TOWARDS AN UNDERSTANDING OF DARK MATTER:
PRECISE GRAVITATIONAL LENSING ANALYSIS
COMPLEMENTED BY ROBUST PHOTOMETRIC REDSHIFTS

by

Daniel Aaron Coe

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

September, 2007

© Daniel Aaron Coe 2007

All rights reserved
Chapter 1

Dark Matter

These days are heralded as the “Golden Age” of Precision Cosmology. But in the future, we may look back and refer to these as the “Dark Ages” of Astronomy. For years now, astronomers have struggled with some basic questions: the age of our universe and its eventual fate. We now believe we have answered these questions, but in doing so we have found that our universe is darker and more mysterious than we had ever imagined.

As we explain below, we have good reason to believe that there is about six times as much matter in the universe as that which we can see. The rest, which we call Dark Matter, is not made of the same material as galaxies and you and I, but instead of particles which have yet to be discovered by physicists. As you can imagine, discovering the Nature of this 80+% of our universe is widely recognized as the most pressing question facing both astronomers and particle physicists today. This thesis is dedicated to helping answer that question.

1.1 The Golden Age of Precision Cosmology

For years astronomers were in the dark about the age and fate of our universe. Different analyses yielded different results, the most embarrassing of which suggested that the universe was younger than the oldest stars observed [Freedman et al. 1994, Jacoby 1994]. Solving these questions can be reduced to determining the expansion history of our universe. If we know how fast the universe has been expanding throughout time, then we can trace this expansion back to when the universe began as nearly a single point in the Big Bang. This gives us the age of our universe.

The first step in this equation involves determining the present expansion rate of our universe, known as the Hubble constant, or \( H_0 \). Since Hubble’s initial measurement in 1931 and through the 1990’s, estimates of the Hubble constant varied greatly between 50 and 100 km/s/Mpc. Finally a concentrated effort was made in the Hubble Space Telescope Key Project to measure \( H_0 \) to within 10%. Recession velocities were compared to distances measured to relatively nearby galaxies.
Their final determination was \( H_0 = 72 \pm 8 \) km/s/Mpc (Freedman et al. 2001). And thus we were halfway toward determining the age and fate of our universe.

The other key ingredient in our equation is the mass density in our universe. Ever since the Big Bang (and after a rapid growth spurt known as “inflation”) the universe is believed to have expanded uniformly to be slowed only by gravity, or more specifically, the mutual gravitational attraction of massive bodies. The more mass in the universe, the more the expansion has slowed down over time. If there is enough mass in the universe (a “critical density”), the expansion of the universe will eventually halt and turn around resulting finally in a “Big Crunch”, which would resemble the “Big Bang” if the film were played in reverse. Not enough mass in the universe and our universe would continue to expand forever. Eventually objects would be so far apart that structures would stop forming, gas would stop condensing to form stars, and the final result would be a “Big Freeze”. Whichever we believe would come to pass, the “Big Crunch” or “Big Freeze” would not be scheduled until many billions of years into the future. Yet in spite of no immediate danger, many were compelled by a fundamental desire to learn what our ultimate fate would be.

Perhaps the best measure of the mass density of our universe comes from observations of Large Scale Structure (LSS). The amount of structure observed in the universe depends strongly on the amount of mass available to seed its growth. This analysis led us to infer a mass density of \( \Omega_m \sim 0.3 \), or just 30% of the “critical density” (Percival et al. 2001) given the above value for \( H_0 \)). Similar results were obtained by other analyses including observations of the frequency of strong galaxy-galaxy lensing events (Chiba & Yoshii 1999; Chae et al. 2002) and mass estimates of galaxy clusters (Borgani et al. 2001). With a mass density lower than critical for the universe, we were ready to conclude that the “Big Freeze” seemed inevitable. We could also estimate the age of the universe. But such a low density flew in the face of inflation theory which predicts the total density should be \( \Omega = 1 \), or equal to critical. And inflation theory was about to receive some very strong supporting evidence. It would take energy density from an unlikely source to restore the total density to \( \Omega = 1 \).

Observations of the Cosmic Microwave Background (CMB) are singularly impressive as direct photographs of the universe when it was a mere 300,000 years old at an epoch known as Recombination, well before the formation of the first stars and galaxies. (For comparison, the earliest images we have of galaxies date back to when the universe was about a billion years old.) Before Recombination, the universe was opaque as a thick fog of electrons prevented photons (particles of light) from traveling far before colliding and scattering off in new directions. Meanwhile, these electrons were eagerly trying to combine with protons to form atoms. But these unions wouldn’t last long before a photon would smash into them, breaking the pair apart. Finally at Recombination, the photons had cooled enough that they were no longer able to disrupt these mergers. With electrons tied up in atoms, photons were now able to stream free through the universe. The vast majority have done so ever since, traveling unimpeded for billions of years. A few completed their long journey in one of our microwave detectors thus giving us our view of the CMB.
What we see in our CMB observations are photons with a very uniform temperature on all areas of the sky. This is a bit of a paradox, as such distant regions of the universe should never have been in physical contact with one another; thus there is no reason for them to appear similar. The widely accepted resolution of this paradox comes from inflation theory (Guth 1981), which explains that the universe experienced a rapid exponential growth spurt in its first $10^{-34}$ s of existence. It is now believed the newborn universe was minted as a pea-sized “instanton” of nearly uniform temperature, which quickly burst to cosmological size via inflation. This explains how opposite ends of the universe can appear so similar. Inflation theory also provides neat explanations for the finer details observed in the CMB, including the Gaussianity and scale-invariance of observed fluctuations. Thus the CMB provides strong support for inflationary theory which, as mentioned above, implies that the total density of the universe is critical, or $\Omega = 1$. But analysis of the CMB also gave more strong evidence that $\Omega_m \sim 0.3$ [de Bernardis et al. 2000].

The strange reconciliation of this dilemma would come from observations of distant supernovae (Riess et al. 1998; Perlmutter et al. 1999). These observations were aimed at more directly measuring the expansion rate of the universe over time. As in the HST Key Project to measure $H_0$, distances would be compared to recession velocities. But this time the galaxies studied would be more distant, thus probing the expansion rate of the universe at earlier times. They expected to find that the expansion of the universe was faster in the past, that since the Big Bang it has been slowing with gravity. Instead they found the opposite: that the expansion of the universe has been accelerating recently. Our understanding of cosmology was suddenly turned on its head.

To say we have an “explanation” for this result would be too generous, but our best guess is that some unknown form of energy adds a “springiness” to empty space that counteracts gravity on large scales. The amount of this supposed energy fits nicely with our theory of inflation, as a value of $\Omega_\Lambda \sim 0.7$ was measured in the supernova studies. This restores the total mass/energy density of the universe to our preferred value of $\Omega = 1$. But this new energy was coined “Dark Energy” by Michael Turner (Huterer & Turner 1999), as it is at least as mysterious as the “Dark Matter” which had already boggled astronomers for over 60 years (see discussion beginning in the next section).

Thus we astronomers have proclaimed this the age of “precision cosmology” in which we have measured to great precision the relative quantities of matter and energy in the universe. (However see Lieu 2007 and references therein, which raise doubts about the certainty of our current cosmological model.) Various complementary observations, including the CMB (Spergel et al. 2007), LSS (Percival et al. 2001; Tegmark et al. 2004; Cole et al. 2005), and distant supernovae (Riess et al. 2004; Astier et al. 2006), have all converged on a “concordance cosmology” which gives values for densities of Normal Baryonic Matter $\Omega_b = 0.042 \pm 0.002$, Dark Matter $\Omega_{DM} = 0.20 \pm 0.02$, and Dark Energy $\Omega_\Lambda = 0.76 \pm 0.02$. The observations also put tight constraints on the Hubble Constant $H_0 = 74 \pm 2 \text{ km/s/Mpc}$ (often given in units of 100 km/s/Mpc: $h = 0.74 \pm 0.02$). Putting this information together gives us the age of the universe, $t_o = 13.7 \pm 0.15$ billion years (now comfortably
older than the oldest observed stars). And given the low matter density of the universe along with
the recent acceleration of its expansion, a “Big Freeze” seems unavoidable as our eventual fate. But
to say we have everything figured out would be to overlook the 95% of the universe made up of Dark
Matter and Dark Energy. In fact we know very little about this great Dark Side of the universe. But
a large effort among both astronomers and particle physicists dedicated to resolving these mysteries
is currently underway.

1.2 Evidence for Dark Matter

Aside from the cosmological evidence for Dark Matter on large scales given above, the
existence of Dark Matter is also inferred in galaxies and galaxy clusters. The first evidence for Dark
Matter was found in 1933 by Fritz Zwicky. By measuring the orbital velocities of eight galaxies in
the Coma Cluster, Zwicky inferred the total mass necessary to hold these galaxies in their orbits.
This mass turned out to be about 50 times higher than that observed based on the brightness of the
galaxies. He thus reasoned that a large amount of unseen Dark Matter must be present.

The analysis required to make this conclusion is not entirely assumption-free. The mass
necessary to hold galaxies within their orbits is given by

\[ v^2 = GM/r, \]  

where \( v \) is the orbital velocity, \( r \) is the radius of the orbit, \( M \) is the mass within \( r \), and \( G \) is
Newton’s gravitational constant. In practice, only one component of the velocity is measured:
the velocity toward or away from us. This velocity is detected by precise measurements of the
galaxy “redshifts” (see Chapter 4). Measurement of velocity across our sky is rendered small and
undetectable by the great distance to these galaxies. To obtain the full velocity from measurements
of a single component, assumptions must be made about the orientations of the orbits. Galaxies
may have orbits of practically any orientation within clusters. But this makes assumptions about
their “average” orientations straightforward and fairly reliable.

The picture is much clearer in individual spiral galaxies, for which most gas and stars
have settled into regular orbits within the plane of the disk. Rubin & Ford (1970) measured the
orbits of gas and stars in the Andromeda galaxy (and later in 16 other galaxies, Rubin et al. 1985)
finding convincing evidence that the enclosed visible mass is insufficient to support these orbits. She
thus inferred the existence of Dark Matter “halos” (perhaps better referred to as “clouds”2) which

---

1This “age crisis” was also resolved in large part due to a reduction of the measured ages of globular clusters
(Chaboyer et al. 1998).

2The term Dark Matter “halo” has led to some confusion. The Oxford American Dictionary, for example, gives
one definition for “halo” as: “Astronomy – a tenuous sphere of hot gas and old stars surrounding a spiral galaxy”.
But rather than residing in a sphere outside a galaxy, a “halo” actually permeates and envelopes galaxies (and galaxy
clusters, alike) more like a cloud or fog.
pervade galaxies and extend well beyond their visible edges. Future observations of other galaxies and galaxy clusters have put these findings on even stronger footing.

Gravitational lensing analyses (which will be discussed more fully in Chapter 3) provide still more direct evidence for “missing mass” in galaxies and galaxy clusters. Gravitational lensing directly measures the distortion of space due to massive bodies according to Einstein’s Theory of General Relativity. These distortions provide an assumption-free measure of the amount of mass within a region. Results (e.g., Broadhurst et al. 2005a) show that galaxy clusters must contain on the order of a hundred or few hundred times as much mass as is visible, a discrepancy even greater than that observed by Zwicky. These “Mass-to-Light” ratios are lower for individual galaxies, on the order of tens or of just a few instead of hundreds.

Is this Dark Matter real, and if so, what could it be? As we discuss below, we do believe it is real, but have little idea what it is. We do have a few clues however. We believe that Dark Matter predominantly consists of diffuse swarms of particles that are collisionless, non-baryonic, and cold.

1.3 Modified Gravity or Collisionless Dark Matter?

Rather than invoking invisible matter, a much more natural explanation would be that our current understanding of gravity breaks down on the large scales of galaxies and galaxy clusters. Perhaps gravity is simply stronger on these scales than we think, and thus able to keep objects in their orbits as well as bending light to the degrees seen in gravitational lensing. This would not be the first time our law of gravity has been rewritten. A young patent clerk named Albert Einstein revolutionized our ideas about gravity with his Theory of Relativity. Along the way, he found that Newton’s law of gravity, which works well here on Earth, is not strong enough in the regions nearby larger masses. Perhaps a similar revolutionary idea could explain Dark Matter (and even Dark Energy) without resorting to yet-to-be-discovered particles.

Much work has been done to rewrite Einstein’s gravity, with MOND perhaps the most popular theory (Milgrom 1983; Sanders & McGaugh 2002; Bekenstein 2004). With a simple modification of the law of gravity, MOND claims, for example, to explain galaxy rotation curves without Dark Matter. Another recent theory of modified gravity is the MOG model (Moffat 2005, 2006). MOG claims to explain galaxy rotation curves (Brownstein & Moffat 2006b), X-ray cluster masses (Brownstein & Moffat 2006a), and the Pioneer 10/11 anomaly along with other observations within our solar system (Brownstein & Moffat 2006c).

But recently, observations have been obtained which seem to prove the existence of Dark Matter once and for all (Clowe et al. 2006, however see Brownstein & Moffat 2007). A pair of galaxy clusters known collectively as the “Bullet Cluster” collided with each other many years ago (Fig. 1.1). The individual galaxies appear to have passed through relatively unfazed (with few if any collisions). The gas of the two clusters however did collide creating shock fronts (observed in X-rays, one resembling a bullet) that lag behind the positions of the galaxies. A gravitational lensing
analysis is performed to map out the mass in the cluster, and we would expect to find mass at both
the positions of the galaxies and at that of the gas (as the gas, in total, is actually more massive than
the galaxies themselves). But instead we find that the vast majority of the mass is aligned with the
galaxies and not with the gas. It appears that Dark Matter does in fact exist, and that it has passed
through the collision unphased along with the galaxies. In fact, based on these observations, strong
upper limits can be placed on the collisional nature of Dark Matter: \( \sigma/m < 1 \text{ cm}^2/\text{g} \) Markevitch
et al. (2004) (with tighter constraints forthcoming from hydrodynamical simulations that attempt
to recreate this specific collision).

While offering strong proof for the existence of Dark Matter, these observations should not
be mistaken for disproving modified gravity. A modification of gravity may seem less necessary if
Dark Matter exists, but there is no reason that MOND/MOG and Dark Matter may both prove
correct. For example, Angus et al. (2007) suggest that a modified (stronger) gravity would make
it more likely to observe an event such as the Bullet Cluster, which appears to have collided with
a great rate of speed hard to account for in traditional theories of structure formation. But then
recent simulations of the Bullet Cluster using Dark Matter show that the collision speed is actually
lower than that observed of the propagating shock front (Springel & Farrar 2007). So the collision
speed of the Bullet Cluster may not be so anomalously high after all.

Meanwhile, Brownstein & Moffat (2007) insist that their modified gravity (MOG) can in
fact account for the offset between the observed mass and gas peaks without requiring any Dark
Matter. Unless Dark Matter particles can be detected some day in the lab and shown to exist
in Nature in the correct quantity to explain the “missing mass”, alternative theories will not only
persist, but should be given due consideration.

1.4 Non-baryonic Dark Matter, and MACHOs vs. WIMPs

On its surface, the idea of Dark Matter is not so hard to swallow, really. There are many
“Dark” objects in outer space, including planets, brown dwarfs, black holes, and all of the gas which
hasn’t condensed to the point of burning. None of these give off light, and are thus all “Dark”. You
and I would even qualify as “Dark Matter” under this simple definition. However we have strong
reasons to believe that the majority of Dark Matter in our universe is much more mysterious.

Searches were carried out to find large massive bodies in the Dark Matter halo of our own
Milky Way. Transit of such a MACHO, or MAssive Compact Halo Object, in front of a star in
our sky induces a brief and small but measurable brightening of that star in what is known as a
microlensing event. Such events were observed, but in much less frequently than that expected if
MACHOs make up all the Dark Matter in our galaxy (Alcock et al. 2000; Tisserand et al. 2006).
Thus the only remaining “normal” candidate for Dark Matter was gas in the interstellar medium
(ISM). But gas would be ruled out as well, as it would be shown that Dark Matter cannot be
anything “normal” at all.
Figure 1.1 The Bullet Cluster offers strong proof of the existence of Dark Matter (Clowe et al. 2006). X-ray concentrations (red) are offset from the mass (blue) inferred from gravitational lensing analysis and from the luminous galaxies (themselves aligned with the inferred mass). This is interpreted as the after-effect of a collision between the two clumps. Dark Matter and galaxies pass through the collision relatively unphased, but the gas has been stripped from the matter and lags behind. Claims against the existence of Dark Matter argue that gravity may simply be stronger than we believe. But if this were true, we would expect the mass in the Bullet Cluster to be found aligned to both the galaxies and the gas (the latter actually heavier in total). However, see Brownstein & Moffat (2007).
All of the objects we observe in our universe, including planets, stars, buildings, cars, you and I, we are all made of the same basic building blocks. Protons and neutrons (collectively known as baryons) make up the nuclei and the mass of our atoms, while lighter electrons orbit about the nuclei in the outskirts of the atoms. We can deduce the total mass of these baryons in the universe from observations of deuterium in distant hydrogen clouds. (Deuterium is the union of a proton and neutron atom. This “heavy hydrogen” was all formed in the primordial universe.) These observations combined with predictions of the ratios of basic particles in the early universe (Big Bang Nucleosynthesis theory), yield a baryon fraction $\Omega_b h^2 = 0.021 \pm 0.002$, or $\Omega_b = 0.041$ for $h = 0.72$ (O’Meara et al. 2001; Kirkman et al. 2003). Observations from the CMB independently confirm this result (§1). But this is far less than the total amount of mass in the universe: $\Omega_m = 0.24 \pm 0.02$ (derived independently from observations of both the LSS and CMB, as discussed in §1.1).

Thus the large remainder of mass must be in the form of some strange non-baryonic matter known as WIMPs, or Weakly Interacting Massive Particles. These are likely particles we’ve yet to discover. Or could WIMPs be particles we’ve only recently discovered to have mass?

### 1.5 Cold Dark Matter Simulations

Recently physicists discovered that one of the fundamental particles in Nature, neutrinos, which were long believed to be massless, do have mass after all (Fukuda et al. 1998). As neutrinos are highly abundant in the universe, could their newfound mass explain Dark Matter? Unfortunately the answer is no, as we find that neutrinos are not massive enough given their density in the universe to explain all the Dark Matter. But they do contribute some fraction. Kahniashvili et al. (2005) find $0.001 < \Omega_\nu < 0.05$, or at most one quarter of the Dark Matter ($\Omega_{DM} = 0.20 \pm 0.02$).

In fact this whole class of particles is ruled out as an explanation of Dark Matter by observations of Large Scale Structure. Such light particles formed in the early universe would have very high (relativistic) velocities, too high to “stay put” within an area the size of a galaxy. Fast-moving particles would thus begin forming structures with the largest most massive structures. These would then need to fragment to form the smaller galaxies we see today. But the structures produced in simulations of such Hot Dark Matter (HDM) don’t resemble those observed in our universe (White et al. 1983). Much more successful are simulations involving Cold Dark Matter (CDM).

The Millennium simulation is the state-of-the-art Dark Matter simulation today of Large Scale Structure formation in our universe, covering a larger volume in greater detail than ever before. Each Dark Matter “particle” in this simulation weighs $\sim 10^9 M_\odot$, or the mass of a small dwarf galaxy. Thus each “particle” is actually representative of many Dark Matter particles. But this mass resolution is unprecedented for the volume of this simulation, a cube 500 Mpc $h^{-1}$ on a side. After setting their initial conditions (positions and velocities derived from observations of the CMB and given our ΛCDM “concordance cosmology”), the Dark Matter particles are allowed to coalesce and form structure under the influence of their mutual gravitational attraction. Some
of the final results are shown in Fig. 1.2 and in Fig. 1.3 alongside real observations of Large Scale Structures in our universe (Springel et al. 2006). The agreement is impressive. Cold Dark Matter in a ΛCDM cosmology is very successful at reproducing the structures observed in our universe, at least on such large scales.

The baryons and gas necessary to form galaxies were not included in the Millennium simulation. (But rumor has it that a “Millennium plus gas” hydrodynamic simulation was completed last October 2006, and that the results are currently being analyzed.) The problem is that there is a lot of complicated and uncertain physics that goes into the condensation of gas, star formation, radiative cooling and heating, and “stellar feedback” (supernova ejection of material). Dark-Matter-only simulations are straightforward, perhaps largely due to our ignorance of Dark Matter, as massive particles are simply allowed to interact via gravity. These assumption-free results are thus very robust. The architects of the Millennium simulation decided to add galaxies only after the simulations were complete. Galaxies were assumed to form where Dark Matter had concentrated. Their properties were determined assuming recipes for star formation, etc. These recipes were then adjusted until the simulated galaxy populations matched well with those observed. We trust that the recipes found to work best were used to produce the new “Millennium plus gas” simulations yet to be released.

1.6 WIMP Dark Matter Candidates from Particle Physics

So far we have seen that Dark Matter particles are cold (travel much slower than light), collisionless (only interact via gravity), non-baryonic (made of different particles than ordinary matter), and, of course, “dark” (do not emit light, and thus do not interact via the electromagnetic force). No particle with such characteristics is known within the current Standard Model of particle physics. But theories beyond the Standard Model aimed at “grand unification” of all forces do include particles which exhibit the properties expected of Dark Matter. Hopefully such theorized particles will be produced in the next generation of particle colliders, such as the Large Hadron Collider (LHC). Depending on the particle, we may also expect to observe its signature directly in collisions with particles here on Earth and/or indirectly via particles produced by its self-annihilation in outer space.

Two of the most popular extensions of the Standard Model with viable Dark Matter candidates are supersymmetry (“SUSY”) and extra-dimensional (such as “string”) theories. In each case, the Dark Matter candidate is the theory’s “lightest particle”, which as such is stable, as it cannot decay to anything lighter. In the simplest SUSY theory dubbed MSSM (the Minimal Supersymmetric extension of the Standard Model), the lightest supersymmetric particles (LSP) are the neutralino and gravitino. In the Kaluza-Klein multi-dimensional theory, the lightest KK particles (LKP) are the first KK excitation of the boson B(1) and the neutrino ν(1).

Our probability of detecting these particles in the lab depends in large part on their masses.
Figure 1.2 Dark Matter structures produced in the Millennium simulation [Springel et al. 2005, reprinted from their Fig. 1]. Each frame shows a 15 Mpc $h^{-1}$ thick slice through the periodic simulation cube 500 Mpc $h^{-1}$ on a side. (The slices are taken at angles so as to avoid duplicating structures shown in the bottom two frames.) The frames zoom in by factors of 4 toward one of the galaxy cluster halos realized in the simulation.
Figure 1.3 Large scale structures observed in our universe (top and left) compared to those produced in the Millennium Simulation (bottom and right). Observed structures are from spectroscopic redshift surveys: the CfA2 \cite{GellerHuchra1989}, Sloan \cite{Gott2005}, and one half the 2dFGRS \cite{Colless2001}. All are plotted to the same scale in redshift versus Right Ascension space. At bottom and right are structures observed in realizations of the Millennium Simulation cut to the same physical areas and magnitude limits as the observational surveys. The densities, power spectrum, etc. of the Millennium Simulation agrees very well with that observed. But it bears mentioning that the recipes used to add galaxies to the Dark Matter simulations (see text) as well as the specific slices shown here were both specifically selected to provide the best possible matches to the observations. Reprinted from Fig. 1 of Springel et al. \cite{Springel2006}.
and their “cross sections” for interaction with regular matter. Until positive detections are confirmed, non-detections can constrain this parameter space and exclude different candidate particles. Some of the more popular Dark Matter candidate particles are plotted in Fig. 1.4. Collectively these candidate particles are known as “WIMPs”, or Weakly-Interacting Massive Particles. This name distinguishes them from an earlier Dark Matter candidate, MACHOs, or MAssive Compact Halo Objects. (As discussed in §1.4, searches for MACHOs via monitoring for microlensing events revealed that they can only contribute a small fraction to the Dark Matter in galaxy halos.)

1.7 Dark Matter Detection

In previous sections, we discussed methods of inferring Dark Matter’s existence by observational methods that measure mass and find it to be much greater than that visible in a given area. Now we turn to methods which can detect or produce specific types of Dark Matter particles. Three techniques are available: indirect detection, direct detection, and production. Indirect detection involves detection of the by-products of Dark Matter particle annihilation in space. Direct detection instead registers the collisions of Dark Matter particles in the halo of our Milky Way with other particles here “in the lab” on Earth. Finally, we hope to one day (if not soon) produce Dark Matter particles in the lab.

These different lines of analysis must work in concert. Dark Matter particle candidates produced in the lab, must be compared to those observed in space both directly and indirectly. If a satisfactory match is found, nucleosynthesis theory will be used to calculate the particle’s abundance in the universe. If this is then found to match that measured for Dark Matter $\Omega_{DM} = 0.20$, we may finally claim to have resolved this great mystery.

Extensive and recent reviews on these subjects may be found in: indirect detection (Carr et al. 2006), direct detection (Gascon 2005; Gaitskell 2004), production (Baltz et al. 2006).

1.7.1 “Indirect” Detection of Dark Matter Particles in Space

Dark Matter in outer space may not produce visible radiation, but by its self-annihilation, it may produce particles and/or radiation that we may detect here on Earth. The collision of two Dark Matter particles may yield neutrinos, gamma rays, and cosmic rays.

This method of observing Dark Matter is generally referred to as “indirect” detection to distinguish it from the “direct” detection of Dark Matter particles interacting with particles in labs here on Earth. However by this nomenclature, our detection of visible matter in space, namely baryons in stars, is just as “indirect”. When baryons join to form Helium atoms, we observe one of the products (photons) that is released. Thus the “indirect” detection of Dark Matter may rightly be thought of as seeing, if not the “light”, at least the radiation given off by Dark Matter. And in the cases of gamma ray and neutrino observations, we may even talk of directly imaging Dark
Figure 1.4 Some well-motivated WIMP-type particles which can be expected to contribute significantly to the Dark Matter density $\Omega_{DM} = 0.20$. Each particle’s theorized interaction cross section with normal matter $\sigma_{int}$ in units of picobarns ($1 \text{ pb} = 10^{-24} \text{ cm}^2$) is plotted versus the particle mass $m_X$ in units of GeV. The neutrino is plotted in red as it is “hot” and thus excluded as a CDM WIMP candidate. The box marked “WIMP” includes various candidates including those from Kaluza-Klein theories. Reprinted from Fig. 1 of Roszkowski (2004).
Matter concentrations, and perhaps lifting them out of the Dark.

Due to the weakly interacting nature of Dark Matter, a high density is required to produce enough collisions to yield a measurable flux of secondary particles. Thus our detection efforts so far have been aimed at the high concentrations of Dark Matter expected to reside in the centers of the Earth, Sun, our Milky Way, and one or two other nearby galaxies. The centers of our galaxy and of nearby galaxies are the preferred target for gamma ray searches, as gamma rays from the Earth and Sun would be absorbed before reaching our telescopes. However, our uncertain knowledge of the central density of galaxies precludes us from accurate predictions of the expected signal. (This provides further motivation for more accurate analysis of galactic rotation curves and improved simulations of Dark Matter galactic halos.) For this reason, the Sun yields the most promising signal for neutrino searches. (The Earth with its lower density would produce fewer DM annihilations and thus fewer neutrinos). Cosmic Rays, being charged particles unlike gamma rays and neutrinos, cannot be traced back to individual sources as their paths to us are inextricably altered by magnetic fields in the ISM. Nevertheless, an excess in positrons has been detected that may be a signal from DM annihilation (DuVernois et al. 2001). Gamma rays have also yielded a possible DM signal (discussed below), while neutrino searches have only yielded upper limits on fluxes thus far.

The main difficulty in such observations, is that all of the proposed particle by-products from Dark Matter annihilation are also produced by many other mechanisms in space, thus adding a large “background” to observed signals. For example, consider the observation of high-energy (∼1 MeV or greater) gamma rays by EGRET on board the Compton Gamma Ray Observatory. Most of this observed radiation is produced by the interaction of charged cosmic rays with the ISM (interstellar medium). Another contribution comes from inverse Compton scattering of cosmic ray electrons off photons. Still more is produced by electron Bremsstrahlung radiation in the ISM. When the expected contributions from all of these mechanisms are tallied and compared to observations, an excess is found at or near our galactic center (Bottino et al. 2004, see our Fig. 1.5). This has been interpreted tentatively as evidence of a signal from Dark Matter annihilation. Caution is mandated here by our lack of knowledge of the physics of cosmic rays and of our ISM which lead to uncertainties in the expected “background” level of gamma rays. Thus it is insufficient here to simply claim that any excess “must” be due to Dark Matter. (Note that this is essentially what is done in, say, a galactic rotation curve analysis, where the masses of luminous objects are known with much more certainty, and any excess mass inferred may all be attributed to Dark Matter.) Instead a specific Dark Matter signature must be predicted and observed, namely the amount and the spectral energy distribution required to make up the deficit of gamma rays. Some analysis suggests that a 40 GeV neutralino would produce the required signal, although the expected amplitude of this signal is highly uncertain as it depends on the Dark Matter density in the galactic center, which again is not well constrained. Finally, an arbitrary “boost factor” is required to bring the expected signal in line with the observations. Use of this boost factor has been justified by the expected clumpiness
of the Milky Way’s Dark Matter halo (the exact degree of which is again uncertain). Since the Dark Matter self-annihilation rate is proportional to the Dark Matter density squared ($\Gamma \propto \rho^2$), a halo including dense clumps will produce more self-annihilations than a smooth halo. Nevertheless, this suggestion of a 40 GeV neutralino provides particle physicists with a candidate to produce or perhaps rule out in the lab. As we will see, 40 GeV is definitely on the low end of that predicted for the neutralino. And if the neutralino’s energy is this low, it probably should have been produced by now in linear colliders.

1.7.2 Direct Detection of Dark Matter Particles in the Lab

Our Milky Way is bathed in a thick fog of Dark Matter particles. The Sun moves through this fog at about 240 km/s, with the Earth’s orbit around the Sun causing a seasonal variation of ±10% in this velocity. The great majority of these particles pass right through the Earth and everything on it, with perhaps a billion WIMPs going through your body every second. To catch such weakly-interacting particles, a wide and thick net must be cast, in the form of large blocks of liquid, gas, or crystals. The most optimistic expectations are that WIMPs will collide with the nuclei of regular matter at the rate of one per 100 kg of material per day. But this rate may instead be as low as one per ton per year! Present detectors typically use about a kilogram (and up to 250 kg) of specialized material to try and catch a few particles of the Dark Matter whizzing by.

If a WIMP does collide with a nucleus in the detector, the nucleus recoils releasing energy detected by methods including thermal phonons, ionization, and scintillation depending on the medium / instrument. But the expected signals may be mimicked by other objects colliding with nuclei in the detector, including stray neutrons, cosmic rays, gamma rays, and alpha and beta radiation. Thus detectors are placed deep underground (to block most cosmic rays), and enshrouded in lead (to block high energy cosmic rays) and polyethylene / water (to block the other contaminants). Some “background” events still leak through and must be either discarded based on their signature, or failing that, discarded statistically based on the number of background events expected.

The most tantalizing possible detection of Dark Matter was reported by the DAMA collaboration (Bernabei et al. 2000, 2003). In 1997, after a year of operation, DAMA detected a seasonal variation of flux which they attributed to the Earth’s motion around the Sun (which alternately increases and decreases the incoming flux of WIMPs). In 2002, after seven years of operation, this conclusion was confirmed and strengthened (see Fig. 1.6). Their data appears to be well explained by a Dark Matter particle of mass 52 GeV and cross section $7.2 \times 10^{-6}$ pb, where one picobarn (pb) = $10^{-12}$ barns (b), and one barn (b) = $10^{-24}$ cm$^2$. However, if WIMPs have this energy and cross section, then other experiments such as ZEPLIN (Uk Dark Matter Collaboration et al. 2005), EDELWEISS (Sanglard et al. 2005), CDMS (Akerib et al. 2006), and CRESST (Angloher et al. 2005) should have also detected them in significant quantities. Yet they have failed to do so. Instead, each of these experiments has detected only a handful (between zero and three) of possible
Figure 1.5 High-energy gamma ray spectrum detected from the galactic center by EGRET, plotted as crosses. The dashed line is an expectation of the background signals described in the text. The dotted blue line is 32 times the signal expected from annihilation of neutralinos with mass $m_X = 40$ GeV in the galactic center given an NFW fit to the Milky Way Dark Matter halo. The sum of these two signals is given as the solid red line. A neutralino signal “boost factor” such as 32 may be justified by clumpiness in the halo, while the specific value of 32 was chosen here to give the best fit to the observed data. Reprinted from Fig. 2 bottom of Bottino et al. (2004).
WIMP-nucleon collisions. From their analysis, they claim the WIMP-nucleon cross section can be no greater than $\sigma < 10^{-6}$ pb, in conflict with the results from DAMA. A detailed discussion of possible causes for this discrepancy can be found in §5.1 of Gaitskell (2004). Among the simplest explanations in favor of the DAMA signal is that some mechanism produces a larger cross section of WIMPs with the NaI used by DAMA rather than the Ge/Xe targets in other detectors. On the other hand, perhaps environmental seasonal variations, such as cavern temperature or ground water levels, are responsible for the annual modulation detection.

Aside from the DAMA results, most other experiments seem to be telling us that the WIMP-nucleon cross section is $\sigma < 10^{-6}$ pb. However, this has not yet begun to probe the predicted range of the WIMP-nucleon cross section $10^{-11}$ pb $< \sigma < 10^{-7}$ pb. The next generation of direct WIMP detection projects, including EDELWEISS-II, CDMS-II, and CRESST-II, hope to push the detection limit down to $\sigma < 10^{-8}$ pb. If the WIMP-nucleon cross section is within this range, then these experiments will detecting significant numbers of WIMPs worthy of study. If not, our study of WIMPs will have to wait until the following generation of detectors already being discussed, which hope to probe down to $\sigma < 10^{-10}$ pb. This should cover almost the entire expected range for the WIMP-nucleon cross section. If WIMPs are not detected by that point, then physicists may have to begin rethinking their theories.

1.7.3 Production of Dark Matter Particles in the Lab

Particle physicists are excitedly heralding the imminent discovery of new particles at particle accelerators, including the Higgs boson and theorized supersymmetric (“SUSY”) particles. The Higgs boson is the only Standard Model particle yet to be discovered, and is dubbed the “God particle”, as it is presumed responsible for giving mass to all other particles. Two-sigma detections of a $\sim 160$ GeV particle, possibly the Higgs boson, have already been reported at the Tevatron at Fermilab, currently the world’s highest energy particle accelerator. This energy scale of $\sim 160$ GeV is at the edge of the Tevatron’s detection (or rather, production) limits. If this signal cannot be confirmed at Fermilab, then it is expected to be confirmed at the higher energy Large Hadron Collider (LHC) scheduled to begin operation in November 2007.

Meanwhile, some theories of Dark Matter particles place their mass on the order of 100 GeV as well, meaning that their discovery may also be imminent. In fact, particle physicists’ theories of electroweak symmetry breaking predict the existence of Dark Matter particles with mass on this order. Independently, calculations of the particle’s expected mass from nucleosynthesis theory arrive at the same $\sim 100$ GeV conclusion (Baltz et al. 2006).

Assuming a Dark Matter particle is convincingly detected at the Tevatron or LHC, its measured mass will be compared with that returned (hopefully) from direct detection of Dark Matter particles in our galactic halo. If these masses are in agreement and find a reasonable home in extensions of the Standard Model, then the value for the mass will be plugged into nucleosynthesis
Figure 1.6 Annual modulation of WIMP-candidate detections collected by the DAMA collaboration over a seven-year period (1996-2002) plotted for three energy bins. This modulation is attributed to the motion of the Earth around the Sun, which increases (decreases) our velocity relative to the Milky Way's Dark Matter halo in the Summer (Winter). The maxima expected at ~June 2nd each year are plotted as dashed vertical lines while the minima are plotted as dotted vertical lines. This data appears to be well explained by a Dark Matter particle of mass 52 GeV and cross section $7.2 \times 10^{-6}$ pb. But this result is in conflict with those from other experiments which put the upper limit at for the cross section at $10^{-6}$ pb. Reprinted from Fig. 10 of Bernabei et al. (2003).
models to predict the current density of this particle in the universe. (Nucleosynthesis theory correctly returns the current densities of elements such as deuterium and lithium, and thus may be expected to do the same for Dark Matter particles.) By comparing this calculated density to the observed density $\Omega_{DM} = 0.20 \pm 0.02$ (see §1.1), we will determine whether this newly-discovered particle in fact makes a significant contribution to the Dark Matter in the universe.
Chapter 2

Dark Matter Halo Simulations

Will a detectable amount of gamma rays, neutrinos, or cosmic rays be produced by Dark Matter self-annihilation in our Milky Way? Will we be able to directly detect collisions of WIMPS and nucleons here in labs on Earth? The answers to these questions depends, respectively, on the density of Dark matter at the Milky Way’s center and that in our local neighborhood within the Milky Way. Expectations for these densities can be derived from detailed simulations of Dark Matter halos, or to be more precise, re-simulations at higher resolution of halos realized in simulations of the universe such as the Millennium simulation described earlier. These simulations probe a very wide range of physical scales, but are not yet able to resolve the very inner regions of galaxies well enough to accurately determine their central densities. Thus attention is paid to the central density slope of these simulated halos, from which the central density may be extrapolated. In simulation after simulation, a finite central slope has always been measured through the innermost resolved radius (§2.1, 2.2, 2.3, 2.4). If a finite slope persists all the way to the center, it suggests a central “cusp” with the central density diverging to infinity. This result is at odds with observations of galaxy rotation curves, which generally reveal a central “core” of constant density (§2.6). It may be that these central cores are simply too small to be resolved in simulations. Or perhaps baryons, absent from most Dark Matter simulations, are the key ingredient necessary to produce an inner “core” (§2.5). On the other hand, observations of galactic rotation curves may have been fooled by non-circular (i.e, elliptical) halos into seeing flat central cores where none exist (§2.7). One way or another this discrepancy must be resolved, both if we wish to estimate the central densities of galaxies, and to save our widely successful CDM model from the “CDM catastrophe” posed by this and other discrepancies.

Note that this chapter will focus specifically on halo profiles as derived in Dark Matter simulations. As we have already discussed in §1.5, comparison of large scale structure observed in simulations and in Nature have provided us with some of our most important clues about the nature of Dark Matter particles.
2.1 NFW Halo Profile

Early high-resolution Dark Matter simulations of galaxies and galaxy clusters concluded that structures of all sizes are well fit by a “universal” radial density profile (Fig. 2.1):

$$\rho(r) = \frac{\rho_s}{(r/r_s)(1 + r/r_s)^2},$$

(2.1)

This profile is now known simply by the authors initials NFW [Navarro et al. 1996]. The two fit parameters $\rho_s$ and $r_s$ were shown to be related and depend on a single parameter, the mass of the halo. A prescription for deriving $\rho_s$ and $r_s$ from the halo mass was provided in the Appendix of [Navarro et al. 1997] and will be discussed below. The relation is such that halos of lower mass (galaxies) have higher central concentrations than halos of higher mass (galaxy clusters). For a halo of a given mass, the central concentration is defined as

$$c_{\text{vir}} = \frac{r_{\text{vir}}}{r_s}$$

(2.2)

where $r_{\text{vir}}$, the virial radius, is the nominal size of the cluster as we define below, and the $r_s$ parameter marks the transition between the inner trend towards $\rho(r) \propto r^{-1}$ for $r \to 0$ and the outer trend towards $\rho(r) \propto r^{-3}$ for $r \gg r_s$. At $r = r_s$, the NFW profile has a slope of $\rho(r) \propto r^{-2}$, or that of an “isothermal” profile (that which gives rise to a flat rotation curve). Galaxies are found to have smaller $r_s$ than galaxy clusters relative to their sizes $r_{\text{vir}}$. This higher central concentration is attributed to the fact that galaxies formed earlier in the universe when the background density was higher. By the time larger galaxy clusters coalesce, the density of the universe is lower, thus leading to a lower central density concentration in galaxy clusters.

The virial radius $r_{\text{vir}}$ is supposed to designate the “edge” of the galaxy or cluster. Within $r_{\text{vir}}$, the halo is “virialized”, meaning that objects within have settled down into regular orbits with radii defined by the amount of halo mass within those orbits. Outside $r_{\text{vir}}$, objects are not in orbit, although nearby objects may still be falling in and accreting onto the halo. In reality, there is no sharp line dividing regions of orbiting and non-orbiting objects. And even if there were, distinguishing such regions would require prohibitively laborious observations. Meanwhile, galaxy clusters have either not yet virialized, or if they have, have done so only recently. In fact, these are the largest structures that have had time to form and virialize since the beginning of the universe. If galaxy clusters have not quite virialized, can we still obtain a reliable measure of their size? For a more readily available and consistent measure of $r_{\text{vir}}$, or halo size, we turn to theories of halo collapse.

Early theoretical work [Peebles 1980] predicted that a sphere of material will collapse if its density is 1.686$(1 + z)$ times that of the background. This overdensity factor depends on the
Figure 2.1 Density profiles of four Dark Matter halos simulated by Navarro et al. (1996, reprinted from their Fig. 3). The halos increase in mass from $\sim 3 \times 10^{11} M_\odot$ to $\sim 3 \times 10^{15} M_\odot$, and yet all are fit well by the functional form given in 2.1 plotted as solid lines. The arrows indicate the gravitational softening lengths of the simulations.
cosmology, with the value given initially calculated for an Einstein-de Sitter universe of \((\Omega_m, \Omega_\Lambda) = (1, 0)\). As such an overdensity collapses, its density will grow to \(\Delta_{\text{vir}} \sim 337\) times that of the background by the time it virializes given the present concordance cosmology \((\Omega_m, \Omega_\Lambda) \sim (0.3, 0.7)\) \cite{Bryan98}. (Note the background density of the universe is decreasing the whole while, thus contributing to this factor of 337.) Between virialization and the redshift at which we view the object, this density is assumed to not change much.

Thus for an observed (or simulated) cluster we can define the virial radius \(r_{\text{vir}}\) as that within which the halo mass equals \(\Delta_{\text{vir}} = 337\) times that of the background:

\[
M_{\text{vir}} = \frac{4\pi}{3} \Delta_{\text{vir}} \rho_m r_{\text{vir}}^3, \tag{2.3}
\]

where \(\rho_m = \Omega_m \rho_{\text{crit}}\) is the background mass density of the universe.

Note that the exact value of \(\Delta_{\text{vir}} = 337\) is not agreed upon in the literature. It is also common to find \(\Delta_{\text{vir}} = 368\). The difference is mainly a result of the different choice of \(\Omega_m = 0.268\) \cite{Spergel03}. A different approximation for \(\Delta_{\text{vir}}\) also plays a role, with the former derived from an expression given by \cite{Bryan98}:

\[
\Delta_{\text{vir}}\Omega_m \approx 18\pi^2 - 82\Omega_\Lambda - 39\Omega_\Lambda^2, \tag{2.4}
\]

and the latter from \cite{Eke98}:

\[
\Delta_{\text{vir}} \approx 178 \Omega_m^{0.45}. \tag{2.5}
\]

Both were derived for a flat universe \(\Omega_m + \Omega_\Lambda = 1\). (Also see \cite{Nakamura97} Eqs. C-18 and C-19, for a set of slightly different approximations.)

And the overdensity is often defined differently in the literature: in units of the critical density \(\rho_{\text{crit}}\), rather than in units of the background mass density \(\rho_m = \Omega_m \rho_{\text{crit}}\). Usually (but not always) in these cases, the overdensity is renamed \(\Delta_c\):

\[
M_{\text{vir}} = \frac{4\pi}{3} \Delta_c \rho_{\text{crit}} r_{\text{vir}}^3, \tag{2.6}
\]

with the two factors being related via \(\Delta_c = \Delta_{\text{vir}} \Omega_m\). Expressed in these units, the overdensity factor is given as \(\Delta_c = 101.1\) for \(\Omega_m = 0.3\) using equation \(2.4\) \cite{Cole96} originally calculated this density factor \(\Delta_c = 18\pi^2 \approx 178\) from theory. And they verified in simulations that it approximately divided the virialized and still-infalling regions of simulated Dark Matter halos in a
$\Omega_m = 1$ universe.\footnote{Navarro et al. (1996) spoke of this factor as the overdensity above $\rho_m$ rather than above $\rho_{crit}$. But as $\Omega_m = 1$ in their simulations, the two densities were equal and thus interchangeable.} A nice round number of $\Delta_c = 200$ was adopted by Navarro et al. (1996) to define the virial radii and thus the central concentrations of their simulated halos. This convention is still often used today to allow for easy comparison with earlier work. I recommend adopting $\Delta_c = 100$ as the new nice round number which approximates the result for the concordance cosmology.

But also note that so far we have assumed we are observing the collapsed halos at the present day. This practice is generally followed in analysis of simulations. That is, results are normally reported for the halos evolved to the present day. But in observations, we should account for the lookback time. In the expressions above, we replace the present day values of $\Omega_{m,0}$ and $\Omega_{\Lambda,0}$ (here “0” subscripts have been added for clarity) with:

$$\Omega_m(z) = \frac{1}{1 + (\Omega_{\Lambda,0}/\Omega_{m,0})(1 + z)^{-3}}$$  \hspace{1cm} (2.7)

and $\Omega_{\Lambda}(z) = 1 - \Omega_m(z)$. For A1689, for example, at $z = 0.1862$, this yields $\Delta_c = 115.6$ using equation 2.4 and $\Omega_m = 0.3$. This is a bit higher than the present day value, but adopting $\Delta_c = 100$ would still be a good approximation (certainly much better than $\Delta_c = 200$). Although note that at high redshift, $\Delta_c$ does asymptote to 178.

With these definitions in hand, we return to the relationship of mass $M_{vir}$ and concentration $c_{vir}$ observed in the NFW and subsequent simulations. A prescription for calculating the NFW fit parameters $\rho_s$ and $r_s$ as a function of halo mass $M_{vir}$ was given in the Appendix of Navarro et al. (1997). A recent recalibration of this relationship (Gentile et al. 2007) based on improved simulations and the current cosmology gives

$$r_s \simeq 8.8 \text{ kpc} \left( \frac{M_{vir}}{10^{11} M_\odot} \right)^{0.46},$$  \hspace{1cm} (2.8)

$$c_{vir} = \frac{r_{vir}}{r_s} \simeq 13.6 \left( \frac{M_{vir}}{10^{11} M_\odot} \right)^{-0.13}.$$  \hspace{1cm} (2.9)

where the NFW parameter $\rho_s$ can be obtained via

$$\frac{\rho_s}{\rho_{crit}} = \delta_c = \frac{\Delta_c}{3} \ln(1 + c_{vir}) - c_{vir}/(1 + c_{vir}).$$  \hspace{1cm} (2.10)

Simulated Dark Matter halos are like snowflakes in that no two are exactly alike. (Of course the same can be said for real galaxies!) Each has a slightly different NFW fit, varying about the above prescription at a rate which can be measured in a given simulation. Bullock et al. (2001) and Wechsler et al. (2002) both\footnote{Cole & Lacey (1996) spoke of this factor as the overdensity above $\rho_m$ rather than above $\rho_{crit}$. But as $\Omega_m = 1$ in their simulations, the two densities were equal and thus interchangeable.} report the following variance about a slightly different $M-c$ relation:

\footnote{Wechsler et al. (2002, footnote 10) claim that the scatter of $\Delta \log c_{vir} \sim 0.18$ reported by Bullock et al. (2001) was a bit too high and should actually be 0.14, thus bringing it in line with their own measured scatter.}
\[ c_{\text{vir}} \simeq \frac{9}{1 + z} \left( \frac{M_{\text{vir}}}{1.3 \times 10^{13} h^{-1} M_\odot} \right)^{-0.13} \], \quad \Delta \log c_{\text{vir}} \sim 0.14, \quad \text{(2.11)}

while Hennawi et al. (2007) find
\[ c_{\text{vir}} \simeq \frac{12.3}{1 + z} \left( \frac{M_{\text{vir}}}{1.3 \times 10^{13} h^{-1} M_\odot} \right)^{-0.13} \], \quad \Delta \log c_{\text{vir}} \sim 0.098. \quad \text{(2.12)}

Despite the fact that over ten years have passed since the NFW profile was observed in simulations, it has still not conclusively been observed in real galaxies or clusters to the exclusion of other profiles. Part of the problem is that a large range of radii must be probed to observe both the inner and outer profile behavior, as well as the predicted turnover. A profile constrained at a limited range of radii may be fit be many different forms. And when NFW profiles are fit to observations, their fit parameters are often discrepant from those observed in simulations (as described above). We will discuss this in more detail in §2.6. But first we discuss models beyond the NFW profile.

Other studies of simulated Dark Matter halos have found that the NFW fit leaves room for improvement. New functional forms have been proposed with most of the attention concerning the exact value for the inner slope (where NFW found \( \rho(r) \propto r^{-1} \)). This value is important, as steeper inner profiles would yield a stronger central density of Dark Matter which may produce more gamma ray flux detectable in existing or upcoming observations. However, a steeper inner slope may also exacerbates the discrepancy between simulated and observed profiles, the latter of which generally have “cored” profiles with inner regions of constant density. A few studies find central slopes shallower than that of NFW (Taylor & Navarro 2001; Ricotti 2003; Hansen & Stadel 2006), while many others find a steeper slope, as we will now discuss.

2.2 A steeper central slope than NFW?

In higher resolution simulations than those originally performed by NFW, Moore et al. (1999) found a slightly different “universal” profile, after fitting simulated Dark Matter halos to the following more general form (Hernquist 1990, his Eq. 43; see also Zhao 1996):
\[ \rho(r) = \frac{2^{(\beta-\gamma)/\alpha} \rho_s}{(r/r_s)^{\gamma}[1 + (r/r_s)^\alpha]^{(\beta-\gamma)/\alpha}}. \quad \text{(2.13)} \]

This profile has an inner “cusp” (a density which diverges to infinity) with a slope approaching \( \rho(r) \propto r^{-\gamma} \) for \( r \to 0 \), and an outer profile which is governed by \( \rho(r) \propto r^{-\beta} \) for \( r \gg r_s \). The transition between the two slopes occurs around \( r = r_s \), and the rate of this transition is governed by the parameter \( \alpha \).
Where NFW found \( (\alpha, \beta, \gamma) = (1, 3, 1) \), Moore et al. (1999) instead found best fits of \((1.5, 3, 1.5)\). While they agree on the outer slope of \( \rho(r) \propto r^{-3} \), the inner slope \( \rho(r) \propto r^{-\gamma} \) of Moore et al. (1999) is steeper than that found by NFW. This difference proves important in comparisons with observed Dark Matter profiles via rotation curves in galaxies and X-ray or gravitational lensing analyses in galaxy clusters. A steeper central slope also implies a stronger central density, increasing the possible (and hopefully observable) gamma-ray flux from Dark Matter self-annihilation in the galactic core ([1.7.1]).

Subsequent simulations have found central inner slopes in between those found by NFW \((\gamma = 1)\) and Moore \((\gamma = 1.5)\). In making such determinations, the fully generalized form in Equation 2.13 is a bit too general with large degeneracies between the free parameters (Klypin et al. 2001). Thus, in their efforts to determine the central slope \( \gamma \), authors generally use one of two constrained versions of Equation 2.13 either a “generalized NFW” profile with \((\alpha, \beta, \gamma) = (1, 3, \gamma)\):

\[
\rho(r) = \frac{\rho_s}{(r/r_s)^\gamma [1 + (r/r_s)]^{(3-\gamma)}},
\]  

(2.14)

or what we might call a “generalized Moore” profile\(^3\) with \((\alpha, \beta, \gamma) = (3 - \gamma, 3, \gamma)\):

\[
\rho(r) = \frac{\rho_s}{(r/r_s)^\gamma [1 + (r/r_s)^{3-\gamma}]}.
\]  

(2.15)

In the highest resolution halo simulation to date – a cluster-sized halo with 127 million particles – Diemand et al. (2005b) find a central slope of \( \gamma = 1.2 \), which remains constant below \( r = 0.01 r_{vir} \), down to their smallest resolved radius \( r = 0.002 r_{vir} \) (see Fig. 2.2). They claim that theirs is the first simulation to have enough resolution to measure a constant inner slope, thus refuting the claims to follow in the next section. However their result has yet to be confirmed by other simulations performed by other authors.

2.3 A new universal profile for both light and Dark Matter?

Recently Navarro et al. (2004) have proposed a better mouse trap. They admit that the central slope of Dark Matter profiles becomes shallower with decreasing radius more slowly than predicted by NFW. But they claim that is incorrect to interpret this discrepancy as requiring a steeper central slope. Instead they propose that the profile slope varies continuously over the entire profile without tending toward any asymptotic value at the center nor at large radius. This behavior has been observed in many simulated halos even while authors continued to look for a single slope which failed to converge to any single value at the inner resolution limits (e.g., Klypin et al. 2001).

\(^3\)This latter form is also often referred to as a “generalized NFW” profile, although strictly speaking it can only exactly reproduce the Moore profile and not that of NFW.
Figure 2.2 Radial profile power-law slopes of Dark Matter halos simulated by Diemand et al. (2005b, reprinted from their Fig. 4). Arrows indicate each simulation’s “convergence radius”, within which the slope flattens due to limited resolution. The highest resolution cluster “DM25” contains 127-million particles. Note that for this simulation, the central slope appears to converge to a single value $\gamma \sim 1.2$ from $r \sim 0.01r_{\text{vir}}$ down almost all the way to the convergence radius. Compare this behavior to that observed in other simulations in Fig. 2.3.
Jing & Suto 2002; Fukushige et al. 2004; however see Diemand et al. 2005b). In fact this idea of a continuously varying power-law slope is not new, but was characteristic of fitting forms for Dark Matter halos popular in the 1980’s before these were replaced by double a power-law form (Dubinski & Carlberg 1991) which evolved to the NFW profile proposed by Navarro et al. (1996) (see the nice brief historical review given in the introduction to Merritt et al. 2006). Navarro et al. (2004) revived this old idea with their proposed functional form

\[
\rho(r) = \rho_{-2} \exp \left\{ \left( -\frac{2}{\alpha_N} \right) \left[ \left( \frac{r}{r_{-2}} \right)^{\alpha_N} - 1 \right] \right\}. \tag{2.16}
\]

The power-law slope of this density profile is -2 at \(r = r_{-2}\) and varies about this value according to another power law

\[
\frac{d \ln \rho}{d \ln r} = -2 \left( \frac{r}{r_{-2}} \right)^{\alpha_N}. \tag{2.17}
\]

The rate of change in the slope is provided by \(\alpha_N\).

The power-law slopes of the CDM halos simulated by Navarro et al. (2004) are plotted in Fig. 2.3. The slopes decrease steadily inward, a trend better reproduced by the profile given in Eq. 2.16 (setting \(\alpha_N = 0.17\)) than by either of the NFW or Moore profiles which converge to constant inner slopes. However, systematic offsets are still present at large radii for the largest simulated halos (clusters). Navarro et al. (2004) rectified this by allowing \(\alpha_N\) to be a free parameter in the fitting of each individual halo. The result is generally higher best fit values of \(\alpha_N\) for clusters than for galaxies, and an overall scatter for all halo sizes of \(\alpha_N = 0.172 \pm 0.032\). However, Merritt et al. (2006) lament that allowing \(\alpha_N\) to vary for each fit results in a loss of profile universality. The universality may yet be restored if \(\alpha_N\) can be found to obey a scaling relation with mass (or perhaps some other parameter(s)). But \(\alpha_N\) did not clearly obey any such relation in Navarro et al. (2004).

(However, see §2.4 and Equation 2.20)

Navarro et al. (2004) did obtain a scaling relation for the best fit parameters \(\rho_{-2}\) and \(r_{-2}\). These relations are not written down as easily as those for the original NFW profile given in Equations 2.8 and 2.9 but a script is available for calculating them upon request from the authors.

The form given in Eq. 2.16 was soon recognized (Merritt et al. 2005) to be the same as the Sersic (1968) profile:

\[
\Sigma(R) = \Sigma_e \exp \left\{ -b \left[ \left( \frac{R}{R_e} \right)^{1/n} - 1 \right] \right\}, \tag{2.18}
\]

except that the latter is a function of surface (two-dimensional projected) rather than three-dimensional density. (Note that \(b\) is not an extra free parameter, but rather a function of \(n\) defined such that half the mass is contained within \(R_e\).) The Sersic profile is commonly used to fit the light
Figure 2.3 Radial profile power-law slopes of Dark Matter halos simulated by Navarro et al. (2004, reprinted from their Fig. 3). Dwarf (red), galaxy (green), and cluster-sized halos are plotted along with the power laws of the NFW (solid) and Moore (dotted) profiles. Note that neither provides a good fit to the data at all radii. Also plotted (dot-dashed) is the power law of the form given in Eq. 2.16. Note that it better reproduces the power law’s tendency to decrease steadily with decreasing radius without converging to any particular value. The clusters however are not fit well at large radius. Navarro et al. (2004) were able to rectify this by increasing the $\alpha$ parameter for those fits, although this results in a loss of profile universality (Merritt et al., 2006). These simulated halos contain about 10 million particles each, and are thus of about an order of magnitude lower resolution than those shown in Fig. 2.2.
distributions of elliptical galaxies and the central bulges of spiral galaxies. This raises the tantalizing possibility that the collapse of Dark Matter particles and that of stars produce similar structures and thus may be understood as very similar processes.

To reinforce this idea, Merritt et al. (2005, 2006) addressed whether the Sérsic profile would fit the surface densities of simulated Dark Matter halos as well as it had fit spatial densities in the Navarro et al. (2004) study. In Merritt et al. (2005), these new fits were found to be good. But to perform a proper comparison, the Sérsic surface density profile was deprojected to three-dimensions. The deprojected Sérsic profile was found to fit the spatial densities of simulated Dark Matter halos as well as the straight Sérsic profile had. A follow-up study (Merritt et al. 2006) of additional simulations found slight differences in the performances of the various profiles. The deprojected Sérsic profile was found to fit clusters best, while the best fit to galaxies was found to be a profile proposed by Dehnen & McLaughlin (2005) (their Eq. 46b), namely the Hernquist (1990) profile (Eq. 2.13) with \([a, \beta, \gamma] = [(3 - \gamma)/5, (18 - \gamma)/5, \gamma]\). Meanwhile, the straight Sérsic profile (fit to the spatial densities) provided the best fits “across the board” to all cluster and galaxy halos.

We briefly note the following implication for gravitational lensing studies. As a deprojected Sérsic profile provided the best fit to the spatial densities of simulated clusters, it follows that the straight Sérsic profile should provide the best fit to surface densities of clusters. Thus when we perform gravitational lensing analyses of galaxy clusters and obtain surface density massmaps, we should expect to find good fits to a Sérsic profile. These fits can then be compared to those derived from simulations. Of course if we are modeling the massmaps via a parametric method, we may choose to model the mass as a Sérsic profile. The lensing equations of this mass profile have been previously derived (Cardone 2004), along with other useful properties presented in (Graham & Driver 2005; Terzić & Graham 2005).

The Sérsic profile blurs the distinction (or perhaps resolves the discrepancy!) between “cusp” and “core”, as this profile does converge to a constant value, but only at very small radius (Graham et al. 2006). Thus while not diverging to infinite density like most “cuspy” profiles, this new profile doesn’t have the extended inner region of constant density expected of a “cored” profile either. Clearer definitions are apparently in order, as Navarro et al. (2004) claim their new profile is “cuspy” while Diemand et al. (2005b) claim their new profile is “cored”.

2.4 Origin of Dark Matter Halo Profiles

While Dark Matter simulations give rise to fairly consistent halo profiles (despite the discrepancies in the details mentioned above), we have little in the way of a physical explanation of the origins for these profiles. Such a physical derivation could in principle resolve the aforementioned discrepancies by providing us with a more accurate prediction of halo shapes. To this end, two main mechanisms for halo formation have been proposed, one driven by significant mergers and the other by more smooth accretion, with perhaps both mechanisms occurring in the order given
Evolution of these processes may be followed semi-analytically, yielding halo profile predictions for a given cosmology. One such study was performed by Salvador-Solé et al. (2007). Among their important findings is that the NFW profile is a poor fit for large halos of mass $M > 10^{14} M_\odot$, resulting in inflated values for the best fit concentration parameter $c$ (Fig. 2.4). This result was previously reported in earlier work by some of these same authors (Manrique et al. 2003), and hints of these deviations were witnessed in subsequent Dark Matter simulations (Zhao et al. 2003; Tasitsiomi et al. 2004). Thus the large NFW concentration parameters reported for massive clusters like Abell 1689 should perhaps not come as a surprise. We comment more on this in §2.6.

Meanwhile, one of the improved Dark Matter profile fitting forms proposed by Merritt et al. (2005), namely the three-dimensional analog of the Sésic (1968) law:

$$\rho(r) = \rho_0 \exp \left[ - \left( \frac{r}{r_n} \right)^{1/n} \right], \quad (2.19)$$

was tested and shown to provide a superior fit to all of the Salvador-Solé et al. (2007) Dark Matter halos over at least ten decades in halo mass from $10^6$ to $10^{16} M_\odot$. They derive a useful scaling relation for the best Sésic fit parameters $n$ and $r_n$ as a function of halo mass (analogous to those given for NFW in Equations 2.8 and 2.9):

$$n \simeq 4.32 + (7.5 \times 10^{-7}) \left[ \ln \left( \frac{M}{M_\odot} \right) \right]^{4.6}, \quad (2.20)$$

$$c_n \equiv \frac{r_{\text{vir}}}{r_n} \simeq \exp \left\{ 13.3 + (7.5 \times 10^{-8}) \left[ \ln \left( \frac{M}{M_\odot} \right) \right]^{5.6} \right\}. \quad (2.21)$$

Future fits of the Sésic form to observed Dark Matter halo profiles can be compared to those fits predicted here as well as to those predicted by Navarro et al. (2004).

### 2.5 Baryons Matter

All of the simulations and semi-analytic derivations described so far in this chapter have neglected baryons, considering Dark Matter particles only. This is representative of the literature. Inclusion of baryons (i.e., gas) is non-trivial, requiring accurate modeling of complicated processes such as star formation, stellar feedback processes (i.e., supernovae), and radiative cooling and heating. Besides, we may not expect baryons to matter so much as they make up only about 10% of the matter in galaxies and 1% or less in galaxy clusters. However, as their collisional nature causes them to “sink” to the halo center, baryons attract Dark Matter to sink with them, which in turn
Figure 2.4 Radial profiles of Dark Matter halo densities (not power-law slopes this time), from semi-analytical formation models [Salvador-Solé et al. (2007), reprinted from their Fig. 2]. The halos range in mass from $10^{11} M_\odot$ to $10^{16} M_\odot$ as shown. Dashed lines give the NFW fit for each profile with the concentration parameters indicated, and the percent deviation from this fit plotted below each panel. Note that the NFW fits are good except for the most massive halos with $M > 10^{14} M_\odot$. The poorness of these fits leads to an inflated best-fit concentration parameter for these massive halos. Meanwhile, a Sérsic profile provides good fits to all of the semi-analytical halos over an even larger range of mass from $10^6 M_\odot$ to $10^{16} M_\odot$ (not shown).
attracts more Dark Matter, etc. Thus the addition of a small fraction of baryons has an effect which is multiplied, and may result in a steeper slope for both the baryon and Dark Matter components.

One recent study on the effects of inclusion of baryons in Dark Matter simulations was performed by Gustafsson et al. (2006). They compare simulated Dark Matter halos with those same halos resimulated with the inclusion of baryons (with reasonable assumptions made about the processes of star formation, etc.). The inclusion of baryons results in a steeper central Dark Matter halo profile as shown in Fig. 2.5. (The profile measured in all cases is of the Dark Matter particles only, to the exclusion of baryons when present.) In the DM-only simulations, the power-law slopes decrease continuously inward, failing to converge to any single slope. This is the same behavior described by Navarro et al. (2004). (But as discussed in §2.2, Diemand et al. (2005b) argue that such convergence is only realized when simulating Dark Matter halos with sufficiently high resolution: over a hundred million particles. The Gustafsson et al. (2006) halos contain a relatively modest 10,000 Dark Matter particles each.) Meanwhile, in the DM+baryon simulations, the Dark Matter profiles show hints of converging to near isothermal profiles ($\rho \propto r^{-2}$) in the center.

But if the addition of baryons leads to a steeper central Dark Matter slope, doesn’t this exacerbate the cusp-core controversy? Gustafsson et al. (2006) say no, because while the Dark Matter forms a steeper cusp, the baryons do form a core, and in the central regions of simulated galaxies, it is the baryons rather than the Dark Matter which prove to dictate the shape of the rotation curve. Thus observed rotation curves which indicate a central core may disagree with Dark-Matter-only simulations, but agree just fine once baryons are included in the simulations. Dutton et al. (2007) agree that baryons can resolve this and many other discrepancies between simulations and observations. But rather than the adiabatic contraction described above, they claim that the opposite effect is required by observations, namely an adiabatic expansion resulting in a shallower central Dark Matter slope. Adiabatic expansion can result as baryon clumps transfer energy to the Dark Matter halo via dynamical friction (Tonini et al. 2006; Dutton et al. 2007). Dutton et al. (2007) claim that only models with such adiabatic expansion give results consistent with all observables, including the Tully-Fisher relation with appropriate normalization, and observed disk sizes and luminosity functions, while also reproducing the flat rotation curves generally observed. Many previously claimed discrepancies between CDM disk formation models and observations suddenly evaporate once adiabatic expansion is considered.

Baryons also have other effects on halo shapes aside from those discussed here regarding the profile and central slope. For instance, baryon gas cooling makes the overall halo less elliptical and more spherical (Kazantzidis et al. 2004).

2.6 Simulated vs. Observed Dark Matter Halo Profiles

Given all of the effort put into determining the exact profile of simulated Dark Matter halos, it is unfortunate that observations have not been able to clearly confirm or distinguish between
Figure 2.5 Radial profile power-law slopes of Dark Matter halos in simulations with and without the inclusion of baryons (Gustafsson et al. 2006, reprinted from their Fig. 3). Dashed blue lines are from the Dark Matter only simulations while red solid lines are from resimulations of the same halos with the inclusion of baryons. The plots on the left are of Milky-Way-sized galaxies \((8 - 9 \times 10^{11} M_\odot)\) while those on the right are of galaxies about nine times smaller. Note that the slopes of the Dark Matter only simulations decrease continuously inward through the smallest resolved radius as described by Navarro et al. (2004) and shown in Fig 2.3. The slopes of the DM+baryon simulations meanwhile show hints of converging with a central slope near \(\gamma \sim 2\), or that of an isothermal sphere \(\rho \propto r^{-2}\). Each halo contains \(~10,000\) Dark Matter particles. Note that the opposite effect (adiabatic expansion resulting in a shallower DM central slope) has also been reported (Tonini et al. 2006; Dutton et al. 2007).
the proposed profiles. The fact is that simulations are able to map Dark Matter halos in much
greater precision and detail than is currently possible observationally. In §3.3 we present a new
analysis technique which may help pick up a bit of the slack on the observational end. This will be
necessary to resolve some important discrepancies now apparent between observed and simulated
halos, discrepancies which may constitute the greatest outstanding challenge to $\Lambda$CDM theory today.

The “cusp” vs. “core” controversy was discussed above, where it was suggested that the
problem may be alleviated by an improved fitting form such as that proposed by Navarro et al. (2004)
and especially by the inclusion of baryons in simulations. But more work can also be done on the
observational end to accurately map the inner Dark Matter profiles of galaxy clusters. While Doppler
shift mapping of gas rotating in nearby galaxies can map their central Dark Matter distributions in
relatively high detail (albeit not entirely assumption free), the Dark Matter distributions of galaxy
clusters remain less well probed. Less cuspy, or even cored, profiles appear to be required to produce
the gravitationally-lensed central images observed in clusters. But cuspy profiles such as NFW are
often fit and fit well to observed cluster halos.

However, the fit parameters derived in observational studies sometimes differ from those
predicted in simulations. For example, Abell 1689 was found to have $c_{200} = 11$ (with $\Delta_c = 200$) in a
combined strong and weak lensing analysis by Broadhurst et al. (2005b), while a cluster of its great
heft ($M \simeq 2 \times 10^{15} M_\odot$) should have $c_{200} = 5.5$, as predicted by Bullock et al. (2001, our Eq. 2.11).
Note that subsequent study appears to have brought the NFW concentration of A1689 more in line,
down to $c_{200} = 7.6 \pm 1.6$ or even $6.0 \pm 0.6$ (weak and strong lensing fits, respectively, from Limousin
et al. 2007; also see previous determinations of $c_{200}$ for A1689 summarized and referenced within).
And we recall from §2.4 that according to semi-analytical modeling (Salvador-Solé et al. 2007), the
NFW fit may “break down” for halos of such high mass, resulting in inflated best fit concentration
parameters.

A comprehensive study of NFW concentration parameters derived for clusters to date was
performed recently by Comerford & Natarajan (2007). They compiled 182 unique measurements of $c$
for 100 galaxy clusters, including 10 new determinations they present for as many clusters. The
mass-concentration relation they detect is shown in Fig. 2.6. The correlation is weak at best, with an
observed power-law slope of $\alpha = -0.14 \pm 0.12$ that is marginally consistent with zero, i.e., no slope
and no correlation. But the large scatter in the data may be expected, as the scatter they measure
($\Delta \log c = 0.15$) is similar to those reported ($0.14$ and $0.098$) by Bullock et al. (2001) and Hennawi
et al. (2007), respectively. In fact, it is impressive that the observed scatter is not greater than
predicted, considering the wide range of authors and techniques used to obtain these measurements.
Or turning this around, we may expect the true scatter of $c$ in clusters is actually less than that
reported here (perhaps more in line with the lower value predicted by Hennawi et al. 2007). The
observed scatter might be reduced by ensuring consistent measurement methods across studies.

Also note that a relatively small range of masses is studied here compared to the range
probed by simulations. Only studies of galaxy clusters are considered here, and then only the most
massive clusters are well represented as they induce stronger lensing which is more easily detected and thus selected for study. Study of a wider range of mass halos might reveal a clearer mass-concentration correlation. There is also some concern with the specific mass range probed here. Most of the clusters weigh in at $M > 10^{11} M_\odot$. At this mass scale, NFW fits may yield inflated concentration values (Salvador-Solé et al. 2007), as discussed in §2.4. Future analysis of cluster and galaxy halos may consider the use of other fitting forms such as the Sérsic profile which shows signs of being more accurate and less biased at all mass scales (§2.3, 2.4). Of course many data points will be required to build up a sample capable of revealing a strong mass-concentration relation whatever the profile formulation.
2.7 Ellipticity & Substructure

The lengthy discussion of Dark Matter halo profiles in this chapter may have lulled the reader into thinking that halos have simple spherical symmetry. But in fact halos realized in simulations are generally elliptical with rich substructure. These characteristics provide additional predictions to be tested against observations.

Simulated Dark Matter halos are generally found to have axis ratios of approximately 0.64 : 0.74 : 1 ([Kasun & Evrard 2005], [Bailin & Steinmetz 2005], [Allgood et al. 2006], [Knebe & Wießner 2006], [Hayashi et al. 2007]) with some scatter found both within and between the different studies. Given this ellipticity, the assumption of spherical symmetry can bias inferred density measurements by 10-20% at large radii ([Knebe & Wießner 2006]). It may even bias measurements of galactic rotation curves toward the appearance of a central core where none exists, contributing to the cusp-core controversy ([Hayashi et al. 2007]). And line-of-sight elongation, while that most difficult to measure observationally, has been invoked as a possible explanation for the high concentrations measured in Abell 1689 ([Oguri et al. 2005]). Such an orientation makes for a stronger lens, thus it may not be surprising for the strongest lens in our sky to be aligned just so. Thus in our gravitational lensing studies, we are well advised to include ellipticity in the radial profiles we use to fit mass density.

Dark Matter halo substructure, meanwhile, can be best studied in “Via Lactea” (Fig. 2.7), the highest resolution galaxy simulation to date with 234 million particles ([Diemand et al. 2007]). The amount of substructure is an order of magnitude greater than that observed in our own Milky Way ([Strigari et al. 2007], and references therein), yet perhaps too small to explain the “flux anomalies” observed in many cases of strong galaxy-quasar lensing ([Metcalf et al. 2004], [Amara et al. 2006], [Macciò & Miranda 2006], [Diemand et al. 2007]).

Reconciliation of the first substructure discrepancy, the so-called “missing satellites” problem, may be the “stochastic bias” solution ([Diemand et al. 2005a]). In a given Dark Matter halo, the densest regions are assumed to attract the earliest condensations of gas and thus the earliest star formation. The radiation from this star formation may heat the surrounding gas sufficiently to significantly suppress subsequent star formation. Thus the smaller pockets of Dark Matter, which should “turn on” later, may simply fail to do so, remaining relatively dark. Proving the existence of such dark subhalos would be an exciting observational coup.

As for flux anomalies observed in lensing, there are still large uncertainties on both the amount of substructure required and the amount realized in simulations. Our new lensing analysis method §3.3 may be able to find macro-structure massmap solutions capable of reproducing many observed flux anomalies without invoking substructure. Even if we can obtain reasonable macro-structure solutions for all observed flux anomalies, substructure would still likely be the correct explanation in at least some cases. That is we cannot expect to rule out substructure as the cause of flux anomalies. But if we can provide alternative explanations (massmap solutions), then we can claim that substructure is required less often and in smaller amounts than previously believed.
Figure 2.7 “Via Lactea” (Latin for “Milky Way”), the highest resolution Dark Matter galaxy simulation to date ([Diemand et al. 2007](#) reprinted from their Fig. 2). Projected density squared within an area $800 \times 600$ kpc and $600$ kpc deep is plotted on a logarithmic scale spanning 20 decades.
Chapter 3

Gravitational Lensing

Strong gravitational lensing is a spectacular demonstration of Einstein’s Theory of Relativity [Einstein 1916]. Einstein explained Newton’s Law of Gravity with the radical idea that mass bends space itself. Light from distant galaxies generally travels straight to us like a golf ball hit firmly along a putting green. But an intervening galaxy or cluster of galaxies can put a “divot” in the light’s path, deflecting its course and thus distorting the galaxy images we see. In the case of a “strong” gravitational lens, light takes multiple paths around the lens (or divot) which reconverge in our telescopes, yielding multiple images of each individual galaxy. A most striking example of this is shown in Fig. 3.1. If this seems too extraordinary to comprehend, you will find that a similar case of lensing is as close as your next glass of wine. Glass (and liquid) refract light in much the same way as gravity, creating distorted and even multiple images of more distant objects (Fig. 3.2).

We are not able to walk around the putting green of space to observe how it is bent as we would to line up a putt. But Nature has provided us with something even better: a series of “practice putts” which allow us to observe how light has followed the curvatures of space. Each gravitationally lensed galaxy shows us the result of one such practice putt, with light from the galaxy following the curvature of space until its arrival at our telescope. Images of a galaxy cluster such as that shown in Fig. 3.1 can reveal anywhere from a few to a hundred or more such practice putts, allowing us to map out the curvature of space, and thus the mass of the galaxy cluster.

This is the most direct and assumption-free method we have for observing the mysterious Dark Matter that makes up over 80% of the material in our universe. In §3.2 we will describe in general terms how this analysis is usually performed, and then in §3.3 how we can do better in terms of both precision and efficiency. In §3.4 we finally dive into a few lensing equations and show how mass maps are reconstructed using our new method. But first, in §3.1 we take a moment to briefly review the history of gravitational lensing.

Figure 3.1 Galaxy cluster Abell 2218, a dramatic example of gravitational lensing, shows multiple images of background galaxies stretched into long arcs in this WFPC2 color composite image (NASA/STScI).
Figure 3.2 More down-to-Earth example of lensing using a wine glass to mimic several multiple image configurations observed in gravitational lensing (credit: Phil Marshall).
3.1 History of Gravitational Lensing

The bending of light due to mass was first contemplated by Isaac [Newton et al. (1782)]. Having determined the equations governing the gravitational attraction of two massive bodies, it was clear that the acceleration experienced by an object is independent of its mass, depending only on the mass of the other object. For example, all bodies fall toward the ground on Earth at the same rate, regardless of their mass. (Or at least they would, were wind resistance were not a factor). Newton thus reasoned that particles of light, even if they are massless, should feel a similar attraction to massive bodies. [Michell (1784)] calculated that light from a distant object passing by the Sun (just skirting its edge) would be deflected by

\[ \alpha = \frac{2GM_\odot}{c^2R_\odot} = 0\degree 88, \]  

(3.1)
given the mass \( M_\odot \) and radius \( R_\odot \) of the sun, Newton’s gravitational constant \( G \), and the speed of light \( c \).

[Einstein (1916)] explained gravity as an effect of the distortion of space, with light (and massive bodies) traveling in paths dictated by this distorted space. Einstein originally calculated that this geometrical distortion would yield the same deflection of light as do Newton’s equations. But in 1916, Einstein revised his calculations realizing that the deflection should be twice that predicted by Mitchell. In General Relativity not only is space bent, but time is distorted as well, actually slowing down near massive objects. This “time dilation” creates an additional deflection of light for the same reason light is refracted by glass or water. Light travels slower in these mediums than in air, and thus attempts to travel around them in search of the quickest path (according to Fermat’s principle\(^2\)). Similarly, light travels slower close to a massive body, and thus goes out of its way to find a quicker path around. The deflection thus induced by time dilation is equal to that due to the geometrical bending of space. Thus Einstein predicted that the deflection due to the Sun would be

\[ \alpha = \frac{4GM_\odot}{c^2R_\odot} = 1\degree 76, \]  

(3.2)
or exactly twice that calculated by Mitchell.

But such deflection of light due to mass had never been observed. Thus Sir Arthur [Ed-dington] set out on a now-famous expedition to detect and measure this deflection during a Solar eclipse in [1919]. The sun was favorably located in front of the Hyades star cluster at the time of the eclipse, thus allowing for measurements of the positions and deflections of these background stars.

\(^2\)To this day, a truly intuitive explanation of Fermat’s principle proves elusive. That is, why does light seek out the quickest paths, and how does it find them?
The results were proclaimed to be just as Einstein predicted (Eddington 1920), thus supporting his strange theory of warped space and time, and catapulting Einstein to celebrity status.

Some subsequent reanalyses of the data obtained on that day indicate that there were in fact large uncertainties which should have tempered any declaration of triumph for Einstein’s Theory of Relativity (e.g., Maddox 1995). However, this experiment has since been repeated, with the new higher precision data robustly confirming Einstein’s predictions. Signals received from spacecraft and pulsars passing behind the Sun have also been observed to slow and bend by the exact amount predicted by Einstein (e.g., Anderson et al. 1975). And other predictions of General Relativity have been confirmed in numerous experiments (e.g., Schäfer 2003 and references therein). Even orbiting GPS satellites must account for the effects of General Relativity.

Lodge (1919) was the first to use the term “gravitational lensing” to describe this deflection of light due to a massive body. But he noted that these “lenses” have no focal length. In fact the focal length increases with distance from the center of the lens. Chwolson (1924) described the “Einstein ring” image produced by an object located directly behind a point mass gravitational lens. A less than perfect alignment would yield two multiple images of the object instead of a full Einstein ring. Einstein was asked by his friend Rudi Mandl, a Czech Engineer, to determine whether we might ever detect such gravitational lensing due to other stars outside our solar system. Einstein (1936) correctly reasoned that such deflections (and thus image separations) would be on the order of milliarcseconds, and thus much too small to detect. But Zwicky (1937) soon realized that an entire galaxy, even given its great distance, would make a strong enough lens to produce detectable deflections, on the order of 10″. The true value turns out to be lower, but only by an order of magnitude, and still very detectable with a decent telescope. Given the fuzzy knowledge of galaxies back in 1937, it is impressive that Zwicky’s estimate was so accurate.

But the first discovery of a gravitational lens which produces multiple images would have to wait until forty years later. In the meantime there was at least one false detection proclaimed by Zwicky, in response to which, a colleague Guido Munch vowed to eat his hat should the object in fact prove to be a lens (Schneider 2006 footnote 1). Finally, the “Twin Quasar” Q0957+561 (Fig. 3.3) was discovered serendipitously by Walsh et al. (1979). The two quasar images were shown to have very similar spectra, suggesting they may be multiple images of the same distant galaxy. And soon after, the intervening lensing galaxy at lower redshift was discovered thus solidifying multiple image theory (Stockton 1980; Young et al. 1980). Since then, various surveys such as CASTLES (Falco et al. 2001), CLASS (Myers et al. 2003), SLACS (Bolton et al. 2006), and CFHTLS-SL2S (Cabanac et al. 2007), have combed the sky discovering numerous such strong galaxy-galaxy lenses (those for which multiple images of a background galaxy are produced by a foreground lensing galaxy).

---

3Apparently Einstein also considered this case in some unpublished notes in 1912 (Renn et al. 1997). Thus the name “Einstein ring” is appropriate.

4A “quasar” (short for “Quasi-Stellar Object”, or “QSO”) is an object that appears point-like as does a star, but is much too distant to be a star. Quasars are now believed to be distant galaxies with bright active cores that dominate the observed images.
Figure 3.3  Q0957+561, the first gravitationally-lensed multiply-imaged object discovered (Walsh et al. 1979) seen here in the H-band (left), courtesy of the CASTLES survey (Falco et al. 2001). Two bright concentrated images of the quasar are seen at the top and bottom with diffraction spikes, while the lensing galaxy is the more diffuse object just above the lower quasar image. Right: the original spectra of the two quasar images as observed by Walsh et al. (1979, reprinted from their Fig. 2). The two spectra were very similar, leading Walsh et al. to propose that they might be multiple images of the same quasar. Support for this claim came the following year with the detection of the lensing galaxy (Stockton 1980; Young et al. 1980).
Meanwhile, clusters of galaxies, despite being at greater distances still, appear to us as even stronger lenses. The strongest known lens in our sky is the galaxy cluster Abell 1689 (Fig. 3.4). It has an Einstein radius of $\sim 45''$. (A single exact value cannot be pinned down, as the lens is not perfectly circularly symmetric.) In deep multiwavelength images taken by ACS (the Advanced Camera for Surveys on-board the Hubble Space Telescope), 106 multiple images of 30 galaxies were detected [Broadhurst et al., 2005a]. This was about an order of magnitude more than the number of multiple images detected in any previous galaxy cluster observation. Such a profusion of multiple images allows for a very detailed “massmap” to be produced of the mass distribution of the galaxy cluster. In fact, as galaxy clusters are predominantly (99% or more) Dark Matter, these massmaps can offer a direct and detailed “image” of Dark Matter. But conventional massmap reconstruction methods were overwhelmed by this large number of multiple images. In §3.3 we explain why, and in §3.3 how our new method overcomes these difficulties.

3.2 Methods of Strong Gravitational Lensing Analysis

The 100+ multiple images detected in recent ACS images of Abell 1689 have been analyzed by at least 6 different groups of researchers, with each using a different analysis method. Here we describe these methods, as they provide us with a good cross-section of the methods of strong gravitational lensing analysis commonly being used today. We will not discuss “weak” gravitational lensing, or that which distorts images but does not yield multiple images.

Strong lensing analysis methods generally propose a series of massmap solutions, checking how well each reproduces the observed multiple images. The procedure (hopefully) converges toward a solution which performs best (if not perfectly). Methods are often classified as “parametric” or “non-parametric”, depending on whether a clear parameterization is used to propose massmaps. A “grid-based” method is often referred to as “non-parametric” even though strictly speaking it does have parameters, namely the mass at every pixel on a grid. Here we would like to make a clearer distinction between two types of methods: those which model mass halos as physical analytical forms and those which do not. Each of these method types, which we will refer to as “model-based” and “model-free”, has advantages and disadvantages as we will discuss.

Halkola et al. (2006), for example, perform a “fully parametric” (or “fully model-based”) massmap reconstruction of A1689. They model the overall cluster halo as two clumps with either elliptical NFW or NSIE (Non-Singular Isothermal Ellipsoid) profiles. To this overall halo they then add a galaxy component, with each cluster galaxy modeled as an elliptical truncated isothermal sphere based on its observed light distribution. Scaling relations for the galaxy velocity dispersions and truncation radii are derived from those found observationally in large surveys and from theoretical considerations. The following parameters are then allowed to vary in search of the mass model which best produces the observed multiple images: the positions and fit parameters of the two cluster halo clumps, and the overall scaling parameter for the truncation radii of the galaxy halos.
Figure 3.4  Galaxy cluster Abell 1689, the strongest gravitational lens in our sky with an Einstein radius of $\sim 45''$. The deep multiwavelength $g'r'i'z'$ ACS images, shown here as a color composite, reveal over 100 multiple images of 30 background galaxies (Broadhurst et al. 2005a).
A similarly physically-motivated mass modeling approach is adopted by Limousin et al. (2007) who construct their own mass model of A1689.

The first published mass model of the A1689 ACS images (Broadhurst et al. 2005a) was not as rigorously physically motivated, and instead a bit more flexible. Each cluster galaxy was modeled as a simple power law mass profile. The resulting total massmap was then fit to a cubic spline, resulting in the “smooth” fit and a residual “lumpy” component. The gravitational potential of the smooth component was allowed to vary in amplitude and by the addition of a third- or fourth-order polynomial in $x$ and $y$. The lumpy component was also allowed to vary in amplitude. The two components were recombined to form each proposed massmap solution. A subsequent paper (Zekser et al. 2006) modified the Broadhurst et al. (2005a) method, modeling each cluster galaxy as a Singular Isothermal Ellipsoid (SIE), and then adding an overall halo, consisting of an NFW profile, shapelets, and a mass sheet. This formulation is a bit more physical, but with the shapelets still allowing extra degrees of freedom.

Finally, two grid-based massmap reconstructions of Abell 1689 have been performed. Diego et al. (2005) construct a massmap on an adaptive grid of 600 pixels and find that which best reproduces the observed images. Attempts to perfectly fit the data with their method lead to instabilities; thus they purposely allow for a small amount of scatter. Using the PixeLens method, Saha et al. (2006) boast perfect reproduction of the image positions, but they are only able to achieve this for 30 images at a time. As resolution of the reconstructed massmap scales directly with the number of multiple image constraints, we may expect that a massmap obtained using all 100+ image constraints to contain about three times the detail of the PixeLens solutions.

There is good reason that both model-based and model-free approaches have been used, as each has advantages and thus serves a purpose. Physical mass models are more likely to resemble the true form of the mass. In principle, they also allow for a more straightforward determination of meaningful physical parameters. For example, Halkola et al. (2007) put constraints on galaxy halo truncation radii (Fig. 3.5), showing that they are more truncated than galaxies in the field. However, to put faith in these galaxy constraints, we must assume that Halkola et al. (2007) have adopted the correct model for the overall cluster halo, as there are obviously degeneracies between the galaxy and cluster halos. In fact they present two different cluster halo models, both with elliptical symmetry (neither performing significantly better than the other). And we must further accept their claim that their poor-fitting models, which reproduce the image positions to within $\sim 2.7$ RMS, are significantly worse than their best model which achieves 2.5 RMS offsets.

On the other hand, model-free massmap reconstruction methods are, by definition, free of assumptions regarding halo shapes, and, importantly, free from assumptions that mass follows light. Some of our most important discoveries about Dark Matter have come and are expected to come from cases in which mass does not follow light. Mass peaks offset from light peaks can provide constraints on the collisional nature of Dark Matter particles, as in the Bullet Cluster (Clowe et al. 2006, discussed in §1.3). And observations of Dark substructure (not associated with light) may...
Figure 3.5 Constraints placed on galaxy halo size by Halkola et al. (2007, reprinted from their Fig. 1a). Poorness of model fit in reproducing the multiple images (measured in RMS offsets) is plotted versus a scale parameter of the halo truncation radii. The truncation radii $s$ are assumed to follow the relation $s = s_o (\sigma / \sigma_o)$, with the velocity dispersions $\sigma$ derived from the observations using the Fundamental Plane relation (Dressler et al. 1987; Djorgovski & Davis 1987; Bender et al. 1992). A different scaling relation of $s = s_o (\sigma / \sigma_o)^2$ was also implemented with similar results (not shown). In both cases, $\sigma_o$ was fixed equal to 220 km/s. The different curves are from realizations utilizing different cluster halo forms (ENFW and NSIE) and for all realizations combined. Note that the best mass models yield images offset from the true positions by 2.5 RMS, while the poorest fits see this increase only marginally to 2.7 RMS.
vindicate CDM simulations which predict much more halo substructure than is visible (§2.7).

Model-free methods are also able to explore a wider range of possible massmap solutions, including (with the advent of LensPerfect) those which perfectly reproduce all 100+ multiple image positions. Of course, exploring the full range of model-free solutions can be a computational challenge. And it is not entirely clear how to sort the physical solutions from those which are “less” physical or aphysical.

Meanwhile, too much model flexibility may be a bad thing, as a flexible model may fit incorrect data without any alarms sounding. For example, Limousin et al. (2007) claim that their model-based method and that of Halkola et al. (2006) were unable to fit some multiple image systems incorrectly identified in the initial Broadhurst et al. (2005a) work, which was a bit more model-free. And Kochanek (2004) objects to the ability of the model-based but flexible Evans & Witt (2003) method to produce a mass model that accounts for the flux anomaly observed in the strong galaxy-quasar lens Q2237+0305 even though this flux anomaly has since been shown to have been due to a microlensing event, which has now passed!

In this latter case, the mass model proposed by Evans & Witt (2003) is but one possible solution among several to the observed flux anomaly. All possible solutions should be considered, and in this case, microlensing is proven to be the true solution.

As for fitting incorrect data, this is certainly a concern, as in any study. But our new LensPerfect method, while using a very flexible massmap parameterization, often sounds a stronger alarm than previous methods if an incorrect multiple image is added. The reason is that most previous methods experienced some scatter in the predicted source/image positions. Thus a misidentified multiple image set might go unnoticed more easily (by both model-based and model-free methods). LensPerfect is less forgiving. Each multiple image puts a rigid constraint on the deflection field. Thus a misidentified multiple image is more likely to cause the deflection field to get tangled, leading to a “less physical” (if not aphysical) solution. We stay alert for such ill-fitting multiple image systems as we add each to our models. And, of course, we must take care to avoid misidentifying multiple images from the start whenever possible.

But the most unsatisfying feature of all previous methods is the fact that none are able to reproduce all the multiple images perfectly. And they generally arrive at their best (least imperfect) solutions only after many computationally expensive iterations. An imperfect fit indicates that some of the data is being wasted. Precise positions can be measured for all of the multiple images, with each encoding information about the shape of the cluster massmap. To not fit the images perfectly is to discard some of the information available in these rich observations.

### 3.3 Perfect Reproduction of All Multiple Images

The problem with most methods of gravitational lensing analysis is that they are too much like a trip to the optometrist to be fitted for new eyeglasses. Different lenses are tried on and selected
for that which best brings the multiple images of each galaxy into focus at a single point. “Lens number one, or number two?” At the optometrist, this process quickly converges on a proper set of glasses. The same is the case for simple cases of gravitational lensing, for example, a single object lensed into four multiple images. A simple lens form is sufficient to reproduce four multiple images at their proper positions. Thus a few parameters are all that are necessary and all that must be tweaked to find a proper lens. But this method breaks down in more complicated cases, especially in that of Abell 1689. Here, an elaborate lens with complex shape is required to accurately bring all 30 galaxies perfectly into focus. Thus a wide range of complicated lenses must be explored. “Lens number 64,372?” If this were a trip to the eye doctor, you could be there all day and still not find a perfect set of eyeglasses!

Rather than proposing different lenses and trying them on for fit, a more direct method is called for. It is as if we were asked “What plus 3 equals 5?” \((x + 3 = 5)\). We might attempt different solutions until we find that which fits best: \(7 + 3 = 10; 1 + 3 = 4; 2.5 + 3 = 5.5\), etc. Or we may simply subtract to quickly and directly find the correct answer: \(5 - 3 = 2\). Our new gravitational lensing analysis method does just that: basically, we have figured out how to subtract.

Such “direct inversion” of multiple images has been performed before, though it was only successfully applied to simple systems with a single lensed galaxy. \cite{Evans2003} used matrix inversion to directly obtain massmap solutions which perfectly reproduced multiple image positions and fluxes. Their massmap solutions were formulated to be isothermal spheres \((\rho \propto r^{-2})\) with multipole angular structure and optional external shear. While providing reasonable solutions for the lensed quasar systems Q2237+0305 and PG1115+080, their solution for B1422+231 was admittedly too “wiggly” to be physical. Developing this method a bit further, \cite{Congdon2005} produced a more reasonable solution for B1422+231, but were less successful with B2045+265 and B1933+503. In both articles, the failures were cited as cases in which clumpy substructure must be present and responsible for the “flux anomalies” (or departures of the fluxes from those predicted by simple mass models). But it is possible that their mass models simply were not flexible enough to perfectly fit the positions and fluxes of multiple images in these systems. We will address this issue with LensPerfect, which is model-free and capable of producing a much wider range of solutions.

Direct inversion has also been used before in weak lensing mass reconstruction methods \cite{Kaiser1995, “KSB”}. However multiple image observations are even better suited to this technique, as observed image positions have much smaller uncertainties than individual weak shear measurements, as the latter are dominated by the intrinsic galaxy shapes. Multiple image positions, meanwhile, can generally be measured to an accuracy of about one pixel in a given image. In ACS images of Abell 1689, this one pixel uncertainty is miniscule compared to the Einstein radius of over 900 pixels. By perfectly matching these tight observational constraints, our massmaps will properly utilize of all the information available in the fine ACS images.
3.4 Deflections in Detail

Now we will give a more detailed explanation about how these mass structures can be resolved, and how we do so directly and precisely with LensPerfect. We will speak in terms of the “deflection field” $\vec{\alpha}$, which, for a given position $\vec{\theta}$ in the plane of the lens, measures how far light has been deflected (Fig. 3.6). In the absence of intervening mass, this deflection field would be zero everywhere. Once mass is introduced, each bit of mass causes light to bend around it according to General Relativity and Equation 3.2. From this, we can derive the total deflection $\vec{\alpha}$ at a position $\vec{\theta}$ due to all these mass bits:

$$\vec{\alpha}(\vec{\theta}) = \frac{1}{\pi} \int d^2 \vec{\theta}' \kappa(\vec{\theta}') \frac{\vec{\theta} - \vec{\theta}'}{|\vec{\theta} - \vec{\theta}'|^2},$$  \hspace{1cm} (3.3)

with the corresponding less-intimidating inverse relation:

$$\nabla \cdot \vec{\alpha} = 2\kappa,$$  \hspace{1cm} (3.4)

where $\kappa = \Sigma/\Sigma_{\text{crit}}$ is the surface mass density in units of the “critical density”. The critical density is generally that required for multiple images of a given source to be produced. In fact, for a circularly symmetric lens which produces an Einstein ring, the average mass inside that ring is exactly this critical density $\Sigma = \Sigma_{\text{crit}}$ (or $\kappa = 1$). The critical density is given by

$$\Sigma_{\text{crit}} = \frac{c^2}{4\pi G} \frac{D_S}{D_L D_{LS}},$$  \hspace{1cm} (3.5)

involving a ratio of the angular-diameter distances from observer to source $D_S = D_A(0, z_S)$, observer to lens $D_L = D_A(0, z_L)$, and lens to source $D_{LS} = D_A(z_L, z_S)$. Angular-diameter distances are calculated as follows [Fukugita et al. 1992] filled beam case; see also [Hogg 1999]:

$$D_A(z_1, z_2) = \frac{c}{1 + z_2} \int_{z_1}^{z_2} \frac{dz'}{H(z')},$$  \hspace{1cm} (3.6)

where the Hubble parameter $H(z)$ tracks the variation of the Hubble “constant” over time (given here as a function of redshift $z$). For a flat universe ($\Omega = 1$),

$$H(z) = H_o \sqrt{\Omega_m (1 + z)^3 + \Omega_{\Lambda}},$$  \hspace{1cm} (3.7)

(Note that for $z = 0$, this reduces to the present-day value $H(0) = H_o$, as $\Omega = \Omega_m + \Omega_{\Lambda} = 1.$)
Figure 3.6 Light ray deflection by a gravitational lens of mass $\kappa$. In the absence of the lens, the galaxy would appear at its true, or “source”, position $\beta$. The intervening mass deflects its light by an amount $\alpha$ to position $\theta$ (as defined in the lens plane, also known as the image plane). $D_S$, $D_{LS}$, and $D_L$ are all measured as angular-diameter distances (see text).
Note that $\Sigma_{\text{crit}}$ is a function of the source redshift. This follows because the deflection angle $\vec{\alpha}$ is a function of source redshift. As source redshift decreases, the light bend angle remains constant requiring the image deflection to decrease by the distance ratio

$$\vec{\alpha} = \left( \frac{D_{LS}}{D_S} \right) \vec{\alpha}_\infty,$$  

(3.8)

where $\vec{\alpha}_\infty$ is the deflection for a source at infinite redshift.

Thus the problem of massmap reconstruction can be reduced to determining the deflection field with all deflections scaled to a common redshift (e.g., $\vec{\alpha}_\infty$), at which point we simply take the divergence and divide by 2 to obtain the massmap (Equation 3.4). The deflection field $\vec{\alpha}(\vec{\theta}) = \vec{\theta} - \vec{\beta}$ may be measured at the multiple image positions $\vec{\theta}$ once source positions $\vec{\beta}$ are determined (see §??). However, in order to take its divergence, the deflection field must be solved for as a continuous function of position (or at least defined on a regular grid). Our interpolated deflection field must also be curl-free:

$$\nabla \times \vec{\alpha} = 0.$$  

(3.9)

The reason for this may be derived from Equation 3.3 as follows. We use the substitution $\nabla \ln \theta = \vec{\theta}/\theta^2$ to define the lensing potential

$$\psi(\vec{\theta}) = \frac{1}{\pi} \int d^2 \vec{\theta}' \kappa(\vec{\theta}') \ln \left| \vec{\theta}' - \vec{\theta} \right|,$$  

(3.10)

such that

$$\vec{\alpha} = \nabla \psi.$$  

(3.11)

The fact that the deflection field $\vec{\alpha}(\vec{\theta})$ may be written as the gradient of a scalar field $\psi(\vec{\theta})$ guarantees that it has no curl: $\vec{\alpha} = (\psi_x, \psi_y)$, $|\nabla \times \vec{\alpha}| = \psi_{yx} - \psi_{xy} = 0$, where $\psi_x = \partial \psi/\partial x$, etc.

So given measurements of $\vec{\alpha}$ at image positions $\hat{\vec{\theta}}$, we must find a curl-free interpolation. Unfortunately, this was not possible until just recently. Fuselier (2006) has developed a method to obtain a curl-free interpolation of a vector field given on scattered data points. This advance in the field of mathematics provides the machinery that makes LensPerfect possible.

In Chapter ??, we will describe the LensPerfect method in full. And in Chapter ??, we present the LensPerfect analysis of Abell 1689. But first we will address the issue of redshifts which enter directly into the lensing equations above and should be measured accurately to obtain accurate massmaps. After introducing the concept of redshifts (§14), we present our photometric redshift
analysis of galaxies in the Ultra Deep Field (§??). This paper describes in detail our measurement techniques. In §??, we apply these methods to obtain a photometric redshift catalog of galaxies in the Abell 1689 field, including the cluster itself, lensed background galaxies, and foreground galaxies.
Chapter 4

Photometric Redshifts

As we have seen, the accurate measurement of distances to other galaxies is key to many cosmological investigations including our own here. Knowledge of the distances to gravitationally lensed galaxies is a key ingredient to accurately mapping the Dark Matter structure of the lensing body.

So how do astronomers measure these distances? Our most precise tools involve measurement of galaxy “redshifts”, which result as light waves stretch with the expansion of the universe over time. Atoms and molecules emit light at very precise and well-known wavelengths. These emissions can be observed in spectra, with shifts from the known wavelengths betraying the object’s redshift. Such “spectroscopic redshifts” can be measured with relatively high accuracy. But the majority of galaxies detectable in deep images are too faint for spectra to be obtained. For these objects, we may fashion a very low resolution spectra from observations in broadband filters, and thus estimate a “photometric redshift”. A large portion of this thesis has been devoted to methods for obtaining such photometric redshifts. These methods were refined and tested on the Ultra Deep Field (mankind’s deepest optical images taken to date, Chapter ??) and then applied to galaxies gravitationally lensed by the galaxy cluster Abell 1689 (§??). Before presenting that research, we will briefly introduce the concept of redshifts and explain how they are measured.

4.1 Redshifts

A galaxy “redshift” may have three (possibly combined) causes: 1) Doppler shift due to its motion relative to ours, 2) the expansion of the universe since the light was emitted (so-called “Cosmological” redshift), and finally 3) Gravitational redshift incurred as light escapes from the gravitational field of a massive body. The second (Cosmological redshift) is that which allows us to measure distances, but we will begin by describing the first (Doppler redshift), as Doppler shifts are probably more familiar to most readers. The third (Gravitational redshift) is usually negligible in
galaxy redshifts (although this is the main cause of fluctuations observed in the Cosmic Microwave Background mentioned in §1.1).

Doppler shift is most famous for altering the apparent pitch of sound, but it also affects the observed wavelength of light. In fact, if a train were approaching us quickly enough, not only would its whistle sound much higher pitch (and beyond our range of hearing), but it would also appear bluer! Conversely, a train receding quickly away from us would have a lower pitch whistle and appear redder. While such “blueshift” or “redshift” may not be observable in Earth-bound trains, it is much more pronounced in nearby galaxies. Most galaxies in our Milky Way’s “Local Group” are observed to be blueshifted, from which we conclude that they are falling in toward us¹ (considerably faster than trains, but not fast enough to make collision with our galaxy imminent). Meanwhile, more distant galaxies are all observed to be receding from us, as part of the Hubble flow, or expansion of the universe.

But the Doppler shift makes only a small contribution to the observed redshifts of more distant galaxies, with the dominant component arising from the expansion of the universe itself. As the universe has expanded over time, so have all light waves within it. Light waves emitted long ago when the universe was half its size, for example, will be observed now as redshifted to double the wavelength. Measurement of a galaxy’s redshift is thus a measure of how much smaller the universe was (relative to the size today) back when the galaxy emitted the light which we are observing now. Given the known rate of expansion of the universe (§1.1), we can convert this measure of universe size into a measure of distance.

Redshift of light can also be produced by gravity, with light losing energy (and thus increasing in wavelength) as it escapes a body’s gravitational pull. Fortunately, such gravitational redshift is much smaller than that observed due to the expansion of the universe, and can be safely ignored.

Thus the question becomes how do we measure such redshift. Spectroscopic redshifts are the most accurate, routinely obtaining an accuracy of $\Delta z \sim 0.001$, where $z \equiv \Delta \lambda / \lambda$ is redshift. Emissions due to specific atoms and molecules are observed at precise wavelengths, with their redshifts from known wavelengths revealing object distances. Hydrogen atoms, for example, are known to emit light at various wavelengths both obtained from theory and measured here in labs on Earth. In a distant galaxy, these “emission lines” (and those due to other atoms / molecules) may be observed at the same relative wavelengths but redshifted from that observed on Earth (Fig. 4.1). It is these relative wavelengths that makes the identification (and thus redshift determination) possible. If the observed spectra has a poorer signal-to-noise ratio, and only one or two emission lines may be observed, then it becomes much more difficult (sometimes impossible) to match these lines with the atoms / molecules that produced them.

Thus for faint objects, hours upon hours of telescope time must be invested to obtain

¹To be precise, the Local Group galaxies are not falling in toward us, but to a point at the Local Group’s center of mass. This is deduced by tracking the Local Group’s galaxies’ slow but steady movement across the sky.
Figure 4.1 Absorption lines due to different atoms / molecules redshifted to $z = 2.534$, $z = 2.873$, and $z = 3.038$, compared to the spectra of image #1.1 of the Sextet Arc (Frye et al. 2007, reprinted from their Fig. 6). It is believed that this spectra shows absorption features from both the interstellar medium of the galaxy itself at $z = 3.038$ and two intervening systems at $z = 2.534$ and $z = 2.873$. Thus the identifications of features in this spectra is complicated, but this figure at least serves as a good example of the redshift of spectral lines.
quality spectra from which spectroscopic redshifts may be derived. For objects fainter still ($i > 25$, or greater than 25th magnitude in an $i$-band filter), quality spectra may be impossible to obtain no matter how long a telescope sits and stares at the objects. The great majority of objects in space-based images are, in fact, too faint to yield spectroscopic redshifts. Fortunately for these objects, another method is available to obtain redshifts.

The method of photometric redshifts takes advantage of the fact that galaxies generally have specific colors. Recently formed galaxies burn bright blue with the light of their hot young stars. Spiral galaxies like ours are a bit more aged and burn yellow like our Sun. Elliptical galaxies are oldest and thus burn reddest. Deviations (or redshifts) from these colors can thus be measured.

In practice, other variables intrinsic to a given galaxy can mimic these redshifts to some degree. These include interstellar dust and metallicity (content of atoms heavier than Hydrogen and Helium). Fortunately, we can actually make use of more information than a single color for each galaxy. Each star emits light in an entire spectrum of color, with a peak intensity at a wavelength that depends on the stars age. The hundreds of millions of stars in a given galaxy thus combine to produce a spectrum with a certain shape depending on the amounts of different-aged stars. We look for these expected shapes, or SEDs (spectral energy distributions), in observations taken through broad-band filters (Fig. 4.2). Of course when we find these SED shapes, they are redshifted from their expected wavelengths, again giving us a measure of distance to the galaxies.
Figure 4.2  BPZ SED fit to the photometry of a member of the galaxy cluster Abell 1689. Here photometry has been measured in 12 filters, making redshift determination especially robust. The observed fluxes are shown as filled circles, with the colors merely distinguishing between different telescopes. The SED is shown as the grey line and the fluxes of this SED as observed through the 12 filters are plotted as blue rectangles. This figure is reprinted and explained further in Fig. ??.
Bibliography


Bekenstein, J. D. 2004, Phys. Rev. D, 70, 083509 [ADS]


—. 2006c, Classical and Quantum Gravity, 23, 3427 [ADS]
Carr, J., Lamanna, G., & Lavalle, J. 2006, Reports of Progress in Physics, 69, 2475 [ADS]
Chwolson, O. 1924, Astronomische Nachrichten, 221, 329 [ADS]
de Bernardis, P., Ade, P. A. R., Bock, J. J., Bond, J. R., Borrill, J., Boscaleri, A., Coble, K.,
Crill, B. P., De Gasperis, G., Farese, P. C., Ferreira, P. G., Ganga, K., Giacometti, M., Hivon,
E., Hristov, V. V., Iacoangeli, A., Jaffe, A. H., Lange, A. E., Martinis, L., Masi, S., Mason,
P. V., Mauskopf, P. D., Melchiorri, A., Miglio, L., Montroy, T., Netterfield, C. B., Pascale, E.,
Piacentini, F., Pogosyan, D., Prunet, S., Rao, S., Romeo, G., Ruhl, J. E., Scaramuzzi, F., Sforna,
Diego, J. M., Sandvik, H. B., Protopapas, P., Tegmark, M., Benitez, N., & Broadhurst, T. 2005,
MNRAS, 362, 1247 [ADS]
Dressler, A., Lynden-Bell, D., Burstein, D., Davies, R. L., Faber, S. M., Terlevich, R., & Wegner,
DuVernois, M. A., Barwick, S. W., Beatty, J. J., Bhattacharyya, A., Bower, C. R., Chaput, C. J.,
Coutu, S., de Nolfo, G. A., Lowder, D. M., McKe, S., Müller, D., Musser, J. A., Nutter, S. L.,
Eddington, A. S. 1919, The Observatory, 42, 119 [ADS]
—. 1920, Space, time and gravitation. an outline of the general relativity theory (Cambridge Science
Classics, Cambridge: University Press, 1920) [ADS]
Einstein, A. 1916, Annalen der Physik, 354, 769 [ADS]
—. 1936, Science, 84, 506 [ADS]


Fuselier, E. J. 2006, PhD thesis, Texas A&M

Gaitskell, R. J. 2004, Annual Review of Nuclear and Particle Science, 54, 315

64


Hansen, S. H. & Stadel, J. 2006, Journal of Cosmology and Astro-Particle Physics, 5, 14 [ADS]


Newton, I., Wilson, P., & Maskelyne, N. 1782, Philosophical Transactions Series I, 72, 58 [ADS]


Renn, J., Sauer, T., & Stachel, J. 1997, Science, 275, 184


Roszkowski, L. 2004, Pramana, 62, 389 [ADS]


Schäfer, G. 2003, Advances in Space Research, 32, 1203 [ADS]

Schneider, P. 2006, in Saas-Fee Advanced Course 33: Gravitational Lensing: Strong, Weak and Micro, ed. G. Meylan, 1–89 [ADS]

Sérsic, J. L. 1968, Atlas de galaxias australes (Cordoba, Argentina: Observatorio Astronomico, 1968) [ADS]


68


Zwicky, F. 1933, Helvetica Physica Acta, 6, 110 [ADS]