Carnegie Observatories Astrophysics Series, Vol. 2:
Measuring and Modeling the Universe, 2004
ed. W. L. Freedman (Cambridge: Cambridge Univ. Press)

Dark Matter Theory

JOSEPH SILK
Department of Physics, University of Oxford

Abstract

I evaluate the dark matter budget and describe baryonic dark matter candidates and their detectability. Dark matter issues in galaxy formation theory are discussed, and I review the prospects for detecting nonbaryonic dark matter. Indirect detection via halo annihilations of the favored dark matter candidate, the SUSY LSP, provides a potential signal. The relic density of dark matter particles specifies the annihilation cross-section within model uncertainties, and indirect detection provides our optimal strategy for confirming the dark matter candidate. Galaxy formation simulations suggest that the predicted clumpiness of the dark halo will facilitate our imaging the dark matter in gamma rays, with cosmic ray signatures providing invaluable confirmation. Similarly, the supermassive black hole in the center of the Milky Way should present a unique signal amplifier by which we can view the dark matter in neutrinos as well as in gamma rays. The astrophysical uncertainties are so large that one has no alternative but to look.

1.1 Introduction

There are two types of dark matter: baryonic and nonbaryonic. Several candidates for the former exist, but the mass fraction is unknown. Conversely, while we have not yet detected the leading nonbaryonic matter (cold dark matter; CDM) candidate, the neutralino, we can calculate its mass fraction and interaction cross-section, within model uncertainties.

Dark matter dominates the Universe, amounting to $\sim 90\%$ of the matter density. Baryonic dark matter, as yet unambiguously identified, comprises up to a third of the baryons, although there are compelling candidates. Elucidating the nature of all of the dark matter is one of the outstanding problems in astrophysics to be addressed over the next decade. As will be discussed in this review, research at the interface of dark matter with galaxy formation has been particularly active in recent years, but has also raised challenges that are leading some to question the entire dark matter edifice.

Dark matter has a venerable history. In the solar system, anomalies in the orbit of Uranus pointed to dark matter in the form of a new planet, and this led to the discovery of the planet Neptune, following the predictions of Adams and Leverrier. The advance of the perihelion of Mercury's orbit also stimulated searches for dark matter in the form of a new planet interior to the orbit of Mercury, dubbed Vulcan. This turned out to be a red herring: the orbital anomalies eventually led Einstein to propose a new theory of gravitation, general relativity.

Similarly, there are parallels that may be drawn today. The modified Newtonian dynamics

(MOND) theory seeks to modify the law of gravity in order to dispense with the need for dark matter. There is little in the way of compelling theory or data to support such a position, but, at the same time, MOND is tenaciously difficult to kill (e.g., Sanders & McGaugh 2002). It is certainly worth bearing in mind that general relativity has been thoroughly tested only in the weak-field limit, and on relatively small scales, now extended up to a Mpc or so by gravitational lensing studies.

The modern dark matter problem was first described in 1933 by Fritz Zwicky (Zwicky 1933), who noted that in the Coma cluster of galaxies, the ratio of mass to light as measured by the virial theorem is about 400 $M_{\odot}/L_{\odot}$. Since individual galaxies are found to have about 10 $M_{\odot}/L_{\odot}$, there is a serious shortfall of luminous matter. When observations of clusters were extended to X-ray frequencies, a significant component of mass was found that was not detected optically. X-ray observations have revealed the presence of a substantial amount of hot intracluster diffuse gas, contributing about 15% of the cluster mass. However, 80% of the cluster mass is not accounted for in any known form. Moreover, this unknown mass component must be nonbaryonic, as the primordial nucleosynthesis measure of global baryon abundance, combined with the mean matter density, is consistent with the observed cluster baryon content in gas and stars.

On larger scales, one can probe dark matter via both Hubble-flow perturbations studied via deep redshift surveys and shear maps from weak gravitational lensing. The dark matter content dominates the total matter content. The general consensus is that $\Omega_m \approx 0.3$, with about 10% of this being in baryons. Most of the baryons must also be dark, although there are persuasive arguments about their nature.

I now describe in more detail the dark matter budget, describe baryonic dark matter candidates and their detectability, and then turn to CDM issues in galaxy formation theory. I conclude with a review of the prospects for detecting nonbaryonic dark matter.

1.2 Global Baryon Inventory

Primordial nucleosynthesis of the light elements demonstrates that the baryon fraction is $\Omega_b = 0.04 \pm 0.004$. In effect, this is measured at $z \approx 10^9$, when the light-element abundances freeze out. There are two other independent measures of the total baryon content of the Universe. The heights of the cosmic microwave background (CMB) acoustic peaks, in particular the odd peaks that correspond to wave compressions and rarefactions on the last-scattering surface scale and at half of this scale, are controlled primarily by the baryon density at $z \approx 1000$. The Lyα forest indirectly measures the baryon density at $z \approx 3$, once a correction is made for the predominant ionized fraction, whose density is inferred from ionization balance by invoking the ionizing radiation field from quasars. All three measures of the baryon density converge on $\Omega_b \approx 0.04$.

At the present epoch, the baryon inventory is rather different, however. Stars account for a modest fraction of all the baryons present, about 0.0026 in spheroid stars and 0.0015 in disk stars and cold gas, in units of the Einstein-de Sitter density (Fukugita, Hogan, & Peebles 1998). These numbers are in approximate agreement with a recent determination, including all cold interstellar matter and stars in low-mass as well as in massive galaxies in the local Universe, which yields 0.0024 ± 0.001, corresponding to $8\% \pm 4\%$ of Ω_b (Bell et al. 2003). Intracluster gas in rich clusters, mapped in X-rays, accounts for another 0.0026 of the closure density, consistent with the conjecture that, outside of the great clusters, most of the hot gas is able to cool and form galaxy disks. In total, only some 20% of the baryons is actually observed at low redshift.

There are strong theoretical indications, however, that most of the baryons, an additional 0.01 to 0.015 in units of the closure density, or a fraction 25% to 40% of Ω_b, are in the form of a warm/hot intergalactic medium (WHIM) in low-density environments outside the rich clusters. A quantitative estimate of the WHIM fraction comes from large-scale numerical simulations of the intergalactic medium (IGM; Davé et al. 2001). The heating is gravitational, due to accretion shocks in filaments and sheets, and generally occurs on the peripheries of galaxies, galaxy groups, and galaxy clusters. The WHIM has recently been detected via excess soft X-ray emission toward clusters of galaxies (Sołtan, Freyberg, & Hasinger 2002) and also via absorption in O VI toward quasars (Simcoe, Sargent, & Rauch 2002), although it is not yet possible to quantify the observed WHIM mass fraction.

The IGM simulations also suggest that there should be some cold intergalactic gas, visible as local, metal-poor, Lyα forest lines, although quantitative predictions are unreliable because most of the cold gas will presumably have formed stars, or at least fallen into galaxies. However, observations of the local Lyα-absorbing intergalactic clouds suggest that they could contribute no more than 0.008 of the closure density, or 20% of Ω_b, toward the baryon budget (Penton, Shull, & Stocke 2000). The net effect is that some 80% of the globally distributed baryons are plausibly accounted for, despite the fact that no more than 20% have hitherto been directly and quantitatively observed. The uncertainties in the gaseous baryon fraction, especially for the warm gas, are such that one can infer that there is no serious global dark baryon problem. However, the situation may be quite different on galactic scales.

1.3 Galactic Baryon Inventory

Rich clusters contain a reservoir of gas that reflects the initial baryon content of the Universe on 10 Mpc scales, and that, for the most part, consists of gas that has been frustrated from forming disks. Processes such as galaxy collisions and tidal harassment inhibit an accumulation of cooling gas in galaxy halos. This frustrates the presumed means by which disks ordinarily accrete cold gas. It is the maintenance of a long-lived supply of cold gas that enables star formation to continue over a Hubble time in low-density regions of the Universe.

The cluster gas fraction is approximately 15%. This plausibly reflects the initial baryon fraction in protodisks. Indeed, simulations of disk formation generally require an initial gas fraction of 15% to 20%. This is necessary for sufficient cooling to have occurred to be able to form the disks. Semi-analytic galaxy formation predicts that halos currently contain large amounts of gas, about as much as eventually forms the disk.

Cold clouds in the halo would be easily observable. The gas, if in the halo, must be at a temperature of several million degrees. However, there is equally a problem for diffuse gas in the halo: the resulting X-ray emission would exceed that observed by up to an order of magnitude (Benson et al. 2000), unless relatively large supernova feedback is implemented (Toft et al. 2002) to eject the gas. If this gas were to end up in the disk, the mass of the disk would presumably exceed observed limits from the rotation curve. Only a small fraction of the disk mass can be present in the halo as hot gas.

The Milky Way stellar disk mass is well constrained by a combination of gravitational microlensing and infrared observations (Klypin, Zhao, & Somerville 2002). The mass of the disk and bulge is $6 \times 10^{11} M_\odot$, whereas the total mass, predominantly in the halo as measured by the rotation curve out to the virial radius, is $10^{12} M_\odot$. If the dark halo profile is NFW-like (Navarro, Frenk, & White 1996), there is no room for any further baryonic component that might be in a noncompact form and not already measured by the powerful combination of bulge microlensing and the Galactic rotation curve. It has been argued that the dark halo must actually have a softer, more isothermal, core than the power-law NFW profile; otherwise, the contribution of the dark matter to the rotation curve in the inner galaxy would be excessive (Binney & Evans 2001).

This leaves us in a quandary: where have the baryons gone, which once were present? There are two possibilities. The baryons may still be present, but hidden. For example, they may be in dense, cold clumps in the outer halo, which are too large, by exceeding the Einstein radius, to give a strong microlensing signal. Or they could be in the form of compact objects, such as massive compact halo objects (MACHOs), since the observational limit on halo MACHOs obtained by microlensing of stars in the Large Magellanic Cloud allows a halo fraction in MACHOs of up to 20%, in compact clump masses below a few solar masses. The positive signal reported by one experiment indeed favors a MACHO mass of about $0.5 M_\odot$. In fact, one only needs to hide a halo mass fraction of about 6%, and possibly as large as 10%, to arrive at an initial baryon mass fraction of $\sim$15%. An alternative possibility is that the unaccounted baryons were ejected from the protogalaxy in an early wind.

I now consider these various possibilities in turn. Cold dense clumps of H_2, of Jupiter-like mass and sizes of order a few astronomical units, have been invoked to account for extreme scattering events (Walker & Wardle 1998) and for unidentified SCUBA submillimeter sources (Lawrence 2001). As such halo clumps orbit the Galaxy, they cross the Galactic plane up to 100 times over the age of the disk. These traversals result in the gas clumps acquiring on the order of 1 magnitude of extinction as they sweep up interstellar dust (Kerins, Binney, & Silk 2002). In order to avoid an excessive rate of collisions and yet keep the clumps large enough to be Jeans stable at a given mass and temperature, the clump covering factor must be on the order of 0.0001. Such clumps are potentially detectable via gaseous microlensing events (Rafikov & Draine 2001), as well as by occultations of background stars in a MACHO-type experiment that monitors millions of stars several times per night.

Of course, such clouds would cool to a few degrees Kelvin and collapse, unless heated, and the only plausible proposed heat source is cosmic rays. It is not clear if the clouds could maintain a stable equilibrium (Gerhard & Silk 1996; Wardle & Walker 1999), although the reemitted radiation is consistent with measurements of the far-infrared background (Sciama 2000). If the clumps collapse, the observational motivation for invoking them is removed.

The MACHO option most likely centers on halo white dwarfs. There have been claims and counterclaims of detections of old, high-velocity white dwarfs. The few detected have generally been attributed to the thick disk. If even a few percent of the halo mass was in the form of white dwarfs, a most unusual stellar initial mass function (IMF) would be required for the precursor stars, peaked in the intermediate-mass range. With a solar neighborhood IMF, one could hide only of order 0.1% of the stellar mass in the Milky Way disk in our halo. If the precursor stars had primordial abundances ($\lesssim 10^{-4}$ solar), one could plausibly suppress helium flashes and associated C or N dredge-up, thereby avoiding the obvious concern about stellar ejecta overpolluting the interstellar medium and protogalaxy by a factor of up to 1000 in the elements primarily produced by intermediate-mass stars. One would still perhaps have to tolerate a greatly enhanced rate of Type Ia supernovae (SNIa). However, the likelihood that most of the SNIa ejecta escape from the galaxy via winds, combined with the unknown scenario for SNIa formation, which itself invariably involves old white dwarfs in binary systems, renders such estimates highly uncertain.

1.4 Outflows

An alternative view is that galaxies such as our Milky Way have undergone massive outflows in the past. There is strong evidence that many galaxies presently undergoing starbursts are driving superwinds, with outflows on the order of the current star formation rate. The high-redshift evidence is especially compelling. At $z \approx 3$, the Lyman-break galaxies, many of which have high star formation rates, occasionally reveal inverse P Cygni profiles and, more generally, broad line widths, with inferred outflow velocities of the order of the escape velocity from the galaxy (Pettini et al. 2002; Shapley et al. 2003). Studies of the Lyα forest in the vicinity of the Lyman-break galaxies, via absorption in the spectra of background quasars, reveal Mpc-sized holes. These are seen both in the H I forest and in C IV absorption, centered on the Lyman-break galaxies (Adelberger et al. 2003). A vigorous wind could excavate such holes, although one cannot exclude a photoionization source.

It may be that spirals have expelled a mass in gas comparable to that remaining in stars, whereas, on the basis of gravitational lensing of background quasars, the time-delay evidence suggests that many massive ellipticals have not lost a significant fraction of their initial baryon content. Indeed, ellipticals need to have conserved the primordial baryon fraction, and possibly even accreted more cold baryons, in order to retain consistency both with the Hubble constant and a NFW-like profile (Kochanek 2003).

This would also be in accordance with the following conjecture, namely that the dark halos in massive ellipticals have pristine dark matter profiles, as generated by major mergers and predicted by N-body simulations, whereas the dark matter concentrations in the halos of many spirals may have been modified by the astrophysics of disk formation. Such modifications could plausibly occur as a consequence of the dynamical heating and associated early outflows, linked to the formation of massive, transient protogalactic gaseous bars (Weinberg & Katz 2002). An alternative possibility appeals to the formation of supermassive central black holes and associated early AGN activity that would have resulted in massive winds (Binney, Gerhard, & Silk 2001).

The desired modifications in halos of spirals, both massive and dwarf-like, consist of the following. The amount of substructure must be suppressed, both to be consistent with that viewed directly by counting visible dwarfs, and inferred indirectly by the near-conservation of the initial angular momentum acquired via tidal torques with neighboring protogalaxies that is necessary to account for the observed sizes of galactic disks. The concentration of dark matter within a couple of disk scale lengths may need to be reduced from the 50% or so by mass predicted by the simulations to the 10% or less inferred from microlensing for the Milky Way Galaxy and from dynamical friction of bars acting on the dark halos for spirals with rapidly rotating central bars (Debattista & Sellwood 2000).

Actually, the latter argument is controversial, as it is subject to ignoring destruction and reformation of bars that could occur several times over the disk age, if gas accretion occurs over a similar time scale and at about the same rate as the disk star formation rate (Bournaud & Combes 2002). This might allow bars to be relatively short-lived, decaying via dynamical friction against the dark matter if the halo profile is NFW-like, or decaying on a similar time scale due to the effects of destabilization via central gas accretion and bulge formation. Provided the disk that reformed via cold gas accretion was sufficiently cold, it would be unstable to bar formation. Presumably the stellar ages of bars would contain clues about the star formation history that could perhaps be unravelled by detailed spectrophotometry.

The dark halos of dwarf spirals generically seem to be mostly cuspless, in some cases in possible contradiction with the extent of a central cusp that is predicted by CDM simulations. The dark matter profiles can be fit either by CDM profiles with anomalously low concentrations or by dark matter profiles with nearly isothermal cores (de Blok & Bosma 2002). Again, this could be a consequence of vigorous early winds, which would be consistent with the low observed metallicities of dwarfs.

Finally, a substantial subset of massive ellipticals has soft cores. These galaxies are associated with slow rotation and boxy profiles. Numerical simulations suggest an origin via major mergers (Burkert & Naab 2004). One could plausibly imagine that such mergers resulted in substantial central gas flows and generated luminous starbursts that drove massive winds. Supermassive black hole formation by mergers of black holes of comparable mass would also heat and scour out the dark matter cusp, producing a soft core or even a central minimum in the stellar luminosity density (Lauer et al. 2002).

It is indeed possible that galaxies have undergone a wide variety of formation histories. Some of these, such as massive gas accretion onto disky ellipticals and major mergers to form boxy ellipticals, may have had a strong but possibly indirect influence on the dark matter profiles, including cusp, substructure and concentration.

One could possibly distinguish between these alternative hypotheses by the prediction that a massive early starburst would result in boosting the rate of Type II supernova production relative to the longer time scale for SNIa explosions, and hence lead to an $[\alpha/\mathrm{Fe}]$ enhancement in the stellar core. Boxy ellipticals should therefore display systematically higher $[\alpha/\mathrm{Fe}]$ than disky ellipticals, which often have central cusps, or power-law profiles in surface brightness, and have also presumably undergone more quiescent accretion and extended periods of star formation that led to disk formation.

Another lensing result from the image ratios of quadruple lens images of quasars can only be explained if the massive elliptical lenses have substructure on subgalactic mass scales (Dalal & Kochanek 2002). A similar conclusion is inferred from the bending of radio jets on milli-arcsecond scales (Metcalf 2002). Some massive ellipticals have substructure consistent with the full predicted power of CDM, that is to say of order a few percent of the halo dark matter in inhomogeneities, with much of this in million-solar mass clumps. Some of the massive ellipticals, those with disky isophotes, as well as low-mass ellipticals and bulges, often have central power-law stellar cusps. It would of course be interesting to know if these properties were indeed correlated, and characteristic of the disky ellipticals and spheroids that are thought not to have undergone major mergers.

There is one crucial issue to be addressed if massive winds with outflow rates on the order of the star formation rate are to be considered seriously in the context of galaxy formation. Simulations fail to produce such winds in massive galaxies. One finds that the winds interact with ambient gas, undergo strong radiative cooling, and are quenched. I have suggested that a fundamental problem with all existing wind simulations is the omission of crucial subgrid physics (Silk 2003). In particular, Rayleigh-Taylor instabilities, as the wind interacts with the surrounding matter and the shocked shell of dense, cooled gas decelerates, increase the interstellar medium porosity. This enables the galactic outflows to proceed with enhanced efficiency. Simultaneously, Kelvin-Helmholtz instabilities entrain cold interstellar gas into the outflows. Ablation and evaporation of the cold clouds adds to the mass loading of the wind. The enhanced efficiency means that even massive galaxies can have winds. The star formation rate peaks in the protogalactic phase. This is when the interstellar cold gas density is greatest, and hence cooling is most important. If a wind is indeed driven, the wind outflow is of order the star formation rate.

However, a successful wind requires that there be sufficient energy in the outflow to drive the gas out of the galactic potential well. Normal supernovae of Type II, associated with the starburst that characterizes the elevated star formation rate, may not suffice to generate enough energy if the galaxy escape velocity exceeds ~ 100 km s^{-1}. However, I have argued that even massive galaxies should have undergone a massive outflow phase. The resolution might lie with hypernovae, with characteristic kinetic energies $\sim 10^{53}$ erg. It has been suggested that hypernovae are associated with the deaths of stars in excess of 30 $M_\odot$, the high energies being generated via release of binding energy from infall onto the forming black hole. There are clues from the observed metallicities in at least one starburst galaxy and in metal-poor halo stars that hypernovae may make an important contribution to the observed chemical abundances. Indeed, hypernovae may even be the dominant contributors to the abundance patterns in extremely metal-poor halo stars (Umeda & Nomoto 2003). If indeed all stars above $\sim 30 M_\odot$ explode as hypernovae in metal-poor environments, the specific energy input into the interstellar medium would be enhanced relative to that of ordinary supernovae by an order of magnitude. This would suffice to drive winds from the most massive protogalaxies.

It has also been suggested that hypernovae are preferentially produced via stellar mergers (Portegies Zwart et al. 2002), which are prime candidates for producing the massive, rapid rotators preferred in hypernova models (Nakamura et al. 2001). Such mergers are likely to be most important in high-density cores associated with starbursts, and might be especially frequent during the protogalactic starburst phase associated with spheroid formation.

1.5 Prospects for Nonbaryonic Dark Matter Detection

CDM, despite the current issues that are being raised concerning substructure and dark matter concentration on small scales, has been remarkably successful in accounting for large-scale structure and in leading to the prediction of the amplitude of the CMB temperature fluctuations. However, detection, direct or indirect, has been elusive, at least for the dominant nonbaryonic dark matter component.

Massive neutrinos are the only form of nonbaryonic dark matter known to exist. However, they are subdominant. The relic neutrino number density is $\frac{3}{11} n_\gamma$, where the relic photon density n_γ is $0.24(kT/\hbar c)^3$ and T is the measured CMB temperature of $2.725 \pm 0.002\,$K. A lower bound to the mass comes from atmospheric $\nu_\mu \to \nu_e$ oscillations, and is $m_\nu \gtrsim 0.1\,$eV. Upper limits come from particle physics and from astronomy. Now, $\Omega_\nu h^2 = \Sigma m_\nu / 91.5\,$eV. The neutrinoless double beta decay limit on the electron neutrino mass, combined with recent reactor constraints (using KamLAND data) on oscillations, set an upper bound $\Omega_\nu h^2 \lesssim 0.07$ (Minakata & Sugiyama 2002). A similar, but stronger bound comes from large-scale structure, and in particular the linear power spectrum derived from the 2dF galaxy redshift survey, such that the sum of the neutrino masses must be less than 2.5 eV

(Elgarøy et al. 2002; Hannestad 2002). Neutrinos can be comparable to the baryon density in contributing to the matter content of the Universe.

Neutrinos decouple relativistically, since decoupling at $kT_{dec} \approx 1\,\mathrm{MeV} \approx 2m_e c^2$ is controlled by neutron freeze-out. Generic massive weakly interacting particle candidates that once were in thermal equilibrium decouple non-relativistically at $kT_\chi \approx m_\chi c^2/20$, and are suppressed in number density by the corresponding Boltzmann factor relative to the photon number density. There is theoretical prejudice that favors the lightest SUSY particle, which is stable if there is a commonly adopted symmetry among the SUSY family of particle partners to the known bosons and fermions. Constraints from the Large Electron Positron Collider set a lower bound on the particle mass of about 50 GeV. A generic upper bound comes from theory: the SUSY scale, and hence the WIMP mass, cannot be much above 1 TeV if SUSY is to be relevant for accounting for the electroweak scale. Since the cross-section for annihilation decreases in the unitarity limit with increasing mass above 100 GeV, one cannot go to too high a mass, typically less than a few TeV, without overclosing the Universe with WIMPs. Realistic models with the observed CDM density (Melchiorri & Silk 2002) $\Omega_{cdm} = 0.12 \pm 0.04h^{-2}$ and including all coannihilation channels (Edsjö et al. 2003) suggest an upper bound of 1500 GeV.

One may hope to detect CDM in the form of the LSP via its annihilations in the halo. The annihilation products include high-energy positrons, antiprotons, gamma rays, and neutrinos. If the halo is uniform, the predicted fluxes fall a factor of 100 or more below observational limits as set, for example, by the diffuse high galactic latitude EGRET gamma ray flux. However, secondary production of gamma rays via $p-p$ and $p-\alpha$ interactions of cosmic rays results in a spectral energy distribution that is softer than the observed diffuse gamma ray flux, suggesting that a more exotic explanation, such as that from annihilations, may possibly be merited.

High-resolution numerical simulations of realizations of galaxy halos predict strong clumpiness on small scales. This would help boost the expected annihilation signal. A reported high-energy positron flux shows a spectral excess relative to predicted cosmic ray secondary positrons near 50 GeV. With a boost factor of 100 to 1000, this could be due to WIMP annihilations in the halo, the fit to the data certainly being improved by a carefully tuned WIMP mass (Baltz et al. 2002).

One consequence is the prediction of a primary high-energy $\bar{p}$ feature that might be potentially observable with the forthcoming PAMELA and AMS experiments. The possible spectral signatures of the diffuse gamma ray and neutrino fluxes from halo annihilations are more elusive. With regard to the gamma rays, one expects a π^0 decay signature both from cosmic ray interactions with interstellar gas and from the annihilations, although the latter would produce a harder spectrum for massive WIMPs. However, one may hope to identify individual halo clumps as pointlike gamma ray sources at the sensitivity and angular resolution of the *GLAST* satellite, to be launched in 2007. Gamma ray lines are another potential signal, that would be a unique tracer of annihilations.

One concern is that the clumpiness of the dark halo may be greatly overestimated by the numerical simulations. This is because significant astrophysics is omitted from halo modeling. Clumps self-destruct via dynamical interactions with each other and especially with the Galactic tidal field. Nevertheless, even the neutrino flux from annihilations may be detectable with a more extreme strategy that targets the Galactic Center.

Observations strongly suggest that the formation of the spheroid stars is closely coupled to the formation of the central supermassive black hole. One observes that $M_\bullet \approx 10^{-3} M_{\rm sph}$ over a wide dynamic range, with only modest dispersion. In a disk galaxy such as the Milky Way, it is likely that the supermassive black hole at the Galactic Center of $\sim 2.6 \times 10^6 M_\odot$ (Ghez 2004) formed by a combination of gas accretion onto a seed black hole and merging of smaller black holes, rather than by major black hole merging events. Cosmology, and in particular studies of primordial star formation, suggests that the seed black hole masses are likely to be around $100 M_\odot$. Inferences about the primordial IMF lead to the likely presence in the protogalactic halo of large numbers of $100 M_\odot$ black holes (Islam, Taylor, & Silk 2002; Volonteri, Haardt, & Madau 2003). Gas accretion in the gas-rich protogalaxy is then the favored formation model for the central supermassive black hole.

Another option would be for the black hole and accompanying spheroid to form by hierarchical clustering and merging of roughly equal-mass black holes that are embedded in mini-spheroids of stars. This seems implausible for the Milky Way black hole and spheroid because the merging time scales would seem to be very long, and also because of the remarkable observed chemodynamical continuity between thin disk, thick disk, and spheroid stars. This is suggestive of a relatively nonviolent history. The last significant merger would have to have occurred at least 12 Gyr ago and preceded disk formation. The stellar ages and chemical properties of the disk and bulge suggest a coeval formation. In order to account for bar and bulge formation, secular evolution rather than merging is usually preferred in Milky Way-type galaxies, and more generally in late-type disks.

A rather different history may be applicable to massive spheroid formation. Of course, elliptical galaxies may well be older, on average, and form in denser environments. Major mergers that triggered bursts of star formation most likely played an important role in the formation of massive ellipticals, as inferred for example from the high central stellar surface densities and much circumstantial evidence from ultraluminous infrared galaxies. Our halo, however, appears to have retained the kinematic substructure characteristic of minor mergers (Gilmore, Wyse, & Norris 2002).

The Milky Way is an especially interesting case. There is a central bar, and there are also indications from the rotation curve and the gravitational microlensing optical depth that no more than 10% of the mass within two disk scale lengths can be nonbaryonic dark matter. The formation of the supermassive black hole adds a further complication that may modify the dark matter profile close to the center. In fact, we can take advantage of this, if the CDM consists of the lightest stable SUSY particle.

The presence of the supermassive black hole allows a potentially important probe of the dark matter density concentration in the inner galaxy. As the black hole forms within the preexisting dark matter potential well, a dark matter cusp develops within the zone of gravitational influence of the black hole, or about 0.1 pc. The cusp slope is $r^{-9/4}$ for a soft galactic core, and is steeper for an initially power-law profile in the galaxy inner halo. Dark matter annihilations are correspondingly boosted, within a radius between 0.1pc and about 100 Schwarzschild radii, within which the dark matter annihilates in less than a dynamical crossing time. One consequence is that a point source of high-energy neutrinos should be detectable toward the Galactic Center (Gondolo & Silk 1999; Ullio, Zhao, & Kamionkowski 2001). However, this is not the end of the story, for there are also radio and gamma ray signatures.

In fact, the advection-dominated accretion model for Sgr A*, for the flow in the accretion disk around the central black hole in the Milky Way, explains all of the observations except in the low-frequency radio and gamma ray regimes (Quataert & Narayan 1999). An additional nonthermal source of energetic particles is needed. Annihilations boosted by the black hole–enhanced cusp provide an attractive possible explanation.

The annihilations into e^+e^- pairs generate synchrotron emission. Remarkably, the predicted spectral shape, dominated by synchrotron self-absorption, matches that observed. Moreover the unidentified EGRET gamma ray source at the Galactic Center can also be fit spectrally, and the magnetic field strength can then be inferred by comparing the radio and gamma ray fluxes (Bertone, Sigl, & Silk 2002). The inferred field strength depends on the initial profile and the WIMP cross-section at a specified dark matter density, both leading to several orders of magnitude uncertainty in the required magnetic field strength, which in fact spans the equipartition estimate. In practice, the profile uncertainties dominate.

The observed radio and gamma ray luminosities can then be explained as being a consequence of the energy input from annihilations if the initial central core density profile is reasonably soft, as compared to the NFW profile. Detection of a neutrino source by the Northern hemisphere underwater neutrino telescopes ANTARES or NESTOR could eventually provide the smoking gun that confirms annihilations as a significant energy source in the inner 0.1 pc of the Milky Way.

1.6 Conclusions

Dark matter has been remarkably elusive. Despite 70 years over which the problem of the dark matter has been recognized, there has been no confirmation of its nature. However, observations are proceeding at a great pace, both from the observational astronomy side, especially, but not exclusively, via gravitational lensing, and on the experimental side by the enormous progress being made in direct and indirect detection experiments. There is increasing recognition that two of the greatest unresolved issues in physical cosmology, the nature of the dark matter and the formation of the galaxies, are intimately connected. Supermassive black holes and active galactic nuclei are important ingredients that need to be incorporated into our modeling before we can understand how protogalaxies evolve.

Imaging dark matter would be a wonderful achievement to crown our studies of these astrophysical puzzles. If our prejudices about dark matter are correct, annihilations provide a potential signal. Direct detection cannot compete: at the Large Hadron Collider any SUSY evidence will have no direct relevance for dark matter, although, of course, the complementarity is an essential strategy. Even direct detection experiments in deep underground laboratories are searching for a SUSY-inspired scattering signal that may be far below 10^{-10} picobarns, and so will be beyond any possible reach in the foreseeable future, where the goal for ton-scale detectors is of order 10^{-8} picobarns.

The relic density of dark matter particles specifies the annihilation cross-section within model uncertainties, and indirect detection provides our optimal strategy for confirming that the SUSY LSP is the dark matter candidate. It may well be that the clumpiness of the dark halo will facilitate our obtaining these images in the gamma ray domain. The cosmic ray signatures will provide invaluable confirmation, as will detection of high-energy neutrinos from the Sun that are generated by annihilations of WIMPs trapped in the solar core. The supermassive black hole in the center of the Milky Way may present a unique signal amplifier by which we can view the dark matter, in neutrinos as well as in gamma rays. The uncertainties in the halo fine-structure and in how the dark matter aggregated and the central supermassive black hole formed are simply too great to believe any theorist's assurance of the outcome. We must search.

Acknowledgements. I thank G. Bertone, G. Bryan, P. Podsiadlowski, A. Slyz, and J. Taylor for helpful discussions.

References

Adelberger, K. L., Steidel, C. C., Shapley, A. E., & Pettini, M. 2003, ApJ, 584, 45
Baltz, E., Edsjö, J., Freese, K. & Gondolo, P. 2002, Phys. Rev. D, 65, 063511
Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D. 2003, ApJ, 585, L117
Benson, A. J., Bower, R. G., Frenk, C. S., & White, S. D. M. 2000, MNRAS, 314, 557
Bertone, G., Sigl, G., & Silk, J. 2002, MNRAS, 337, 98
Binney, J., & Evans, N. W. 2001, MNRAS, 327, L27
Binney, J., Gerhard, O. E., & Silk, J. 2001, MNRAS, 321, 471
Bournaud, F., & Combes, F. 2002, A&A, 394, L35
Burkert, A., & Naab, T. 2004, in Carnegie Observatories Astrophysics Series, Vol. 1: Coevolution of Black Holes and Galaxies, ed. L. C. Ho (Cambridge: Cambridge Univ. Press), in press
Dalal, N., & Kochanek, C. S. 2002, ApJ, 572, 25
Davé, R., et al. 2001, ApJ, 552, 528
Debattista, V. P., & Sellwood, J. A. 2000, ApJ, 543, 704
de Blok, W. J. G., & Bosma, A. 2002, A&A, 385, 816
Edsjö, J., Schelke, M., Ullio, P., & Gondolo, P. 2003, J. Cosmol. Astropart. Phys., 04, 001
Elgarø, Ø., et al. 2002, Phys. Rev. Lett., 89, 061301
Fukugita, M., Hogan, C. J., & Peebles, P. J. E. 1998, ApJ, 503, 518
Gerhard, O. E., & Silk, J. 1996, ApJ, 472, 34
Ghez, A. M. 2004, in Carnegie Observatories Astrophysics Series, Vol. 1: Coevolution of Black Holes and Galaxies, ed. L. C. Ho (Cambridge: Cambridge Univ. Press), in press
Gilmore, G., Wyse, R. F. G., & Norris, J. E. 2002, ApJ, 574, L39
Gondolo, P., & Silk, J. 1999, Phys. Rev. Lett., 83, 1719
Hannestad, S. 2002, Phys. Rev. D, 66, 125011
Islam, R. R., Taylor, J. E., & Silk, J. 2003, MNRAS, 340, 647
Kerins, E., Binney, J., & Silk, J. 2002, MNRAS, 332, L29
Klypin, A. A., Zhao, H. S., & Somerville, R. S. 2002, ApJ, 573, 597
Kochanek, C. S. 2003, ApJ, 583, 49
Lauer, T. R., et al. 2002, AJ, 124, 1975
Lawrence, A. 2001, MNRAS, 323, L147
Melchiorri, A., & Silk, J. 2002, Phys. Rev. D, 66, 041301
Metcalf, R. B. 2002, ApJ, 580, 696
Minakata, H., & Sugiyama, H. 2002, Phys. Lett. B, 532, 275
Nakamura, T., Umeda, H., Iwamoto, K., Nomoto, K., Hashimoto, M., Hix, W. R., & Thielemann, F.-K. 2001, ApJ, 555, 880
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563
Penton, S. V., Shull, J. M., & Stocke, J. T. 2000, ApJ, 544, 150
Pettini, M., Rix, S. A., Steidel, C. C., Adelberger, K. L., Hunt, M. P., & Shapley, A. E. 2002, ApJ, 569, 742
Portegies Zwart, S. F., Makino, J., McMillan, S. L. W., & Hut, P. 2002, in Stellar Collisions, Mergers and their Consequences, ed. M. Shara (San Francisco: ASP), 95
Quataert, E., & Narayan, R. 1999, ApJ, 520, 298
Rafikov, R. R., & Draine, B. T. 2001, ApJ, 547, 207
Sanders, R. H., & McGaugh, S. 2002, ARA&A, 40, 263
Sciama, D. W. 2000, MNRAS, 312, 33
Shapley, A., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, ApJ, 588, 65
Silk, J. 2003, MNRAS, 343, 249
Simcoe, R., Sargent, W. L. W., & Rauch, M. 2002, ApJ, 578, 737
Sołtan, A. M., Freyberg, M., & Hasinger, G. 2002, A&A, 395, 475
Toft, S., Rasmussen, J., Sommer-Larsen, J., & Pedersen, K. 2002, MNRAS, 335, 799
Ullio, P., Zhao, H. S., & Kamionkowski, M. 2001, Phys. Rev. D, 64, 3504

Umeda, H., & Nomoto, K. 2003, Nature, 422, 871
Volonteri, M., Haardt, F., & Madau, P. 2003, ApJ, 582, 559
Walker, M., & Wardle, M. 1998, ApJ, 498, L125
Wardle, M., & Walker, M. 1999, ApJ, 527, L109
Weinberg, M. D., & Katz, N. 2002, ApJ, 580, 627
Zwicky, F. 1933, Helv. Phys. Acta, 6, 110