Lecture 3:
The theory of galaxy formation

Dr. James Mullaney

February 5, 2018

1 Introduction

Before looking in detail at piecing the evolution of galaxies together from observations, we'll start with considering how we simulate galaxy evolution, since much of our theories are based on such models.

2 Broad theories of galaxy formation

There are two broad theories of galaxy formation:

- **Monolithic collapse** (Eggen, Sandage & Lynden-Bell 1962)[1]: Here, a single, large cloud of gas collapses to form a single, massive galaxy. Density fluctuations in the cloud lead the denser regions to collapse to form a loose cluster, or halo, of stars. As the cloud collapses further, it conserves angular momentum; the dominant component of rotation wins-out over the others and a disk is formed. Gas continues to collapse within this disk to form more stars. In this scenario, all of the mass that eventually forms the galaxy is present from the outset in one big "lump".

- **Hierarchical galaxy formation** (White & Rees 1978): In this scenario, larger galaxies are built-up over time by the repeated merger of smaller galaxies. The smaller galaxies are created by the collapse of gas clouds to form collections of stars. These then coalesce to form a larger galaxy, dragging their gas content with them, enabling some star-formation to continue in the resulting larger galaxy. This idea of hierarchical merging fits well with our models of a Universe that is dominated by collisionless dark matter. It is currently the most widely-accepted model of galaxy-formation and evolution.

While the hierarchical model of galaxy formation is currently our best theory to explain the observed properties of galaxies, it needs a lot of "tweaking" in order to accurately reproduce what we see. Many of these tweaks *are not* physically motivated, and are instead set by hand to reproduce observations. For example, among of the most important tweaks are the efficiencies of various "feedback" processes, such as how light from young stars heats their surrounding gas, thus preventing it from forming further stars. The physics governing these feedback processes are still too complex to

[1] On researching this, I came across a report that suggests Olin Eggen was a bit of a kleptomaniac. After his death in the last '90s, scores of rare books that had gone missing from the Royal Greenwich Observatory three decades earlier were found in his office. He had always insisted that he never had them.

model, so theoretical astronomers have to estimate their efficiencies until their models reproduce the properties of galaxies in the real Universe. As such, although our models are very effective at reproducing observations, it *does not* mean we fully understand galaxy formation/evolution.

3 Methods of modelling hierarchical galaxy evolution

There are two main methods of modelling galaxy formation and evolution: **hydrodynamic** and **semi-analytic**. As you may expect, both have their benefits and pitfalls (otherwise, we'd only use one). In the following two subsections, we'll explore both these methodologies.

3.1 Full hydrodynamical simulation

Of the two methods, this is probably the easiest to understand conceptually. Hydrodynamic simulations try to model the motion of dark matter, stars and gas under the influence of physical forces. For dark matter and stars this is relatively easy since they are both *collisionless*; in other words, they only interact via their gravity (the chances of two stars colliding are extremely small due to their tiny size compared to the separation between them). As such, dark matter and stars can be modelled using "straight-forward" N-body simulations.

Gas, on the other hand, is *highly dissipational* (i.e., it dissipates gravitational potential energy easily in the form of radiation). Gas also readily absorbs energy, which affects its ability to collapse and form stars. As such the behaviour of gas is governed by many more processes than gravity: e.g., heating, cooling, pressure, etc. To get a sense of how much more complicated modelling gas is compared to considering only gravity (as for stars and dark matter), consider the complex motion of air from a hairdryer to the simple orbits of the planets. To be modelled fully, gas needs to be treated *hydrodynamically*.

There are two main types of hydrodynamic models: **Lagrangian** and **Eularian**. The easiest way to think about the difference between the Lagrangian and Eularian aproaches is to consider their most common examples: **Smoothed Particle Hydrodynamics (SPH)** (Lagrangian) and **Mesh models** (Eularian). In SPH, the gas is treated as a population of particles that interact with other gas particles via a *smoothing length*. The larger the smoothing length, the more the gas interacts with itself (in SPH codes, stars and dark matter are treated as particles with a zero smoothing length). Taken to the extreme, an *ideal* SPH model would consider every individual molecule or atom of gas as a particle. In reality, however, we typically have to assume individual gas particles that are many parsecs across and contain many solar masses of gas (our computers are a long way from being able to simulate every atom of gas in a galaxy).

In Mesh models, the modelled volume is split up into very small cells, and *continuity equations* are used to calculate how much gas enters a given cell from its neighbours, and how much leaves the same cell to its neighbours. If one is different from the other, the density and pressure of the gas in that cell must change. The smaller the cells, the more precise your model. However, decreasing the size of the cells in one dimension by a factor of two increases the number of cells by a factor of $2^3 = 8$, with a corresponding increase in the number of calculations required (and thus the total time it takes for the model to run).

In each type of model, the physical conditions of the gas (such as temperature, density, pressure) are calculated at each *timestep*. The shorter the timestep in the model, the more accurate it is, but the more calculations are needed (and thus, again, an increased running time). After each timestep, the properties of the gas represented by a given particle or contained within a given cell are compared to their neighbours and calculations are made to determine how it interacts with its neighbrouing particles or cells. For example, heat may be passed from one particle/cell to another, there may be bulk motion from one particle/cell to its neighbours etc. Also, "prescriptions" (which may be well-defined, or simply guesses) are used at each timestep to calculate how much the gas should be heated, cooled, and how much will have collapsed to form stars, or fallen into a black hole.

3.2 Semi-analytic models

Semi-analytic models (SAMs) take the philosophy of replacing the most complicated aspects of hydrodynamic simulations with simple analytic expressions. As such, they tend to be far quicker than hydrodynamic models, but rely on far more assumptions.

As with hydrodynamic models, SAMs are not restricted to astrophysics; they are also commonly used to model the Earth's climate. In general, they use relatively simple N-body simulations to model a core component of the physical situation, then use analytical expressions to model the detailed processes on top of this underlying core. In galaxy evolution, they exploit the widely accepted concept that the dominant form of matter in the Universe is dark matter. The argument goes that, since dark matter dominates, we can model the dark matter – which is relatively easy, since it is non-interacting – and then use analytic expressions to populate this dark matter with gas and stars.

One key benefit of the SAM approach is that the big N-body calculation, i.e., that in which the dark matter is modelled, only needs to be performed once. After that, any number of different analytic expressions can be used to generate populations of galaxies (in a relatively short time, since they are analytic), which can then be compared against observations of the real Universe. For example, one model could be that every "blob" (or, "halo") of dark matter contains a galaxy with a stellar mass (i.e., the sum of the mass of all its stars; M_*) that is 1% of the mass of the dark matter halo ($M_{\rm Halo}$). The analytic expression for this would be:

$$M_* = 0.01 \times M_{\rm Halo} \tag{1}$$

As you can see, this is a really simple expression, but it would result in a population of galaxies with given masses. Even if our simulation contained a billion dark matter halos, it would only take a few seconds for our semi-analytic model to the masses of the galaxies. If, on comparison against the real Universe, we then realised that this was a bad model, we could easily try 2% (without having to run the whole N-body dark matter model again) and see if that were any better. Today's SAMs are very sophisticated, with anayltic expressions used to populate dark matter halos with gas, to control the cooling of this gas, to control how stars form, to control feedback processes, etc., but the principle remains the same.

In what follows, we will go through the steps needed to generate a more typical SAM of galaxies in the Universe:

3.2.1 Choosing initial conditions

Any simulation needs a starting point: a set of initial conditions at time $t = 0$ that the simulation can then evolve to the next time step. In cosmological SAMs, the initial condition is set by the distribution of dark matter 379,000 years after the Big Bang. This time is chosen as it is the point at which the Cosmic Microwave Background (CMB) was emitted. The CMB has been well-studied by satellites such as COBE, WMAP and Planck. As a result, the temperature fluctuations of the CMB are now very well-defined. These temperature fluctuations give a representation of the matter distribution in the early Universe, which cosmological SAMs use as their starting point. On measurement, the probability of a given density fluctuation δ is given by a Gaussian field:

$$p(\delta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\delta^2}{2\sigma^2}\right) \tag{2}$$

SAMs use this equation to populate their models with an initial distribution of dark matter (i.e., at each point in your $t = 0$ model allocate a density selected randomly from a Gaussian distribution). Since simulations are limited by computing power, they must have a limited resolution; in the most widely used cosmological SAM – the Millenium Simulation – this resolution corresponds to each dark matter "particle" having a mass of a billion solar masses.

3.2.2 Generate a merger tree

Once the initial conditions of our dark matter are set, we can start our simulations running. Since we're dealing only with dark matter at this stage, we only have to contend with two factors: (a) the expansion of the Universe due to the Big Bang and, later, dark energy (these are all specified by our cosmological parameters, which are now pretty well-defined) and (b) gravity.

At the end of the simulation (i.e., often, but not always, when $t =$ today), all of the clumps of dark matter (known as dark matter halos) are identified, and all the dark matter particles that end up in each halo are traced back through the simulation to their starting point. This creates what is referred to as a "merger tree", since at $t = 0$ there are lots of separate particles (i.e., branches) that first merge to form mid-sized halos (i.e., limbs) and then large halos (i.e., tree-trunks).

The merger-tree represents the end-point of the N-body dark matter simulation. It describes the full merger-history of all dark matter particles in the simulation, and is all we need if we want to use analytic expressions to create a population of galaxies. This is because, since we're assuming that dark matter dominates over everything else, we don't have to worry about the detailed motions or positions of the halos when using analytics to populate them with baryons (i.e., gas, stars, galaxies etc). Everything after this stage is the "analytics" part.

3.2.3 Cooling of gas in dark matter halos

With the merger trees in-hand (and remember, they provide everything we need for the analytics stage), we can use analytic expressions to population them with gas. First, we use a prescription to allocate gas to a halo; this is usually as simple as saying that each halo gets its "fair share" of baryons. In other words, each halo is given a mass of baryons proportional to its dark matter mass (usually $M_b \approx 0.15 M_{\rm DM}$; i.e., the ratio of baryonic matter to dark matter).

Next, the gas is assumed to fall toward the centres of their dark matter halos. As it does this, the gas gets shocked and heated to the virial temperature, given by:

$$T_{\rm vir} = \frac{1}{2} \frac{\mu m_H}{k} \frac{G M_{\rm Tot}}{r_{\rm vir}} \tag{3}$$

where m_H is the mass of the Hydrogen atom, k is the Boltzmann constant, $M_{\rm Tot}$ is the total halo mass, and $r_{\rm vir}$ is the virial radius (here, the radius within which a cloud of gas is destined to form a galaxy).

Then, the hot, shocked gas is assumed to cool from the inside outwards with a disk of cool gas forming due to the conservation of angular momentum. How quickly the gas cools is also defined analytically and can depend on many factors, such as the temperature and composition (i.e., metallicity) of the gas. It can therefore be quite complicated (and thus subject to a lot of tweaking). Of course, after the first round of star formation in the model, we can include analytic expressions that control how the metallicity of the gas changes due to stellar reprocessing, which will affect later gas cooling. Thus, the model can become highly self-interacting, or *dynamic*.

3.2.4 Star-formation

Since we now have cooling gas, we should think about the natural consequence: star-formation. Precisiely how gas clouds collapse to form (populations of) stars remains one of the biggest unanswered questions in astrophysics and is an active area of research (just ask Simon Goodwin, whose research focusses almost entirely on this area). Without a complete theory of star-formation, semi-analytic models rely on empirically-defined relationships between the amount of available cold gas and the rate of star-formation. The most well-known of these is the Schmidt-Kennicutt relationship:

$$\Sigma_{\rm SFR} = (2.5 \pm 0.7) \times \Sigma_{\rm gas}^{(1.4 \pm 0.15)} \tag{4}$$

where $\Sigma_{\rm SFR}$ is the star-formation rate per unit area (in $\mathrm{M}_\odot\ \mathrm{yr}^{-1}\ \mathrm{pc}^{-2}$) and $\Sigma_{\rm gas}$ is the surface-density of gas (in $\mathrm{M}_\odot\ \mathrm{pc}^{-2}$). So, with our analytic prescription for how much cool gas there is in a galaxy, we can use Eqn. 4 to calculate the rate of star-formation and build-up our galaxy stellar masses in our simulation over time.

3.2.5 Feedback processes

When the first large, cosmological SAMs were run, it was found that they tended to make galaxies that were far more massive than we observe in the real Universe. This implied that there were some processes in the real Universe that was preventing galaxies from forming too many stars that were not being included in our models.

One of ways that gas can be prevented from forming stars is by heating it, so it stays too warm to collapse. Another way is to expel it from the potential well at the centre of the dark-matter halo. It was found that both these effects could be achieved by including "feedback" processes – so called because it is the actual act of gas cooling that causes these feedback processes to be triggered.

There are a number of potential feedback processes, but the two most important are due to supernovae and Active Galactic Nuclei (**AGN**). In the case of supernovae, once the first stars in the simulation have been formed, an analytic expression is used to determine what fraction will go supernova at a given time. These supernovae heat, and potentially expell, the surrounding gas, preventing it from forming new stars. It therefore provides *negative feedback* since the process of star formation actually acts to prevent further star formation (via their end-of-life supermovae).

In the case of AGN, gas cools and falls to the centre of the dark matter halo, where some of it will ultimately fall toward the supermassive black hole at the centre of the nascent galaxy. As it does so, the gas forms an accretion disk, which heats up and releases vast amounts of energy (we'll cover AGN in much more detail later in the course). In some SAMs, this energy is used to heat up and expell some of the surrounding gas, thus preventing it from forming stars (again, it is a form of negative feedback).

Feedback processes are some of the most uncertain and difficult-to-model features of models (SAMs and hydrodynamic). As such, they rank as some of the most argued-upon features of modern astrophysics. They almost certainly exist in the real Universe, but we don't have a very good grasp of how they work, and under what circumstances.

3.3 The benefits and pitfalls of Hydro models and SAMs

Now we've had a fairly comprehensive introduction to the two main types of cosmological galaxy evolution models, we can consider the benefits and pitfalls of each.

Hydro models

Benefits:

- Track the detailed evolution of dark matter, gas and star motions as galaxies evolve via gas accretion and mergers.

Pitfalls:

- The resolution is relatively poor ($\sim 20 - 100$ pc)
- At smaller scales than the resolution, we still have to make major assumptions about the physics ("sub-grid physics").
- While fewer in number, these sub-grid assumptions are akin to the analytic prescriptions in SAMs.
- As such, while it is less *degenerate* than SAMs, there is still room to tweak sub-grid parameters to match observations.
- Take a long time to run, so it is difficult to "try-out" lots of different sub-grid prescriptions to investigate how they alter the outcomes.

Semi-analytic models:

Benefits:

- Once the dark matter N-body simulation has been done, it is relatively inexpensive (i.e., quick) to try-out different analytic prescriptions.
- If you're not too interested in the detailed physics, they can give you a useful "mock universe" to plan observations etc.

Pitfalls:

- Don't track in detail the motion of gas or stars (since stars form from gas).
- Huge amount of free parameters (and more can be easily added), which can make them highly degenerate (i.e., the effects of changing one analytic parameter can be countered by changing another).
- Some of the analytics to describe the physics are highly uncertain (and some are just added to get the "correct result" without much physical intuition.

4 Learning objectives

In this lecture, we've considered the main *methods* of modelling galaxy formation and evolution. The key learning objectives are:

- Have and knowledge understanding of the main models for galaxy formation and evolution.
- Appreciate the differences between the semi-analytic and fully hydrodynamic approaches to modelling galaxy evolution.
- Have knowledge of the main steps involved in forming galaxies in SAMs: initial conditions, an "N-body" dark-matter simulation, merger trees, gravitational accretion and cooling of gas, shock heating, star formation and feedback.
- Have an understanding of the benefits and pitfalls of SAMs vs. hydrodynamic simulations.