction	(2.8.48)
Star formation: Quiescent			Lagos et al. (2011)
ν_{SF}	$0.74\ \text{Gyr}^{-1}$	Efficiency factor for molecular gas	(2.4.14)
P_0	1.7×10^4	Normalisation of pressure relation	(2.4.13)
α_p	0.8	Slope of pressure relation	(2.4.13)
Star formation: Burst			Baugh et al. (2005)
f_{dyn}	20	Multiplier for dynamical time	(2.4.16)
$\tau_{\text{burst, min}}$	0.1 Gyr	Minimum burst timescale	(2.4.16)

(continued)

Table 2.1 (continued)

Parameter	Value	Description	Reference
Supernova feedback			Cole et al. (2000)
V_{SN}	320 km s^{-1}	Normalisation of mass loading	(2.5.18)
γ_{SN}	3.2	Slope of mass loading	(2.5.18)
α_{ret}	0.64	Gas reincorporation timescale multiplier	(2.5.19)
AGN feedback and SMBH growth			Bower et al. (2006)
f_{BH}	0.005	Fraction of mass accreted onto SMBH in starburst	Malbon et al. (2007)
α_{cool}	0.8	Ratio of cooling/free-fall time	(2.5.20)
f_{Edd}	0.01	SMBH Eddington luminosity multiplier	(2.5.21)
ϵ_{heat}	0.02	SMBH heating efficiency	
Disc stability			
F_{stab}	0.9	Threshold for disc stability	(2.6.24)
Galaxy mergers			Jiang et al. (2008)
Size of merger remnants			Cole et al. (2000)
f_{orbit}	0	Orbital energy contribution	
Starburst triggering by mergers			Baugh et al. (2005)
f_{ellip}	0.3	Mass ratio threshold for major merger	
f_{burst}	0.05	Mass ratio threshold for starburst	
Dust model			Granato et al. (2000)
f_{cloud}	0.5	Fraction of dust mass in clouds	
t_{esc}	1 Myr	Escape time of stars from clouds	(2.9.58)
β_{b}	1.5	Dust emissivity slope in starbursts	(2.9.65)

Note P_0 is in units of $k_{\text{B}} \text{ cm}^{-3} \text{ K}$

parameters in the quiescent star formation law are set by observational results (e.g. Leroy et al. 2008; Bigiel et al. 2011). Once the parameter values are set the model is fully specified, this forms the starting point for the work presented in this thesis and unless explicitly stated they are not changed throughout.

The Nature of Dusty Star-Forming Galaxies

Cowley, W.

2017, XXXV, 196 p. 74 illus., 57 illus. in color.,

Hardcover

ISBN: 978-3-319-66747-8