DARK MATTER IN COSMOLOGY

ANTHONY AGUIRRE
Department of Physics
University of California at Santa Cruz
1156 High St.
Santa Cruz, CA 95064 USA
E-mail: aguirre@scipp.ucsc.edu

The last two decades in cosmological research have been an exciting time, and produced an exciting product: we now have in hand a "standard model" of cosmology. While several aspects of this model remain mysterious, its predictions are in remarkable accord with a vast range of observational data. A key aspect of this model, and one of the aforementioned mysteries, is the dark matter: a cold, collisionless consituent of the universe with $\sim 30\%$ of the cosmic energy density. In this article I broadly review the standard cosmological model, and the role and place of (non-baryonic) dark matter in it.

1. Introduction

Since the lectures in this volume were given, there has been great progress of our understanding of the role of, evidence for, and constraints on, dark matter. While we still have no real idea what dark matter is (and indeed must now postulate a new "dark energy" component of unknown nature as well), a rather precise and increasingly well-tested (and testable) picture of dark matter's role in cosmology has emerged.

While our understanding of all of the issues discussed in the Jerusalem summer school lectures has been advanced, I will focus on the topic which has perhaps advanced the most, and received least attention in the original lectures: the role of cold dark matter (CDM) in the formation of large-scale structure and galaxies. This is a vast subject and I hope here only to give an overview of the "big picture" and indicate directions for further study. Likewise, I have made no attempt to make comprehensive references; I have instead given for most subjects a few references that I find particularly seminal or useful.

I will first review in Sec. 2 the initial conditions for the standard cosmological model, and outline our theoretical understanding of the role of dark matter in structure formation. I will then discuss the confrontation of this theory with observations of the Cosmic Microwave Background (CMB) (Sec. 3), the Lyα forest and the large-scale distribution of galaxies (Sec. 4). Finally I will address the general picture of galaxy formation in CDM cosmology in Sec. 6.

2. Dark matter and Structure formation

2.1. *Initial conditions and the standard cosmological model*

The current standard model of cosmology posits that at a very early time, the universe was nearly homogeneous and isotropic, radiation-dominated, and nearly flat. Its geometry is thus described by the Friedmann-Robertson-Walker (FRW) metric

$$ds^2 = dt^2 - a^2(t)[dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)], \tag{1}$$

where $a(t)$ is a scale factor evolving according to

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\bar{\rho} + \bar{p}) \tag{2}$$

in terms of averages of the density ρ and pressure p. Two galaxies at small fixed *comoving* separation Δr will have physical separation $d = a(t)\Delta r$, and move apart physically at a rate $v = \dot{a}\Delta r = Hd$, where $H = \dot{a}/a$ is Hubble's constant. The observation of this relation led, of course, to the development of the big-bang cosmology.

Deviations from homogeneity are described by a random variable $\delta(\vec{x}, t)$, defined as

$$\delta(\vec{x}, t) \equiv \frac{\rho(\vec{x}, t) - \bar{\rho}}{\bar{\rho}}, \tag{3}$$

where $\vec{x}$ are *comoving* coordinates (like r, θ, ϕ in the metric) which are fixed for a particle at rest with respect to the cosmic fluid.

These perturbations are generally assumed to be Gaussian, i.e. the Fourier modes

$$\delta_{\vec{k}}(t) \equiv \int d^3\vec{x} \exp(i\vec{k}.\vec{x})\delta(\vec{x}, t) \tag{4}$$

at fixed t are each described by a Gaussian probability distribution of zero mean and variance σ_k (note that k is a *comoving* wavenumber and has units of inverse length). When computing statistical properties of δ for large volumes we can approximate[29] $\sigma_k^2 \simeq |\delta_k|^2$; the latter is often referred to as the power spectrum and taken to have a power-law form:

$$P(k) \equiv |\delta_k|^2 = Ak^n. \tag{5}$$

Such a power spectrum can be translated into a more physically suggestive measure by integrating $|\delta_k|^2$ for modes below the inverse of some length scale r; then the variation in mass M within a sphere of radius r goes as

$$\frac{\Delta M}{M} \propto M^{-(n+3)/6}. \tag{6}$$

Theoretically, n could take a number of values: $n = 4$ is maximal in that the non-linear but momentum-conserving dynamics of particles on small scales would build a large-scale $n = 4$ "tail" to the power spectrum; $n = 2$ would result from randomly throwing particles down within equal-mass cells, while $n = 0$ (Poisson fluctuations) would correspond to just randomly throwing particles down (see, e.g., Peacock[30]). The *observed* value for small k is $n \simeq 1$, which is called "scale-invariant" because the perturbation to the Newtonian gravitational potential is equal on all length-scales; it also has the pleasing property that the perturbation amplitude on the scale of the cosmological horizon is always the same.[31]

This described flat, homogeneous universe with $n = 1$ Gaussian density perturbations is widely thought to have resulted from a period of inflation in the early universe (see the article by Press & Spergel in this volume), though in principle some other process could give rise to it – the reader is encouraged to look for one!

The remainder of the cosmological model is then specified by describing the material and energetic contents of the cosmic fluid at some early time. This is conveniently and conventionally done in terms of the ratio Ω_i of the ith species' present day energy density to the current critical energy density $\rho_{\mathrm{crit},0} \equiv 8\pi G/3H_0^2$, where H_0 is the current Hubble constant. Extrapolation of each density component to a smaller $a(t)$ (e.g., $\rho \propto a^{-3}$ for pressureless matter) then gives each energy density at earlier times.

Current observations (to be described below) indicate that our universe contains $\Omega_r \sim 10^{-5}$ in radiation, $\Omega_b \simeq 0.04$ in baryons, $\Omega_{\mathrm{dm}} \simeq 0.23$ in cold, collisionless, non-baryonic particles (i.e. the Dark Matter), and $\Omega_{\mathrm{DE}} \simeq 0.73$ in some yet-more-enigmatic substance called "Dark Energy" with $p \simeq -\rho$. The repeated postulation of mysterious substances is tolerated by most cosmologists only because of the striking success of the theory these postulates engender.

2.2. *Evolution of perturbations*

Understanding the growth of the perturbations δ_k rigorously is an intricate subjects requiring a careful treatment of perturbation theory in General relativity; see Padmanabhan[29] for a detailed treatment. It can, however, be understood at two less rigorous but more tractable levels.

The first is somewhat heuristic (though in fact it can be made relatively precise).[31] Consider a density perturbation δ of comoving scale λ in a matter component (i.e. baryonic or dark matter) during a time when the universe can be considered to be dominated either by radiation or by pressureless matter. This can be thought to describe an over/underdense sphere of radius $\sim \lambda$ embedded in a uniform FRW-universe of density $\bar{\rho}$. Birkhoff's Theorem (the relativistic generalization of Newton's "spherical shell" theorem) indicates that the embedding space can be ignored and the sphere treated as an independent universe.[a] If $\delta > 0$ its expansion will slow relative to the outer region so that its density relative to $\bar{\rho}$ (i.e. δ) will increase if the inner and outer regions are compared at a later time in such a way that the expansion rates are equal.[b] Working this out reveals that $\delta \propto a^2$ during radiation domination ($a \propto t^{1/2}$), and $\delta \propto a$ during matter domination ($a \propto t^{2/3}$). If λ exceeds the horizon length λ, this analysis captures much of the dynamics, since on these scales different fluid components cannot evolve separately, and pressure support cannot prevent the growth of perturbations (instead, pressure adds to the source term for Einstein's equations).

In either a matter- or radiation-dominated epoch, the the horizon grows as t and hence faster than a, so any perturbation of fixed comoving scale will eventually enter the horizon if the epoch last sufficiently long. At this point two new effects become important. First, pressure: prior to horizon entry the sound crossing time across a perturbation ($> \lambda/c$) always exceeded the

---

[a]Actually, it states only that a spherically symmetric vacuum solution to Einstein's equations is the Schwarzschild solution; but it can safely be interpreted in this more liberal way.

[b]Perturbation theory in general relativity can be tricky because of the "gauge" ambiguity in choosing a surface of constant time; e.g. for small perturbations one can always choose a surface in which the universe is homogeneous (see Press & Vishniac[33] for an amusing presentation of some of these issues.) This difficulty can be overcome by carefully choosing a fixed gauge[33] or by working in carefully chosen gauge-independent variables.[2,18]

dynamical time $\sim t$; now it will not if

$$\lambda < \lambda_J \equiv \sqrt{\pi}\frac{c_s}{(G\rho)^{1/2}}, \tag{7}$$

where c_s is the sound speed (or velocity dispersion, in the case of collisionless particles) of the medium. This leads to a minimal "Jeans mass", dependend upon the temperature and density of the medium, below which fluctuations cannot grow. Second, perturbations in different fluid components – such as matter and radiation, or collisionless and collisional matter – may grow at different rates. For example, dark matter, which interacts negligibly with radiation and baryonic matter, can – and in some situations do – grow even if baryonic perturbations are supported against collapse by their pressure. However, even a pressureless perturbation cannot grow if its dynamical timescale $t_{\rm dyn} = 1/\sqrt{G\rho}$ is longer than the expansion timescale t (which is the dynamical time of the dominant fluid component).

Perturbations inside the horizon can be treated using the equations of motion for a fluid in an expanding universe (for those versed in General Relativity, these can be derived directly from the covariant conservation of the energy-momentum tensor: $\nabla_\mu T^{\mu\nu} = 0$) along with the weak-field version of Einstein's equations:

$$\nabla^2\Phi = 4\pi G(\rho + 3p/c^2), \tag{8}$$

where $\nabla\Phi$ gives the acceleration of a slowly-moving test particle. In a medium dominated by a non-relativistic fluid with sound speed $c_s \equiv \partial p/\partial\rho$, the analysis gives:

$$\ddot{\delta}_k + 2\frac{\dot{a}}{a}\dot{\delta}_k = \delta_k\left(4\pi G\bar{\rho} - \frac{c_s^2 k^2}{a^2}\right). \tag{9}$$

The appearance of the Jeans length can be seen in the r.h.s.: if the physical wavelength of the perturbation, a/k, does not exceed λ_J, the solution is oscillatory; otherwise the term in c_s can be neglected and using the fact that $\bar{\rho} \propto t^{-2}$, the solution splits into a growing mode $\delta_k \propto t^{2/3}$ and a decaying mode $\delta_k \propto t^{-1}$. For a radiation-dominated phase with pressure gradients neglected,

$$\ddot{\delta}_k + 2\frac{\dot{a}}{a}\dot{\delta}_k = 32\pi G\bar{\rho}\delta_k, \tag{10}$$

yielding two solutions $\delta_k \propto t^{\pm 1}$. For large λ the growing radiation- or matter-dominated solutions are in agreement with the heuristic model, but on smaller scales the behavior of each fluid component will depend crucially on whether or not the perturbation exceeds λ_J for that component, and whether the component dominates the expansion.

These considerations factor together into an overall picture as follows. In a hot big-bang, the universe is radiation-dominated until some time $t_{\rm eq}$. During this epoch, perturbations outside the horizon grow as $\delta \propto a^2$. Upon entering the horizon, perturbations in the radiation and baryons are held up by pressure and fail to grow (in fact they oscillate with near-constant amplitude). The dark matter, being both cold and collisionless, would "like" to grow as $\delta_{\rm dark} \propto a$, and is prevented only by the rapid expansion (dominated by the radiation); the perturbations turn out to grow, but only logarithmically ($\delta_{\rm dark} \propto \ln a$.)

At $t_{\rm eq}$, matter begins to dominate the expansion, and $\delta_{\rm dark} \propto a \propto t^{2/3}$. Baryons, however, are still coupled to the radiation; this provides strong pressure support (i.e. a high Jeans mass) so that the perturbations in the baryons, like those in the photons, cannot grow but instead oscillate (as discussed in more detail below). During this epoch the dark matter perturbations can thus grow substantially relative to those in the baryons, a fact which will be of great importance.

Finally, at some time $t_{\rm dec}$, the rate of collisional ionization become too low to maintain the ionization of the baryonic fluid, and the nuclei and free electrons combine to form atoms. With few free electrons, the baryons decouple from the photons, the baryonic Jeans mass drops drastically, and baryonic perturbations can subsequently grow as $\delta_{\rm baryon} \propto a \propto t^{2/3}$ on scales $\lambda \gtrsim \lambda_J$ (the latter now being determined by the baryonic pressure rather than that of the photons); on smaller scales they continue to oscillate. In fact, due to their growth during $t_{\rm eq} < t < t_{\rm dec}$, the dark-matter perturbations are now larger than those in the baryons. The baryons can then "fall in" to the existing dark-matter perturbations; thereafter their amplitudes will be equal for $\lambda \gtrsim \lambda_J$, where λ_J is now calculated using the combined baryonic and dark matter density.

The resulting power spectrum of the matter (baryons+dark matter) after decoupling carries the imprint of this earlier epoch, and this imprint is encapsulated in the "transfer function" T_k:

$$T_k \equiv \frac{\delta_k(z=0)}{\delta_k(z)D(z)}, \tag{11}$$

where $\delta_k(z)$ is the power spectrum at some very early z before which any relevant perturbation had entered the horizon, and $D(z)$ is the "linear growth function" $D(z)$, which is a general expression for the linear growth of perturbations in a homogeneous background, absent effects such as free-streaming or pressure support: $D(z)/D(z_0) = \delta(z)/\delta(z_0)$ for some reference redshift z_0. This is given by[12]

$$D(z) = \frac{5\Omega_m}{2}(1+z_0)g(z)\int^z \frac{1+z'}{g^3(z')}dz',$$

$$g^2(z) = \Omega_m(1+z)^3 + \Omega_\Lambda + (1-\Omega_0-\Omega_\Lambda)(1+z)^2.$$

Figure 1. Transfer function T for standard cosmological model (top), and power-law index of T (bottom).

A full and precise calculation of T_k must be done numerically, and several publicly available codes for doing so exist[28]. However, there are approximations that are sufficiently good for many purposes.[3,16,12] The key features are exhibited in Fig. 1, which shows the transfer function for cold dark matter computed using Eisenstein & Hu.[12] At large scales (small k) it is constant and the power spectrum at late times exactly reflects the "primordial" $n \simeq 1$ power spectrum. At large k, corresponding to scales below the horizon size at T_{eq}, T falls as k^{n_T} so that $P(k) \propto k^n T^2(k) \propto k^{n+2n_T}$, with n_T decreasing with k to $n_T \simeq -2$ at the smallest scales (see bottom panel of Fig. 1). If the ratio Ω_B/Ω_{DM} is fairly small, the main effect of baryons on this transfer function is to change the effective density of the dark matter in a scale-dependent way; taking this into account for $\Omega_B/\Omega_{DM} = 0.044/0.226$ and $h = 0.71$ gives T_k as shown in Fig. 1 (top panel).

The power spectrum at decoupling can be directly connected to several important cosmological phenomena. First, the power on large scales continues to grow $\propto a$ and describes the large-scale distribution of matter in the universe, as reflected for example in the distribution of galaxies on scales $\gtrsim 10\,\mathrm{Mpc}$. Second, on smaller scales the perturbations provides the seeds for the nonlinear collapse of the fluid into galaxies and clusters. Third (less important for the universe but more important for our knowledge of it), the perturbations leave a detailed imprint in the CMB. The signature of dark matter in all three of these phenomena is important and testable, and will be discussed in the next several sections.

3. Tests and constraints from the microwave background

Even before the confirmation of the precise thermal spectrum of the CMB by COBE, it was recognized that anisotropies in the observed CMB temperature would provide a snapshot of the density inhomogeneities that existed the time when the universe first became transparent to photons at $t_{\rm dec}$ (often also called the "recombination time" or "time of last scattering"). A number of excellent reviews of the physics of CMB anisotropies have been written; see, e.g. Hu & Dodelson[15]; here I will qualitatively review the basics, then focus on the role of dark matter.

The observed anisotropies of the CMB may be divided into *primary* anisotropies, which result from density fluctuations at recombination, and *secondary anisotropies* that are imprinted during later cosmological evolution. Primary anisotropies result from three main effects: the gravitational red/blueshift of photons emerging from potential wells, the Doppler shift of photons emitted from a medium with an inhomogeneous velocity field, and the lower-temperature emission of photons from regions that are overdense and hence recombine slightly later. These three effects are often termed, respectively, the "Sachs Wolfe effect", the "Dopper effect", and "intrinsic fluctuations." Since in linear theory the velocity field can be directly related to the density field, all three effects essentially capture the imprint of density inhomogeneities on the last-scattering "surface".

The analysis of the observed CMB generally proceeds by the decomposing the temperature $T(\theta, \phi)$ into a sum of spherical harmonics Y_{lm} of amplitude a_{lm}, and computing the angular power spectrum

$$C_l \equiv \frac{1}{2l+1} \sum_{m=-l}^{m=l} |a_{lm}|^2.$$

Roughly speaking, C_l gives an estimate of the power on angular scales $\approx 180/l$ degrees.

This angular power spectrum contains a multitude of information about the physics and constituents of the the universe when the the fluctuations were imprinted at $z \approx 1100$, as well as some information about the subsequent evolution of the universe. The latter come primarily from the angular diameter distance $d_A(z)$ (defined as D/θ, where θ is the angle subtended by an object of physical size D at redshift z), which connects physical scales at the recombination epoch to angular scales in the observed CMB anisotropies. This distance measure contains an integral over redshift that involves the energy densities of all energy components (see Hogg [17] for explicit formulas and other distance measures). Since the physical size of the last scattering surface is known, this gives one constraint on the energy densities that turns out to be quite sensitive to curvature; it is from this measurement[40] that we now know that the universe is geometrically approximately flat ($|\Omega - 1| < 0.05$).

While the angular diameter distance to $z \approx 1100$ sets the overall scaling for C_l, because of the complicated interplay (described in Sec. 2.2) of different components in the evolution of density perturbations between when the enter the horizon and when they are imprinted in the CMB, the measured power spectrum also yields information on, among many other things, the prevalence of dark matter and baryons. Consider first a scenario without dark matter. When fluctuations in the baryon density enter the horizon (or, really, the horizon expands to encompass them), they are pressure supported and oscillate until decoupling, when the CMB is last scattered. At that time the largest scale just able to compress (before rarefying due to pressure) imprints extra power at that scale; this leads to a peak in the power spectrum on a physical scale of the horizon at recombination, or an angular scale of about 0.5 degree. Higher harmonics are represented by a series of peaks at higher-l; the second peak corresponds a scale of maximum rarefaction. (The anisotropies measure the amplitude, rather than the value of the density fluctuations, which is why this shows up as a peak.) Higher-l peaks alternate between compression and rarefaction. With no further effects accounted for, the CMB would resemble a flat line at $l \lesssim 50$ connected to a squared sinusoidal curve at $l \gtrsim 50$. Due, however, to damping at high-l due to the finite thickness of the last scattering surface, and the ability of photons to diffuse out of small-scale density wells, the peaks fall off steadily in amplitude as l increases.

Collisionless dark matter changes this picture by adding a compression component that has no restoring force, so that the compression modes have greater amplitude than the rarefaction modes. With sufficient dark matter content, there is thus a fall in peak amplitude from the first to second peak, then a *rise* to the third peak. In addition, the additional compressing force makes the peaks somewhat narrower. Such effects provide a signature of dark matter, and their nature can be seen visually by inspecting some of the many reviews of CMB physics in which the cosmological parameters are varied.[15]

A detailed comparison of the WMAP data to a suite of models [44] shows that even with all other parameters left to vary, $\Omega_{\rm DM}h^2 = 0.10 \pm 0.02$ is required, and when only 6 "standard" parameters are free, $\Omega_{\rm DM}h^2 = 0.12^{+0.02}_{-0.02}$.

It is worth commenting here on possible alternatives to the dark matter hypothesis such as that propounded by Milgrom (this volume), in which gravity is modified so as to become stronger at small acceleration scales. Are these now ruled out by the CMB? Perhaps, but it is not as yet entirely clear. It is nearly impossible to see how such models would account for an alternation of peak amplitudes except by extreme luck or contrivance, but the data regarding the third peak is (as of this writing) insufficiently precise to warrant iron-clad conclusions. For the first two peaks, an alternative to dark matter might hope to reproduce the CMB by positing that gravity is unmodified at early times, so the only difference from the standard scenario is the absence of collisionless matter. In this case "no-CDM" models can be generated that provide a quite good qualitative fit to the observed power spectrum.[24] Still, the data on the first two peaks is very high quality, and it does not appear possible to fit it without a substantial contribution by massive neutrinos[44,24] (which would, of course, be non-baryonic dark matter, albeit of a familiar type).

A second test of the presence and importance of dark matter is in the connection between the CMB power spectrum and the power spectrum of galaxies that are the result of structure formation at later times. A given cosmological model provides a precise translation between the power spectrum of initial perturbations, and the power spectrum of matter at late times $z < 4$, to which we now turn.

4. Tests and constraints from the Ly-α forest and Distribution of Galaxies

Long before the CMB anisotropies were ever observed, attempts were made to understand the large-scale distribution of matter using the distribution of galaxies, and to use this information to infer the mechanism of structure formation. The primary tool used has, again, been the power spectrum $P(k)$. For galaxies it it useful to relate this to the more easily measured 2-point correlation function

$$\xi(\vec{r}) \equiv \langle \delta(\vec{x}) \delta(\vec{x} + \vec{r}) \rangle, \tag{12}$$

which can written in Fourier-transformed terms in a statistically isotropic universe as:

$$\xi(r) = \frac{V}{(2\pi)^2} \int P(k) \frac{\sin kr}{kr} 4\pi k^2 dk, \tag{13}$$

where V is a volume over which the averaging in Eq. 12 is done. This can be inverted to yield an expression for the power spectrum.

Galaxy surveys can give a measurement of the distribution of the cosmic (mainly dark) matter, but only to degree that either galaxies (or light from galaxies) trace mass, or that the "bias" in the relation between light and mass is independently known. The study of the power spectrum of galaxies has become quite mature with the completion of substantial parts of the SLOAN and 2dF galaxy surveys[45,13]. These surveys find that bias depends on galaxy type, but is near unity when averaged over all galaxies. Further, both surveys derive a mass power spectrum on comoving scales $\sim 20 - 200\,\mathrm{Mpc}$ that is a good match to the prediction obtained by evolving the CMB-measured power spectrum forward in time to the present epoch, using standard cosmological theory. This provides an excellent consistency check on the standard cosmological model. These comparisons are now being used to probe the primordial power spectrum over a wide range of length-scales. [44]

A second way of measuring the mass power spectrum on scales $\lesssim 10\,\mathrm{Mpc}$ that has recently been developed uses absorption spectra of high-redshift quasars. These spectra are filled with a "forest" of Lyα absorption features caused by density fluctuations in the highly-ionized IGM at $z \sim 1 - 4$. Theoretical arguments[37] and numerical simulations[10] indicate that there is a tight correlation between Lyα absorption and the density of the absorbing gas, so the correlation function of absorption in a quasar spectrum can be rather directly converted into a 1-d power spectrum of the intergalactic medium.[7] On large enough scales this is expected to closely track the dark matter distribution, and so gives an independent mass power spectrum on rather small scales. Encouragingly, this power spectrum agrees fairly well with the galaxy power spectrum where they overlap, and with the forward-evolved CMB-inferred primordial power spectrum (although there are tantalizing hints that the combination of CMB and Lyα forest data may call for a non-standard primordial spectrum[40]). Two groups are now undertaking a detailed study of the Lyα power spectrum using the SLOAN data.

Although the power specra from galaxy surveys and Lyα absorption are excellent tools for studying dark matter, structure formation, and their interplay, they do not directly address the original purpose for which dark matter was proposed, and in which the behavior of dark matter is least well-understood: its role in galaxy formation and evolution.

For this, we must return to the story of structure formation to the point at which an initially overdense perturbation forms a collapsed dark matter halo that can host a galaxy.

5. Dark Matter and galaxy formation

5.1. *Halo formation*

Once a perturbation grow sufficiently large, it will separate from the background cosmic expansion and collapse to a self-gravitating "halo". Under the assumption of spherical symmetry, this can be seen to happen at a time when the overdensity as calculated by linear theory reaches a critical level Δ that depends (weakly) on the background geometry and the constituents of the fluid; in a universe of only dark matter, $\Delta = 1.69$.

5.2. *the Halo mass function*

A truly accurate calculation of the properties of dark matter halos requires direct N-body simulation. Such calculations are well-developed and publicly available codes exist (such as GADGET [42]) that can numerically evolve the cosmological dark matter distribution from a very early time through the age of galaxy formation.

A basic understanding of the distribution function of halo masses can, however, be gained from a simple model pioneered by Press & Schechter [32] that, as it turns out, provides a surprisingly good characterization of the mass function of halos found numerically. In this approach, it is assumed that a structure of mass M collapses when a density perturbation smoothed over that mass scale, as calculated by linear theory, reaches a critical density δ_c. In this picture every overdense region will collapse eventually, and the probability P that a randomly chosen point will be in such a fluctuation of mass M is just

$$P(\delta > \delta_c; M) = \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{\delta_c}{\sqrt{2}\sigma(M)}\right)\right],$$

because $\sigma(M)$, as described in Sec. 2, is just the width of a Gaussian probability distribution governing the density contrast in regions of mass M. There are then two more steps. First, it is assumed that the *under*dense gas simply accretes onto the collpsed halos, the net effect of which is just to double $P(\delta > \delta_c; M)$ (although just a fudge by Press & Schechter, this is justified in the more rigorous approached mentioned below). Second, it is reasoned that we should "attribute" each mote of dark matter to the most massive collapsed region of which it is part: $2P(\delta > \delta_c; M)$ then describes the fraction of the dark matter that is incorporated into halos of mass $> M$. This is then a cumulative probability function that can be differentiated[30] to yield a mass function $f(M)$, where $f(M)dM$ is the comving number density of halos of mass between M and $M + dM$:

$$Mf(M) = \frac{\rho_0}{M}\left|\frac{d\ln\sigma}{d\ln M}\right|\sqrt{\frac{2}{\pi}}\exp(-\nu^2/2). \tag{14}$$

Here $\nu \equiv \delta_c/\sigma(M)$, and ρ_0 is the comoving dark matter density.

This function takes the form of a power law with an exponential cutoff at high-mass, and as such provides the potential for accurately fitting luminosity functions of galaxies, which tend to have this "Schechter function" form. However, the observed luminosity function exhibits an approximate behvior of $f(M) \propto M^{-1}$, whereas Eq. 14 gives $f(M) \propto M^{(n-9)/6}$, using Eq. 6. Since on the scale of small galaxies, $n < -2$, the predicted number of very small halos greatly exceeds the number of observed faint galaxies. This points to the "satellite problem" that has been greatly discussed as one of the challenges for the CDM paradigm in galaxy formation; but of course there is no assurance that small halos should necessarily all form galaxies around them.[c]

[c]The predicted dark matter halo distribution and the galaxy luminosity functions are

Although the Press-Schechter approach often suffices for everyday use, there are more advanced treatments[5,38] (still short of direct simulation) that provide greater accuracy, and can yield additional information such as (a statistical description of) the merger history of halos.

5.3. *Halo profiles*

Numerical studies[27] have shown that within the standard CDM model, halos collapse to a nearly universal form with a spherically-averaged density profile of

$$\rho(r) = \frac{\rho_c}{r[1 + (r/r_c)^2]}, \tag{15}$$

where ρ_c is the central density and r_c is a "core radius". These parameters can be expressed in various combinations of the virial radius r_v (the radius within which the mean density is, say, $200\bar{\rho}$), the concentration parameter $c \equiv r_v/r_c$, and the halo mass M_v within r_v. The physical reason that the halo density profile takes this universal form is unclear, though some attempts to derive it analytically have been made.

The profile of Eq. 15 has some features of interest for the theory of galaxy formation. First, the corresponding circular velocity profile is approximately flat for $r \sim r_c$ (as one would hope, in order to explain the flat rotation profile of observed galaxies), but slowly falls off at large-r. Second, it contains a rather steep central cusp of $\rho \propto r^{-1}$. The exact slope of this central cusp has been a matter of some debate and consternation,[27,23,35] because a number of (particularly small, low surface-brightness) galaxies appear to have rotation curves inconsistent with such a central cusp [11,43]. Although a very active area, the issue of whether the fault for this discrepancy lies in the CDM predictions, the CDM model itself, or in the accuracy of the observations is at present still rather unresolved.

5.4. *Angular Momentum*

Also key for galaxy formation is the angular momentum $\vec{J}$ of the collapsing halo, the magnitude of which can be expressed in dimensionless terms as

$$\lambda \equiv \frac{|\vec{J}|E^{1/2}}{GM_v^{5/2}}, \tag{16}$$

also discrepant at the *high*-mass end; this can be seen very clearly in clusters which do *not* resemble gigantic galaxies.

where E is the binding energy of the object[30]; $\lambda = 1$ would correspond to angular momentum fully supporting the halo against collapse. In numerical simulations[46], $\ln \lambda$ is found to be distributed normally among halos, with $\langle \ln \lambda \rangle \approx -3.2$ and $\sigma \approx 0.5-0.6$. The angular momentum is thought to result from tidal torquing by nearby perturbations during collapse, but may also have contributions from the accretion of smaller halos during subsequent evolution.[48]

5.5. *From a dark halo to a galaxy*

The canonical description of galaxy formation after the collapse of a halo goes as follows. The gas in the halo, unlike the dark matter, can dissipate energy by cooling. This leads to a contraction of the gaseous halo to a radius $\sim \lambda r_v$. At this point, the halo is supported by angular momentum and cannot contract further in the plane perpendicular to the total angular momentum vector. This leads to the formation of a thin, axisymmetric disk with a density profile that is determined in part by the initial angular momentum distribution of the gaseous halo. During the collapse process, the increased baryonic density at small radii tends to contract the dark halo, potentially enhancing even further the steep density profile predicted by simulations.

What next happens to the disk depends upon what instabilities exist in the disk density structure. A disk without a dark halo is highly unstable to the formation of a bar – indeed this was another early argument for dark halos – and may form one even in a halo's presence. On a local level, density perturbations in the disk are stabilized by shearing in the disk (which manifests in an $r-$dependence of the angular rotation speed Ω) and thermal pressure. They can grow only when the dynamical time of some region characterized by a surface density Σ is shorter than both the sound-crossing time and the shearing time. This gives rise to the Toomre $Q-$parameter[4]

$$Q \equiv \frac{v_s \kappa}{\pi G \Sigma}$$

where $\kappa^2 = \frac{d}{dr}\Omega^2 + 4\Omega^2$, and v_s is the sounds speed. For $Q > 1$, local perturbations are unstable against growth, and star formation can presumably proceed.

5.6. *Current status of galaxy formation theory*

The simple picture just outlined neglects an enormous set of complicated physical process that play a part in galaxy formation. Two particularly important ones are, first, that halos accrete and collide with other halos and second, that energy released from star formation affects the physics of the gas.

A great amount of work has been performed to attempt to treat these and other complicated processes to assemble a reasonably comprehensive picture of galaxy formation that can be compared to galaxy observations. There are two basic approaches in this project. In the numerical approach, numerical simulations including gas dynamics are evolved from an early time to produce an *ab-initio* calculation of galaxy properties today. In the "semi-analytic" approach in, simplified prescriptions for physics such as gas cooling, star formation, and feedback from stellar energy release are added to already-completed dark-matter-only simulations (or, in simpler models, extensions of the Press-Schechter approach) to produces a set of statistical predictions for galaxy properties.[d] In both approaches predictions can be made for the luminosity and mass function of galaxies, the global star formation history, and other observables.

Overall, both programs have met with a great deal of success. Many of the observed bulk properties of galaxies, as well as trends in those properties with time or galaxy mass, are reasonably reproduced. This would by no means be assured in any alternative to the standard CDM model. Nonetheless, there are several outstanding difficulties in the details comparison of CDM galaxy formation theory to observations that are sufficiently severe that they have led some theorists to contemplate modifying the dark matter properties or abandoning the notion of dark matter altogether.

5.6.1. *Outstanding problems, and Alternatives to (Cold) Dark Matter*

The first possible problem was mentioned in Sec. 5.2: CDM theory predicts a number of low-mass halos that is much larger than the number of low-mass galaxies we observe. This had long been noticed in semi-analytic galaxy formation model,[19,39] in which feedback was invoked to reduce the small-halo abundance. It was made more acute when simulation groups produced dark-matter simulations meant to resemble the Milky Way halo

[d]Due to limited resolution, numerical simulations also must add parametrized prescriptions for processes such as star formation.

and discovered $\sim 100-1000$ simulated dark-matter satellites, compared to only ~ 10 detected satellite galaxies.[20,22]. This problem, however, has a number of quite plausible solutions – there is no particular reason to believe that very small halos should have stars, and good reasons to believe they should not: cosmic radiation after reionization could evaporate them, and feedback could blow away their star-forming gas. In addition, the "missing" dark matter subhalos may now be showing up observationally in the form of flux-ratio anomalies in multiply-images lensing systems[8] which indicate dark substructure in galaxies, with approximately the density predicted in CDM theory.

The second potential problem was mentioned in Sec. 5.3: dark matter halos are expected to have a steep $\rho \propto r^{-1}$ density cusp in their centers, yet observed dark matter density profiles, as inferred from the dynamics of the central gas and stars, tend to be better fit by a model with a constant-density core, and are often outright incompatible with an r^{-1} cusp. [11] The status of this problem is still not entirely clear, though the observers and simulators are rapidly improving their results and understanding to what degree there is conflict.

The third potential problem concerns the halo angular momentum discussed in Sec. 5.4. Early calculations showed that $\lambda \sim 0.05$, typical for galaxy halos, would lead to a disk size comparable to observed spiral galaxy disks *if* the angular momentum of the gas was strictly conserved. [50] This nice general idea, however, breaks down when implemented in more detail. First, numerical simulations of galaxy formation find disks that tend to be far too small. This is believed to occur due to tansfer of angular momentum from the gas to the dark matter, but it is as yet unclear whether this is a correct physical effect or an artifact of limited numerical resolution.[26] Semi-analytic models also have difficulties, in that the net amount of angular momentum in halos is approximately right (*if* conserved) but the distribution of angular momentum in simulated dark halos, if applied to the gas and conserved parcel-by-parcel, does not lead to an exponential disk. [6] Like the density profile, this problem is subject of significant current attention[34].

A final potential problem, slightly harder to crisply define, concerns the systematic properties of galaxies. Despite the array of complicated and stochastic effects expected to be integral to galaxy formation (e.g., merger, starbursts, galactic feedback and winds, environmental effects, random formation times, etc.), the properties of spiral galaxies appear to be remarkably regular. For example, the Tully-Fisher relation between luminosity and asymptotic rotation speed is compatible with being nearly exact – i.e.

the scatter in the relation could plausibly be entirely observational error. [47]. In more detail, the simple relation proposed by Milgrom (this volume) as a formulation of modified gravity fits the systematics of galaxies extraordinarily well – given the observed gas/star density profile, the observed rotation curve can be accurately predicted using only at most one free parameter.[36] If CDM theory is correct this requires a very tight (and probably not-quite-understood) coupling between the visible and dark matter.

This rash of problems initially provoked a number of proposed modifications of dark matter, e.g. to make it slightly warm,[14,49] or self-interacting.[41] The idea of all of these was to reduce the small-scale structure – whether in halo cores or in tiny subhalos. These models appear to have fallen largely out of favor, partially due to the ameliorization of the "subhalo problem", and partly because modifying the cores of dwarf galaxies (where rotation curves are well-measured) without significantly altering the density profile in the cores of clusters is difficult.[9,21] Also, because it probes small scales, the Lyα can place direct constraints on dark matter itself; if dark matter particles were light enough that they were not completely cold, the free streming of the particles would erase small-scale power in the density field. Current measurements from the Lyα forest put a lower bound of $\sim 750\,\mathrm{eV}$ on the dark matter particle mass.[25]

The final, rather radical alternative to CDM theory that bears mentioning is MOND (see this volume), conceived as a modification of gravity that would obviate the need for dark matter. As mentioned above, MOND has great success in accounting for the observed dynamics of galaxies and aacounting for the systematics of galaxy properties. However, because it is not a full theory, it is far less predictive than CDM: the CMB anisotropies, large-scale galaxy distribution, galaxy mass function, etc., cannot be reliably calculated and tested, and in some cases where MOND does make some firm predictions (cluster dynamics, absence of substructure in galaxies, etc.) it run somewhat afoul of observations.[1] The success of MOND seems to be pointing to something, but whether or not it points to the need to modify gravity and banish dark matter, I leave to the reader to decide.

6. conclusions

Although much has happened in astronomy and astrophysics since the 1986 Jerusalem Winter school, several things remain the same. First, the fundamental nature of the dark matter is, as of this writing, still completely unknown. Its elucidation ranks as one of the foremost tasks in astrophysics, and given the enormous effort currently being put forth by many observational and experimental groups, we have a reasonable hope that it may be forthcoming in the relatively near future. Second, although many techniques of studying the dark matter have become somewhat more sophisticated in detail, they are the same in their basic structure, and rely on the same basic physics, as when the following lectures were compiled nearly two decades ago. The student will, therefore, find a great deal in the following chapters that will help build a foundation for understanding a range of topics in modern astrophysics and cosmology.

[[proofread]]

References

1. Aguirre A., astro-ph/0310572
2. Bardeen J. M., *PhRvD* **22**, 1882 (1980).
3. Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., *ApJ* **304**, 15 (1986).
4. Binney, J., & Tremaine, S. *Galactic Dynamics.* Princeton, NJ: Princeton University Press (1987).
5. Bond J. R., Cole S., Efstathiou G., Kaiser N., *ApJ* **379**, 440 (1991).
6. Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A., Porciani C., Primack J. R., *ApJ* **555**, 240 (2001).
7. Croft R. A. C., Weinberg D. H., Pettini M., Hernquist L., Katz N., *ApJ* **520**, 1 (1999).
8. Dalal N., Kochanek C. S., *ApJ* **572**, 25 (2002).
9. Dalcanton J. J., Hogan C. J., *ApJ* **561**, 35 (2001).
10. Davé R., Hernquist L., Katz N., Weinberg D. H., *ApJ* **511**, 521 (1999).
11. de Blok W. J. G., *RMxAC* **17**, 17 (2003).
12. Eisenstein D. J., Hu W., *ApJ* **511**, 5 (1999).
13. Hawkins E., Maddox S., Cole S., et al., *MNRAS* **346**, 78 (2003).
14. Hogan C. J., Dalcanton J. J., *PhRvD* **62**, 063511 (2000).
15. Hu W., Dodelson S., *ARA&A* **40**, 171 (2002).
16. Hu W., Sugiyama N., *ApJ* **444**, 489 (1995).
17. Hogg, D., astro-ph/9905116
18. Hu W., Seljak U., White M., Zaldarriaga M., *PhRvD* **57**, 3290 (1998).
19. Kauffmann G., White S. D. M., Guiderdoni B., *MNRAS* **264**, 201 (1993).
20. Klypin A., Kravtsov A. V., Valenzuela O., Prada F., *ApJ* **522**, 82 (1999).
21. Miralda-Escudé J., *ApJ* **564**, 60 (2002).
22. Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., *ApJ* **524**, L19 (1999).
23. Moore B., Quinn T., Governato F., Stadel J., Lake G., *MNRAS* **310**, 1147 (1999).
24. McGaugh S., astro-ph/0312570 (2003).
25. Narayanan V. K., Spergel D. N., Davé R., Ma C., *ApJ* **543**, L103 (2000).
26. Navarro J. F., Steinmetz M., *ApJ* **478**, 13 (1997).

27. Navarro J. F., Frenk C. S., White S. D. M., *ApJ* **490**, 493 (1997).
28. Zaldarriaga M., Seljak U., *ApJS* **129**, 431 (2000).
29. Padmanabhan T., *Structure Formation in the Universe.* Cambridge, UK: Cambridge University Press (1993).
30. Peacock J. A., *Cosmological Physics.* Cambridge, UK: Cambridge University Press (1999).
31. Peebles P. J. E. *Principles of Physical Cosmology.* Princeton, NJ: Princeton University Press (1993).
32. Press W. H., Schechter P., *ApJ* **187**, 425 (1974).
33. Press W. H., Vishniac E. T., *ApJ* **239**, 1(1980).
34. Primack J. R., astro-ph/0312547.
35. Primack J. R., astro-ph/0312549.
36. Sanders R. H., McGaugh S. S., *ARA&A* **40**, 263 (2002).
37. Schaye J., *ApJ* **559**, 507 (2001).
38. Sheth R. K., Mo H. J., Tormen G., *MNRAS* **323**, 1 (2001).
39. Somerville R. S., Primack J. R., *MNRAS* **310**, 1087 (1999).
40. Spergel D. N., Verde L., Peiris H. V., et al., *ApJS* **148**, 175 (2003).
41. Spergel D. N., Steinhardt P. J., *PhRvL* **84**, 3760 (2000).
42. Springel V., Yoshida N., White S. D. M., *NewA* **6**, 79 (2001).
43. Swaters R. A., Madore B. F., van den Bosch F. C., Balcells M., *ApJ* **583**, 732 (2003).
44. Tegmark M., et al., astro-ph/0310723 (2003).
45. Tegmark M., et al., astro-ph/0310725 (2003).
46. van den Bosch F. C., Abel T., Croft R. A. C., Hernquist L., White S. D. M., *ApJ* **576**, 21 (2002).
47. Verheijen M., *Ap&SS* **269**, 671 (1999).
48. Vitvitska M., Klypin A. A., Kravtsov A. V., Wechsler R. H., Primack J. R., Bullock J. S., *ApJ* **581**, 799 (2002).
49. White M., Croft R. A. C., *ApJ* **539**, 497 (2000).
50. White S. D. M., Rees M. J., *MNRAS* **183**, 341 (1978).