Structure Formation in an Expanding Universe

Last time we talked about how extra-dense regions can collapse when the background is static. However, we know that the universe is expanding. The effects of expansion are the topic of this lecture.

The Evolution of Density Perturbations

Consider a spherical Newtonian explosion in vacuum. When the sphere has radius r, a particle on the edge of the explosion has speed $v(r)$, and the mass of the material interior to the particle is $M = M_{\rm crit}(1+\delta)$, where $M_{\rm crit}$ is the mass that would make $v(r)$ exactly equal to the escape speed at r and $\delta \ll 1$. How does δ depend on r?

To determine this, consider the specific energy of the particle, which is its energy per unit mass. This is

$$\frac{1}{2}v^2 - GM/r = C \; , \tag{1}$$

where the total specific energy C is a constant. If $M = M_{\rm crit}$, then the total energy is zero (by definition of $M_{\rm crit}$), so we can rewrite this as

$$\frac{1}{2}v^2 - GM_{\rm crit}(1+\delta)/r = \frac{1}{2}v^2 - GM_{\rm crit}/r - GM_{\rm crit}\delta/r = -GM_{\rm crit}\delta/r = C \; . \tag{2}$$

Since C is a constant and so is $M_{\rm crit}$ for $\delta \ll 1$, it follows that $\delta \propto r$.

Let's take stock of what this really means. It says that if a patch of space currently is slightly overdense, say by 1%, then the scale factor of the universe must double so that the patch is now overdense by 2% (note that this is *entirely independent* of the size of the patch). But because spacetime is expanding, the critical density is decreasing. This means that the actual density of the patch, measured say in kg m^{-3} is *decreasing* during this period. That's rather different than what we explored in the last class, where the density would just go up.

However, the density will not decrease indefinitely. In fact, a useful perspective is as follows: as Newton showed, matter inside a spherical shell feels no force from the shell. The extension in general relativity, called Birkhoff's theorem, then means that we can treat a slightly overdense spherical region of space as if it is a universe in its own right, with its own values of Ω, Ω_k, and Ω_Λ. Suppose we ignore Ω_Λ, which is okay for the redshifts $z > 10$ when structure first formed. Then a patch of space slightly denser than normal evolves as if it had positive curvature, meaning that the expansion slows, stops, and turns around.

In more detail, we can do a simplified (yet still realistic) treatment to determine the overdensity at the moment of turnaround. Suppose that the average density of the universe is exactly the critical density, and that a spherical region of mass M has a constant slight overdensity. We will assume that the matter in this region is cold dark matter: the "cold" means pressureless, so that in practice the matter just feels gravity but doesn't collide. Ignore dark energy, and consider the total energy of a particle of mass m at the edge of the region (of radius r). Energy is conserved, and m is conserved, thus so is the specific energy:

$$U = \frac{1}{2}\dot{r}^2 - GM/r < 0 \; . \tag{3}$$

As a result, $\dot{r}^2 = 2U + 2GM/r$. **Ask class:** how do they expect the turnaround time to depend on U? Suppose that the region starts with a radius $r_{\rm min}$, and expands to a maximum radius $r_{\rm max}$, at which point $\dot{r} = 0$ and the region stalls before turning around. The time T that this takes is

$$T = \int_{r_{\rm min}}^{r_{\rm max}} dr/\dot{r} = \int_{r_{\rm min}}^{r_{\rm max}} \frac{dr}{\sqrt{2U + 2GM/r}} \; . \tag{4}$$

We will now assume that $r_{\rm min} \ll r_{\rm max}$, so that we can set $r_{\rm min} \to 0$ in the integral.

Since $r_{\rm max}$ is given by the condition that $\dot{r} = 0$, it means that $r_{\rm max} = GM/(-U)$ (recall that $U < 0$, so this is positive as it should be!). We can then write the time as

$$T = \int_0^{-GM/U} \frac{dr}{\sqrt{-2U}\sqrt{\frac{-GM/U}{r} - 1}} \; . \tag{5}$$

We change variables to $x \equiv r/(-GM/U)$, leading to

$$T = \frac{GM}{-U\sqrt{-2U}} \int_0^1 \sqrt{\frac{x}{1-x}} dx \; . \tag{6}$$

Doing the integral then yields

$$T = \frac{\pi GM}{-2U\sqrt{-2U}} \; . \tag{7}$$

To determine the overdensity, though, we need to know the radius r_0 to which the background would have expanded, then the density ratio will be $(r_0/r_{\rm max})^3$. Because the background is at exactly the critical density, it means the energy is zero, so that $\dot{r}^2 = 2GM/r$. Therefore,

$$T = \int_0^{r_0} \frac{dr}{\sqrt{2GM/r}} \; . \tag{8}$$

Solving gives $r_0^{3/2} = (3\pi/4)(-GM/U)^{3/2}$. The density contrast is then

$$\rho/\rho_{\rm crit} = (r_0/r_{\rm max})^3 = (3\pi/4)^2 \; . \tag{9}$$

Wow! What a miracle! The energy U does not enter the calculation *at all*, so this result applies for all scales.

Well, maybe it isn't such a miracle after all. In fact, this is a reflection of the general point that Newtonian gravity has no scale to it: it's a simple power law no matter what. One way of thinking about this is that you could reverse the direction of time, starting with a uniform density sphere at rest and letting it collapse. The process of collapse looks the same regardless of the initial size; yes, the total time depends on the density, but everything scales with that. Therefore, it should not be a surprise that a dimensionless quantity (the ratio of densities) is the same no matter what.

Collapse and Virialization

What next? When the sphere has reached maximum expansion, it starts collapsing again. By symmetry it will take exactly the same time T to get to zero radius, by which point the background will have expanded to a radius $2^{2/3}$ times larger than before (hence a volume 4 times larger).

In reality, though, the cold dark matter particles will not reach zero radius. The reason is that since the mass is not distributed with perfect smoothness, a given particle will be closer to one particle than others, hence the gravitational acceleration it feels is not exactly radial. That particle will therefore acquire some small angular momentum, as will others (even if the sum total is zero). Energy is also transferred from one particle to another. The result is that as the collapse happens, there is a redistribution of energy and angular momentum. The net result is that the system undergoes one of my favorite processes: violent relaxation (makes me think of paintball on the weekend!). One one free-fall time, the system distributes itself so that although individual particles still orbit, the system as a whole does not contract or expand. This is called *virial equilibrium*, and the process is called *virialization*. The state of the system is, not surprisingly, governed by the virial theorem:

$$\frac{1}{2}\frac{d^2 I}{dt^2} = 2K + \Omega \tag{10}$$

where $K > 0$ is the kinetic energy of the system, $\Omega < 0$ is the gravitational potential energy of the system, and I is the moment of inertia of the system. If the system is more or less steady (moving, but not changing in bulk), then the left hand vanishes and we have $\Omega = -2K$. Note that this is satisfied by a circular orbit, but here we consider particles going every which way: the theorem is a statement about the bulk properties, not individual particles.

Ask class: how can we use this to determine the final radius of the region after virialization? We simply note that if the region got to radius $r_{\max}$ with zero speed, then the total energy is $-CM/r_{\max}$, where C is a constant that depends on the radial distribution of matter. The collapse is self-similar (meaning that the particles maintain their relative scaled positions during the collapse), so if the collapse is to a radius $r_{\rm final}$ then the energy equation is

$$-CM/r_{\rm max} = -CM/r_{\rm final} + K \ . \tag{11}$$

We know that $K = CM/(2r_{\rm final})$ from the virial theorem, which tells us that $r_{\rm final} = r_{\rm max}/2$. That's it!

This thus tells us that at the point of virialization, the density contrast is

$$\rho_{\rm vir}/\rho_{\rm crit} = (3\pi/4)^2 \times 8 \times 4 = 18\pi^2 \ . \tag{12}$$

The first factor is what we had at turnaround; the second is the increase in density due to virialization; and the third is the volume increase of the background expansion.

Of course, if the background universe has *less* than the critical density, then an overdensity might as well. In that case, there does not need to be a turnaround. More specifically, the current universe is underdense on average and has a repulsive term operating (the cosmological constant). If this continues, it means that the only gravitationally bound structures have already formed; no new ones are in the offing. This is in contrast to a universe with $\Omega = 1$ and $\Omega_k = \Omega_\Lambda = 0$, where in principle overdensities can keep on collapsing.

Theory and observation (particularly of the cosmic microwave background) tell us that the initial overdensities are larger on smaller scales. That means that small structure forms first, and large structure has to wait until lower redshifts. For this reason, observations of the largest structures in the universe can tell us about structure formation and also provide a new type of constraint on dark energy. We will therefore turn our attention to clusters of galaxies.

Zel'dovich Pancakes and Clusters

Mmmm...pancakes! Before discussing clusters per se, we should relax one of our assumptions: that everything is spherically symmetric. This is, of course, a classic physics assumption; cows and everything else are spherical until proven otherwise, just the same way that in cosmology all populations of objects are unevolving standard candles until we are forced to confront reality!

We will therefore take a tentative step towards the actual universe by considering an overdensity that is ellipsoidal instead of spherical. Suppose that the axes are $a > b > c$. **Ask class:** in which of the three directions will collapse happen first? It will be easiest for the collapse to happen along the shortest axis, c. This leaves a system that is pretty thin along one axis, and extended along the others, hence a "pancake" named after Yakov Borisovich Zel'dovich, a great Russian astrophysicist. The collapse will then happen along b, leading to filamentary structure. Finally, collapse occurs along a to get to a roughly spherical distribution (which could be modified if there is rotation present).

Note, in fact, that the figure shown in the last lecture demonstrates at least the filamentary nature of large-scale structure, but the spherical nature of smaller-scale structure. The reason, as we shall discuss shortly, is that initial large-scale perturbations are smaller than small-scale perturbations, so small scales can collapse and become spherical more quickly. In fact, clusters of galaxies are the largest spherical virialized gravitationally-bound systems in the universe (see Figure 1, which shows X-ray and optical images of a cluster).

Hierarchical Structure Formation

What, then, is the overall picture we have about how matter got together? Since cold (i.e., pressureless) dark matter dominates the mass, most simulations have ignored baryons although a few are starting to be more comprehensive. From the standpoint of dark matter, the issue is that small structure forms first because the initial perturbations were greatest, then as time went on (and the scale factor of the universe went up), larger structure formed. However, you should not think of these as isolated incidents, where in one place small structure formed and in a distinct place large structure appeared. Instead, it is a very dynamic dance indeed. In a large overdense region there are small extra-overdense regions that form structure. Then, those structures got together to form yet larger structures. The picture is then of dark matter blobs smashing together all the time to form bigger and bigger things. Nice movies of this are at http://star-www.dur.ac.uk/∼moore/movies.html. One aspect of all these collisions that appeals to me is that because many dark matter collections are thought to have black holes at their centers, collisions and dynamical friction (settling of heavy stuff) probably means that there are many inspirals and mergers of supermassive black holes with each other. Future space-based gravitational wave detectors will be able to see this, and thus give us a unique view of this crucial period in the history of the universe.

Intuition Builder

We described collisions of dark matter balls (and their associated baryons) as central to structure formation. Is that happening now? In particular, consider a galaxy cluster which currently has 1000 galaxies in a region of space 1 Mpc across. If you checked back in a trillion years, how would you expect it to look?

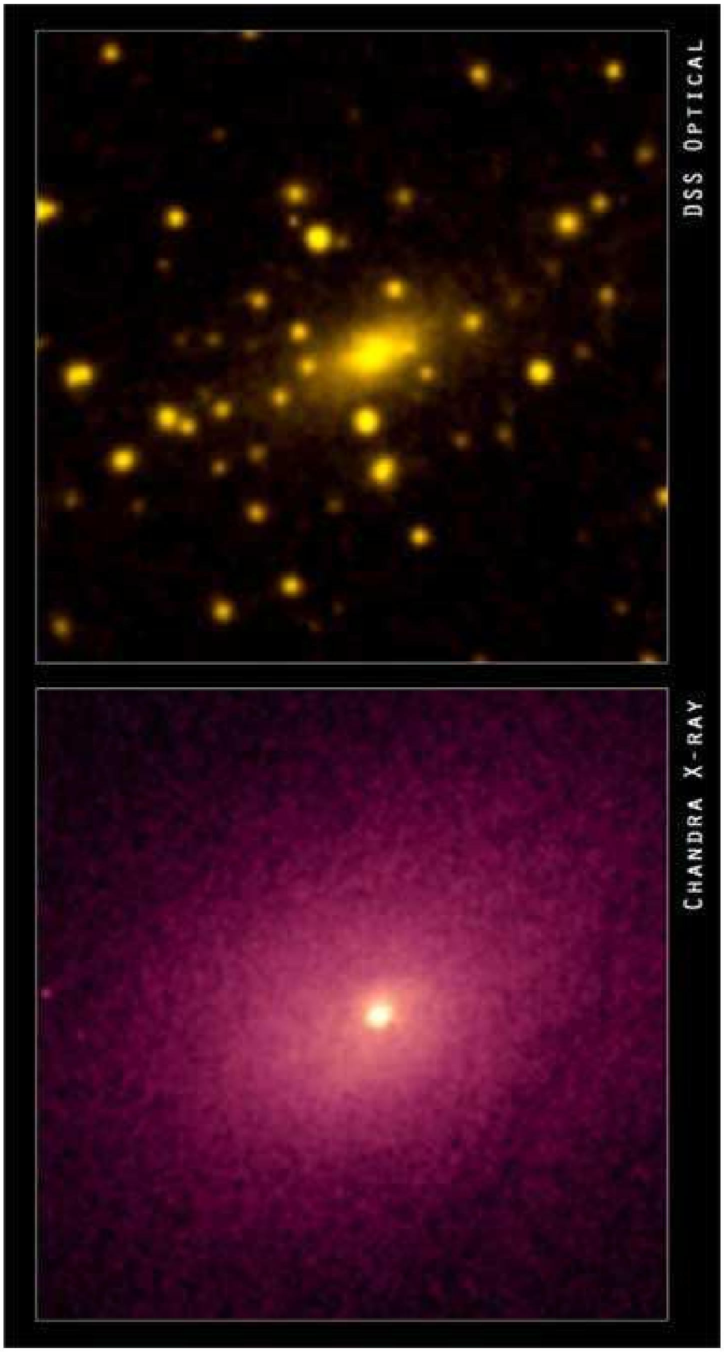

Fig. 1.— X-ray (bottom) and optical (top) images of the galaxy cluster Abell 2029. Hot gas contains several times more mass than exists in the cluster stars. From http://www.sr.bham.ac.uk/exgal/images/99150main_abell2029_comp_m.jpg