Clusters: Context and Background

We're about to embark on a subject rather different from what we've treated before, so it is useful to step back and think again about what we want to accomplish in this course. We have been accumulating a set of tools with which we can approach problems in high-energy astrophysics. For a given situation we should be able to ask some fundamental questions about the reliability of the answers. For example, if we read a paper about black holes we should ask "will general relativity be important"? If so, do the authors apply it correctly? **Ask class:** what are a couple of questions one might, similarly, ask about neutron stars? For neutron stars we can also ask about GR, and also ask whether strong magnetic fields may play a role. This means that if we see a paper about merging neutron stars that treats gravity in the Newtonian limit, we should be suspicious about the results.

Development of physical insight is improved if we consider a variety of physical regimes. In this course we've thought about extremely high densities and strong magnetic fields, which are not in the normal run of environments considered even in astrophysics. In the same spirit of broadening our exposure to different regimes, we will consider in the next three lectures a very different environment, yet one that also has been illuminated greatly by high-energy observations. This is the subject of clusters of galaxies. Clusters have great importance in a variety of contexts, but especially as probes of cosmology and of the formation of the first structure in the universe. This will therefore bring in many topics and questions that we have not yet considered. In this lecture we'll think primarily about context, in the next lecture we'll talk about the observational properties and what they imply, and in the third lecture we'll talk in detail about the Sunyaev-Zeldovich effect, which is a growing field with many cosmological applications.

Quick review of properties

Before heading into the context, here's a quick summary of some of the properties of clusters of galaxies. Clusters were first cataloged in a systematic way by Abell in 1958. He identified some 1700 clusters on the Palomar Observatory Sky Survey plates, by eye, and produced a richness criterion based on the galaxy count within a radius of $1.5\,h^{-1}$ Mpc, where $h = H_0/100$ km s^{-1} Mpc^{-1} ≈ 0.7. In such a volume, a cluster typically contains $\sim 10^3$ galaxies (compared with < 1 in a similar average volume in the universe), has a mass $M \sim 10^{14-15}\,M_\odot$, a gas temperature $T \sim 10^7 - 10^8$ K, and a total (bolometric) luminosity $L \sim 10^{43-45}$ erg s^{-1}. The number density of clusters in the local universe is about $n_{\rm cl} \approx 10^{-5}h^3$ Mpc^{-3}. Approximately 10% of galaxies are in clusters, the rest being "field" galaxies. Although "clusters of galaxies" were first discovered optically, by the, well, clustering of galaxies, galaxies comprise a small fraction of the total mass; most of the mass is actually dark matter (i.e., a collisionless component that is not specifically identified), and most of the rest is in hot gas spread throughout the cluster. It is worth keeping in the back of your mind that the inference of some of these quantities is not a straightforward thing, and that, e.g., getting the mass of a cluster can be tricky. In fact, **Ask class:** what are ways in which cluster mass could be inferred? Typically one gets the mass from (1) motions of galaxies within the cluster, assuming they are bound, (2) the temperature of the gas, and (3) gravitational lensing. The approximate agreement of all of these is encouraging.

Cosmology

As we said earlier, clusters have many cosmological implications. To understand them better, we'll skim the surface of a few important aspects of cosmology.

The first, and most important, thing is that the universe is expanding. As you may remember, it was this prediction of general relativity that caused Einstein to balk and introduce the cosmological constant. The line element for the expanding universe can be written in many ways, one of which is

$$ds^2 = -dt^2 + \frac{dr^2}{1 + kr^2/R^2} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\ . \tag{1}$$

Here $k = -1, 0,$ or 1 and the angular part indicates that this is a universe with spherical symmetry. In this expression R is the radius of the universe, and is time-dependent. $k = 0$ indicates a spatially "flat" universe, which will barely expand forever, $k = -1$ indicates a "closed" universe that will eventually recollapse, and $k = 1$ indicates an "open" universe, which will expand forever with some room to spare. Incidentally, the time-dependence of R means that there is no global energy conservation in an expanding universe. If this shocks and horrifies you, note that the stress-energy tensor *does* still obey the conservation law $\nabla \cdot T = 0$, so it's not as if we have abandoned all such laws! One of the fundamentally useful quantities in cosmology is the redshift z, which is defined as $1 + z$ equaling the ratio of the measured wavelength to the wavelength in the emitter's rest frame.

Much about the evolution and fate of the universe can be encompassed in just a few numbers. These are usually described in terms of their present-day contributions. It is, for example, of interest to know if the universe will expand forever or ultimately recollapse. For this, one can define three related parameters. One is the total mass density of nonrelativistic matter relative to a critical density, $\Omega_m = \rho/\rho_{\rm crit}$. If there is nothing else in the universe but ordinary matter, $\Omega_m < 1$ means an open universe, $\Omega_m > 1$ means a closed universe, and $\Omega_m = 1$ means a flat universe. Similarly, we can define the radiation energy density divided by the critical density as Ω_r, the neutrino energy density divided by the critical density as Ω_ν, and so on. We can also define a curvature parameter Ω_R, such that $\Omega_R > 0$ means an open universe, $\Omega_R < 0$ means a closed universe, and $\Omega_R = 0$ means a flat universe. Finally, one can define a similar contribution from the cosmological constant, Ω_Λ (or more generally from dark energy). The sum of all of these is unity, $\Omega_m + \Omega_r + \Omega_\Lambda + \Omega_R + \ldots = 1$, at any redshift. At high redshift, regardless of what Ω_m, Ω_r, Ω_Λ, $\Omega_R, \ldots$ are now, the universe acts as if $\Omega_R \approx 0$. Moreover, at $10 \lesssim z \lesssim 1000$, Ω_m is the most important component, at $z \gtrsim 10000$ Ω_r is the most important component, and there is a transition in between; Ω_Λ is only important at $z \lesssim 1$. These orderings are important to remember, because they simplify things dramatically at redshifts $z \gtrsim 10$.

Another important constant is the Hubble constant, which measures the current rate of expansion of the universe. This, by the way, is a misnomer, because the Hubble constant isn't constant with redshift; at any given age of the universe, it is constant with location.

The effects of the expansion of the universe are many. One of them is something we encountered briefly in the gravitational lensing lecture: whereas at low redshifts distance can be measured in many independent ways that all agree with each other (e.g., luminosity distance, angular diameter distance, proper distance, and so on), at high redshifts they diverge from each other and you have to be very careful about which one you're using at a given time. For example, when people report the evidence for dark energy from supernova observations, they're using a luminosity distance. This is defined as the distance one would have to go in Euclidean space in order to make an object appear as faint as it does at the given redshift. Even though the proper distance (i.e., measured with a ruler from you to the object of interest) approaches a constant value as $z \to \infty$, the luminosity distance does not, because due to redshifts the flux drops as $(1 + z)^{-2}$ (one factor from the lowered energy of the photons, one from the lowered rate of reception of photons). The effect on flux can be expressed with the distance modulus. For an Einstein-de Sitter universe, in which $\Omega_m = 1$ and $\Omega_\Lambda = \Omega_R = 0$ currently (and for all time), the distance modulus is

$$m - M = 25 + 5\log_{10}\left[6000(1 + z)(1 - (1 + z)^{-1/2})\right] - 5\log_{10} h \ . \qquad (2)$$

Recall that the absolute magnitude M is as measured at a distance of 10 pc, and that 5 magnitudes larger means fainter by a factor of 100. Using this, one can figure out the distance if the true luminosity is known. When $\Omega_\Lambda \neq 0$ the expression is changed, and it is the apparent deviation from the $\Omega_m = 1$ form that is the evidence for a nonzero cosmological constant. The current best guess is that presently $\Omega_m \approx 0.3$, $\Omega_\Lambda \approx 0.7$, and $\Omega_R = 0$. Remember, though, that although the cosmological constant may dominate now, at higher redshift the matter contribution is most important. In fact, with increasing z the matter part scales like $\Omega_m(1 + z)^3$, whereas the cosmological constant term is just Ω_Λ (both need to be normalized so that $\Omega_m + \Omega_\Lambda + \Omega_R = 1$, and if $\Omega_R = 0$ now it was always so and will always be so).

Our final bit about general cosmology is dark matter. One of the grand successes of the Big Bang model is that it predicts the abundances of light elements. Hydrogen, deuterium, helium 3 and 4, lithium 6 and 7, and beryllium were produced in the first few minutes, and their relative primordial abundances are all related to just one parameter: the fractional contribution Ω_b of baryons relative to the critical density. Measurements of primordial abundances are tough, but the best current estimate is that $\Omega_b \approx 0.02\, h^{-2}$ (with an average physical density of 3.6×10^{-31} g cm^{-3}). Since the overall matter budget is $\Omega_m = 0.3$, this means that 80-90% of matter must be something other than baryons. Other evidence suggests that this matter must be effectively non-self-interacting and non-luminous, and is therefore called "dark matter". One must, however, be careful, because the evidence for all this, though highly suggestive, is far from a done deal.

Structure formation

But what does all this have to do with clusters, you may ask? Patience, we're getting there! Clusters are among the largest gravitationally bound systems in the universe, so we'd like to know how they formed. This brings up the more general issue of structure formation: evidence is that early in the universe, the universe was astonishingly uniform. Now, of course, it isn't (we are more than 30 orders of magnitude denser than the average density of the universe). How did this happen?

Let's think first about gravitational collapse in a region where there's no funny business about an expanding universe. Suppose we also ignore pressure, so nothing can stop the collapse. Then a slight overdensity will lead to collapse. **Ask class:** what, approximately, will be the time scale of the collapse? It will be comparable to a free fall time scale. Since the free fall timescale goes like $\sqrt{r^3/GM}$, as the region collapses (i.e., r decreases) the time scale decreases, and there is a rapid and exponential collapse.

But what happens in an expanding universe? **Ask class:** qualitatively, what are the differences? Then, if the overdensity is slight, so that $\delta\rho/\rho \ll 1$, the matter tries to collapse but the universe expands from under it. The result is that the fractional overdensity increases (because, relative to the background density, an overdense region won't expand as fast), but as long as $\delta\rho/\rho \ll 1$, the locally measured size of the region actually *increases.* This is a very different situation from collapse on a static background. It is useful in this and many other contexts to introduce a scale factor for the size of the universe, $a(t)$, which is normalized to be unity in the present day. Note that $a \propto (1+z)^{-1}$. Then, for linear collapse ($\delta\rho/\rho \ll 1$), $\delta \propto a$. That is, the fractional overdensity of a linear region increases proportionally to the size of the universe. This is much slower collapse than it would be in a static universe, and it means that the initial size of perturbations at $z \approx 1000$, where radiation decoupled from matter, had to be of the order of 10^{-5} on angular scales of a few degrees, otherwise gravitational instability would not have had time to form structure at z of a few.

The cosmic microwave background, which comes from $z \approx 1000$, is a wonderful record of the early days when all perturbations in the universe were linear. This is very simple to treat compared to the current nonlinear universe. Note that, as in the case of black holes, "simple" does not mean "mathematically trivial". It just means that there are a limited number of variables one must consider, so it is possible to do a more mathematically rigorous and (in principle) trustworthy set of analyses than in the current universe, where things are really complex. The information contained in the CMB is potentially a goldmine, which is why there is so much current effort directed toward measuring it. Generally, an important thing to remember is that (with lots of details!) the initial fractional amplitude of perturbations at small length scales tends to be greater than at large length scales. Therefore, all else being equal, small things form before big things in this picture.

At some point, structures will become nonlinear ($\delta\rho/\rho \gtrsim 1$). Their further development depends on a number of factors. Ignoring everything else, when a mass concentration becomes nonlinear, the further collapse proceeds more quickly than before because the universal expansion is of diminishing importance compared to the contraction, so when a perturbation becomes strongly nonlinear the collapse is almost the same as it would have been on a static background. Therefore dark matter, if it is cold and pressureless, will continue contracting for a while. The baryons, however, may or may not. **Ask class:** what could stop the contraction for baryons? Their self-pressure may be able to stall contraction. This happens if the mass of the region is less than the Jeans mass, which is the mass at which a sound wave can cross the region faster than the free-fall time (i.e., compression by gravitational contraction creates an acoustic compression wave, which tries to expand; equivalently, it is the mass such that the total energy is negative [the negative gravitational potential energy outweighs the positive thermal energy]). Numerically, in a region with temperature T and number density n, the Jeans mass is

$$M_J \approx 5 \times 10^7\ M_\odot (T/10^4\ \mathrm{K})^{3/2} (n/1\ \mathrm{cm}^{-3})^{-1/2}\ . \tag{3}$$

In the early universe, from $z \sim 1000$ to $z \sim 50-100$, the baryons have a temperature tightly coupled to the temperature of the cosmic microwave background, $T_r = 2.7(1+z)$ K, and thus in that era the Jeans mass was independent of redshift, and had a value of $M_J \sim \mathrm{few} \times 10^5\ M_\odot$. At lower redshifts, the matter temperature decreased below T_r and thus the Jeans mass decreased to a few times $10^4\ M_\odot$. The first structures may therefore have had masses comparable to this mass. Later, larger structures formed. The first structures are thought to have formed at redshifts ~ 30, but larger, galactic-mass, structures probably needed to wait until $z \sim 10$ (probing this era is a major goal of studies using the James Webb Space Telescope). Clusters, with masses $\sim 10^{14-15}\ M_\odot$, formed significantly later yet, and their formation and history relates to the values of Ω_m and Ω_Λ, because unlike the formation of smaller objects, clusters formed at a time when dark energy was dynamically important.

The last point about structure formation is that even dark matter does not collapse to a point. **Ask class:** even if the dark matter does not interact with itself in any way but gravitationally, what might stop the collapse? It will have some relative motion or rotation, so it collapses until it virializes, that is, until orbital velocities are comparable to radial velocities. This probably happens when the overdensity $\delta\rho/\rho \sim 200$. When this does happen, it appears likely that the first highly nonlinear structures to form are not spheres, but are instead flattened: Zeldovich "pancakes". It is thought that these pancakes then collapse along another axis to form spindles, and finally along that axis to eventually reach equilibrium as approximate spheres. There are numerous large-scale simulations that explore such structure formation extensively.