Lecture 4: Angular Momentum and Galaxy Formation, continued

*The formation of elliptical galaxies*

Using the approximate relation \( \lambda \approx 0.4(V_m/\bar{\sigma}) \) for elliptical galaxies, where \( V_m \) is the maximum rotation velocity and \( \bar{\sigma} \) is the mean velocity dispersion, we find that bright ellipticals have \( \lambda \lesssim 0.1 \) (Davies et al. 1983):
So bright ellipticals are not significantly supported by rotation. Thus dissipation and angular momentum loss are important. As their sizes are roughly the same as spirals, if they conserved angular momentum and mass, they should fall on the same $J/M - M$ diagram as the spirals, but they don’t (Fall 1983):
This observation, that angular momentum loss is important, leads one to wonder how this angular momentum loss is achieved. It appears that the best hypothesis (still!) is that merging of disk galaxies produces ellipticals. Because this problem is basically an N-body problem (possibly including gas), it is not particularly tractable analytically. Therefore, we must turn to simulations...
Toomre & Toomre (1972) made the first electronic computer simulations of merging disk galaxies to attempt to explain the morphologies of galaxies found in the “Atlas of Peculiar Galaxies” (Arp 1966), such as NGC 4038/9 (Arp 244), “the antennae”: 

![Image of NGC 4038/9 (Arp 244), “the antennae”]

![Diagram of the antennae effect in galactic mergers]
Actually, the first simulations of gravitational encounters were made by Holmberg (1941) using *light bulbs* to trace the gravitational potential! This is because you can calculate the “gravitational force” at any point by measuring the total intensity of all of the light bulbs at that point (since light, like gravity, goes as $r^{-2}$ at a distance $r$ from the source).
Toomre & Toomre (1972) go on to imagine that the final product of such collisions might eventually settle down into a galaxy that looks *something* like an elliptical galaxy, and that “some elliptical galaxies” may have formed this way:

NGC 7252, from Kormendy (1982). Note that the inner isophotes (c) look nearly elliptical, even if the outer isophotes (d and f) look very disturbed. In fact, the azimuthally-averaged surface brightness profile is virtually that of an elliptical galaxy, following a de Vaucouleurs (1948) $r^{1/4}$ law (lower right).
The Toomres’ realization that ellipticals could form from mergers of disk galaxies touched off an on-going study of the interactions and mergers of galaxies. We’ll come back to the observations of this phenomenon later in the course, but first let’s examine some of the more recent models and simulations of elliptical galaxy formation.
First, let’s consider what happens in a merger. (This explanation is derived from considerations of N-body models by Barnes & Hernquist 1992.)

In a head-on collision of two spherical galaxies, the orbit of the two galaxies decays due to the gravitational compression arising when the two galaxies nearly coincide. This compression causes a slightly greater forces to be felt between them as they try to separate than at the same distances during approach.

![Image 1](image1.png)

*Fig. 8. Surface density as a function of time for two colliding polytropes, $v_{in}(0)=3(25)^{1/2}$. At $t=6$ the time is reversed and all particle orbits are reversed for $t=0$. Note compression at $t=1.5$ followed by expansion.*

![Image 2](image2.png)

*Fig. 6. Collision between two polytropes with $v_{in}(0)=0$. At $t=9$ the time is reversed and all particles are reverted to $t=0$.*

The collision of two polytropes, from van Albada & van Gorkom (1977).
This mechanism stirs up each galaxy at the expense of the orbital energy and thus they merge in only a few passages. In off-center collisions, stars that orbit their respective galaxy in the same direction get promoted into less-bound orbits, receiving binding energy and angular momentum. These produce the tails (see the figures from Toomre & Toomre earlier). If the galaxies rotate internally in the same direction as their passage, the orbital decay is even more rapid.

During the merger, binding energy is transferred between different components of the system, but the redistribution is incomplete. The centers and outskirts of the remnant remain dominated by stars from the centers and outskirts of the progenitor galaxies. Therefore gradients present in the progenitors should remain at some level in the product.
Ignoring a small amount of mass loss (if any) the scale of the remnant can be estimated from the total energy of the system. Given two identical galaxies merging through a parabolic encounter, the total mass and binding energy ($U$) of the merger product will be twice that of either victim. Then the gravitational radius ($r_g \equiv GM^2/|U|$) will be double that of one of the victims, the mean velocity dispersion will be equal to that of one of the victims, and the characteristic surface density will be half. But this only applies to the remnant as a whole, and not to the central properties of the remnant.
Of course, these dynamics of spherical galaxy mergers without dissipation are those of the dark halos and not disks within those halos. When disks are considered, life gets more interesting. These internal, tightly-bound components are not braked by the tidal forces; rather, these components lose angular momentum to their own halos. Thus, the visible parts of the galaxies can miss each other completely at first and yet still merge quickly because their halos merge.
As in a merger of spherical systems, the ordering of the binding energy is only blurred. The tightly-bound luminous components will remain individually concentrated within a more diffuse common halo until they find each other and merge. Then the luminous matter in the merger remnant tends to have a higher velocity dispersion, by up to \( \sim 40\% \) above the progenitors. However, the core of a remnant in a \textit{gas-free} merger simulation tends to have lower phase-space density than a real elliptical, so something is needed: either a pre-existing bulge or dissipation. This is a bit circular, actually—to get an elliptical from two disk galaxies, the galaxies already need an elliptical-like component. So gas is very likely to be needed, to increase the phase-space densities (Mihos & Hernquist 1994).
Where does the angular momentum go in a (gas-free) merger?

Left: the evolution of angular momentum in the disks (top), globular cluster system (middle), and dark halo (bottom) in a merger simulation, from Hernquist & Bolte (1993). Right: the evolution of angular momentum in a bulge with a dark halo (top) and without (bottom), from Barnes (1988).

Clearly, angular momentum is transferred from the luminous components into the dark halo. This solves our problem of why disks have $\lambda \approx 0.5$ and yet ellipticals have $\lambda \approx 0.07$: the merger conserves angular momentum but moves it into the (unobservable) dark halo.
Now let’s examine a merger involving gas. The stars will evolve collisionlessly, in the manner which we’ve already outlined, but the gas will cool and eventually form stars.


The majority of the gas is clearly driven into the center and forms stars, thus increasing the phase-space density in the core, as desired.
The exact timing of the star formation and the consequent distribution of gas and newly-formed stars apparently depends heavily on the internal structure of the progenitors, but not very much on the encounter geometry. Galaxies with only disks and halos will have several major star formation events in the merger, prompted by the formation of a bar instability, which efficiently drives gas into the center of the remnant. Galaxies which also contain bulges will have one significant star formation episode in the merger and are stable against the (internal) bar instability.

The star formation histories of gas-star mergers, from Mihos & Hernquist (1996).
We can put all of this together and insert it into a cosmological simulation (Steinmetz & Navarro 2002). We can now watch the morphological transformation of a disk into a bulge, due to a major merger at $z = 3.3$:
...into a disk with a bulge, as it reaccreses gas:

![Image of a disk with a bulge](image1)

...into a barred disk, due to the close passage and eventual accretion of a satellite between $z = 1.6$ and $z = 1.2$:

![Image of a barred disk](image2)

...and finally into an elliptical due to the last major merger at $z = 0.6$:

![Image of an elliptical galaxy](image3)

The evolution of the most massive halo in a very high-resolution cosmological ($\Lambda$CDM) simulation (Steinmetz & Navarro 2002).
Let’s examine this elliptical in a little more detail (Meza et al. 2003). The mass-accretion and star-formation histories are sort of what we might expect for an elliptical galaxy that formed from a set of mergers:

The mass evolution and star formation history of a massive elliptical galaxy in a ΛCDM simulation (Meza et al. 2003).

The mass in stars is dominated by stars formed at early times, although as we’ll see later, the light of this galaxy is likely to come to from both the younger and oldest stars.
The structure and kinematics are also basically what we expect, if somewhat more compact with a higher central surface brightness than is typical for giant ellipticals.

The surface brightness profile (left) and amount of rotational support (right) of the remnant elliptical (blue squares), from Meza et al. (2003).
So is this really the way that ellipticals form?

There is an alternate school of thought for the formation of elliptical galaxies, elucidated most clearly by Larson (1969, 1974a,b, 1975; Larson & Tinsley 1974): the “monolithic collapse” of an initial gas cloud that collapses and forms stars along the way. These models were motivated by the observations that elliptical galaxies are basically old, smooth objects with little sign (at least in the early 1970s) of inhomogeneities in stellar populations, kinematics, or morphology. Although this sort of collapse is unnatural in a hierarchical formation scenario, it is instructive to consider these models briefly—because rapid hierarchical collapse in the ancient Universe may resemble monolithic collapse to the modern observer.
Larson takes as his starting point a spherical, non-rotating but turbulent gas cloud that collapses and forms stars. Most authors (including those who make hierarchical clustering simulations) assume that the star formation rate follows the Schmidt (1959) law:

\[ \text{SFR} \propto \rho^n \]

where SFR is the star formation rate per unit volume and \( \rho \) is the gas (volume) density. (The power \( n \) is usually taken to be between 1 and 2.) Larson generalizes this assumption, allowing the gas to form stars at a rate that is proportional to both the local density and velocity dispersion of the gas. Then he writes down the Boltzmann equation (i.e., the fluid dynamics) for the stars and gas individually, increasing stellar density and decreasing the gas density as the gas is transformed into stars. From the moments of the Boltzmann equation, he can then follow the evolution of the mass in stars and gas, their densities, and their velocity dispersions.
In these models, one must assume an initial mass, radius, and velocity dispersion (i.e., temperature) for the collapsing gas cloud. One also must assume that the required turbulence happens on scales much smaller than that of the galaxy; indeed, Larson assumes that the appropriate scale is the Jeans’ mass for the gas. This assumption has the rather nice property that the gas is already collapsing into the densities required to form stars.
Next, one can add other knobs to the model: gas cooling, initial mass function of stars, stellar mass loss, supernovae explosions, and metal production. The surface density profile (ignoring SN) of a typical model in this scheme is strikingly similar to a real elliptical, after tuning the initial conditions and a few constants:

A monolithic collapse model tuned to fit NGC 3379, from Larson (1974a).
The star formation history of this model is quite reminiscent of the elliptical from the hierarchical collapse simulations. Moreover, the metallicity gradients seen in some ellipticals are naturally reproduced.

(Left) The star formation history at different radii and (right) the metallicity gradient in the final galaxy in a monolithic collapse model (Larson 1974a).
We’ll end this section with a question. Did ellipticals form from hierarchical collapse of small structures, or did they form from the monolithic collapse of a single gas cloud? Do these models make *concrete, testable, and falsifiable* predictions for observable properties of elliptical galaxies?

Before moving away from models of the formation of galaxies, we should briefly survey the two classical models of the formation of our own Galaxy, the Milky Way.

Eggen, Lynden-Bell & Sandage (1962＝ELS) reformulated the conclusions of Roman (1955) that high-velocity stars near the Sun tend to be metal-poor (cf. Oort 1926). They showed that the eccentricities $e$, energies of oscillations perpendicular to the disk $E_z$, and angular momenta $L_z$ of these stars are correlated with stellar metallicities such that as

\[
\begin{align*}
[\text{Fe/H}] & \downarrow & e & \uparrow \\
[\text{Fe/H}] & \downarrow & E_z & \uparrow \\
[\text{Fe/H}] & \downarrow & L_z & \downarrow
\end{align*}
\]
ELS concluded that either the MW had a violent formation history or that most low-metallicity stars did not form in a centrifugally-supported disk. This is the argument: Both $e$ and $L_z$ are *adiabatic invariants*, that is, their values for a given star will not change if the potential in which the star moves changes only slowly. Hence, either low-metallicity stars *formed* on orbits with high $e$ or the potential of the MW changed rapidly since these stars were formed.

They therefore presented a formation scenario in which the MW formed from a (roughly) spherical, rotating protogalactic cloud that was metal-poor and in freefall. As it collapsed, angular momentum was conserved and its spin increased. The most metal-poor stars and the halo globular clusters formed from the initially free-falling gas. Supernovae explosions increased the metallicity of the gas, and new stars formed. After shrinking, the remaining gas formed a centrifugally-supported, metal-rich disk that started forming the disk stars and clusters.
Searle (1977) and Searle & Zinn (1977=SZ) proposed a very different scenario for the formation of the MW’s halo. Motivated by the discovery that the metallicities of halo globular clusters span a wide range and do not correlated with distance from the Galactic center, they suggested that the Galaxy formed from the collapse of individual, independent gas clouds. The metallicity of these protoglobulars would depend on the number of supernovae explosions that occurred before the remaining gas is swept away (or the entire cluster is destroyed) by the kinetic energy of the supernovae.
Sandage has referred to SZ as “ELS+noise” but this is a gross overstatement, as the SZ picture really brought a fundamentally new perspective to galaxy formation. Moreover, the ELS picture of smooth collapse in a slowly-changing potential is clearly wrong, as shown by the N-body+gas experiments we’ve been examining. On the other hand, it’s also clear that the halo of our Galaxy didn’t form entirely from the mergers of smaller systems, as the nucleosynthetic history of the halo differs significantly from the Local Group dwarf galaxies, as we will see later.