

Dark Matter and Galaxy Formation: Challenges for the Next Decade

Joseph Silk

Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 4LN

Abstract. The origin of the galaxies represents an important focus of current cosmological research, both observational and theoretical. Its resolution involves a comprehensive understanding of star formation, galaxy dynamics, the cosmology of the very early universe, and the nature of the dark matter. In this review, I will focus on those aspects of dark matter that are relevant for understanding galaxy formation, and describe the outlook for detecting the most elusive component, non-baryonic dark matter.

INTRODUCTION

Dark matter and galaxy formation are intimately related. This applies equally to baryonic and to nonbaryonic dark matter. In this talk, I will review the global budget for baryons and discuss the issue of dark baryons. I will describe the role of nonbaryonic dark matter in galaxy formation, and give an overview of the prospects for detection of cold dark matter.

A confluence of data on the cosmic microwave background temperature fluctuations, large-scale galaxy redshift surveys, quasar absorption line structure of the intergalactic medium, and distant supernovae of Type Ia have led to unprecedented precision in specifying the cosmological parameters, including the matter and energy content of the universe. The universe is spatially flat, $\Omega = 1.02 \pm 0.02$, and dominated by dark energy $\Omega_\lambda = 0.70 \pm 0.3$ with equation of state $w \equiv \frac{p}{\rho c^2} = -1.02 \pm 0.16$, nonbaryonic dark matter amounting to $\Omega_m = 0.27 \pm 0.07$, and the baryon content $= 0.0044 \pm 0.004$. The latter number incorporates a value of the Hubble content $H_0 = 72 \pm 5 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

A major assumption underlying the quoted errors is the adoption of priors. In particular, primordial gaussian adiabatic, scale-invariant density fluctuations are adopted. If, for example, an admixture of 30 per cent isocurvature fluctuations is included, consistency with CMB data is still obtained but the error bars are expanded by up to an order of magnitude [1]. Another assumption is that the fine-structure constant is actually constant. Allowing this to vary also gives further freedom, especially in the baryon density.

A strong case for the dominance of dark matter in galaxy clusters was made as long ago as 1933. It is remarkable that our understanding of its nature has not advanced since then. Of course, modern observations have led to an increasingly sophisticated exploration of the distribution of dark matter, now confirmed to be a dominant component relative to baryonic matter over scales ranging from those of galaxy halos to that of the particle horizon.

2. GLOBAL BARYON INVENTORY

There are three methods for determining the baryon fraction in the high redshift universe. The traditional approach is via primordial nucleosynthesis of ^4He , ^2H and ^7Li . The primary uncertainties lie in the systematic errors associated with ionisation corrections for ^4He , and extrapolation to primordial values via corrections for synthesis of ^4He and destruction of ^2H and ^7Li in stars. A unique value of $\Omega_b = 0.04 \pm 0.02$ is generally consistent with recent data, although there is some tension between ^2H , on the one hand, which in principle is the most sensitive baryometer and favours a higher Ω_b , and both ^4He and ^7Li . This tension has recently been increased [2] by the demonstration of a $[\text{Li}/\text{Fe}]$ gradient of $\sim \frac{0.1 \text{ dex}}{\text{dex}}$ in extremely metal-poor halo stars with $[\frac{Fe}{H}] < -2$, indicative of a role for pregalactic stellar destruction of primordial Li, as well as by determinations of $\frac{^6\text{Li}}{^7\text{Li}} = 0.05 - 0.08$ that indicate a 10-15% spallation

contribution to ${}^7\text{Li}$ in this metallicity range. Hence more generous error bars may be preferred, at least until the role of systematic effects such as atmospheric depletion are fully understood.

A completely independent probe of Ω_b comes from measuring the relative heights of the first 3 peaks in the acoustic temperature fluctuations of the cosmic microwave background. With the conventional priors, the data yields excellent agreement between the baryon abundance at $z \approx 1000$ and $z \approx 10^9$. Relaxation of the priors increases the error bars, but the central value is relatively robust. Yet another independent measure of Ω_b , this time at $z \approx 3$, comes from modelling the Lyman alpha forest of the intergalactic medium. This depends on the square root of the ionizing photon flux, in this redshift range due predominantly to quasars. The inferred value of Ω_b is again 0.04, with an uncertainty of perhaps 50%. Finally at $z \sim 0$, one only has a reliable measure of the primordial baryon fraction in galaxy clusters, which may be considered to be laboratories that have retained their primordial baryon fraction. The observed baryon fraction in massive clusters is about 15%, which is consistent with $\Omega_b = 0.04$ for $\Omega_m = 0.28$, the WMAP-preferred value.

Let us now evaluate the baryon fraction at the present epoch, both on galactic scales and in the general field environment. The following is an updated summary of the baryon budget recently presented by Fukugita and Peebles [3].

Stars in galactic spheroids account for about twice as much baryonic mass as do stars in disks. Disks dominate the (blue) light but spheroids have higher mass-to-light ratios. The total stellar contribution is about 15% of the total baryonic abundance of 0.04. Rich clusters only account for 5% of the galaxies in the universe, and so all of the hot diffuse gas in clusters, which account for 90% of cluster baryons, only accounts for about 5% of the total baryonic budget.

Cold intergalactic gas at the current epoch is mapped out in Lyman alpha absorption towards quasars. Identified with the Lyman alpha forest observed at high redshift, the low redshift counterpart is sparser. Its detection is more difficult, requiring a UV telescope such as HST or FUSE. However it is found to dominate the known baryon fraction today, and amounts to about 30% of the total baryon fraction [4]

In summary, some fifty percent of the baryons in the local universe have been detected and mapped. There are indications, motivated as much by theory as by observations at this stage, that the remaining baryons are in the warm intergalactic medium (WIM) at a temperature of $10^5 - 10^6\text{K}$. Simulations of structure formation indicate that some intergalactic gas is shocked to a temperature of $10^5 - 10^6\text{K}$. Much of this gas has not yet fallen into galaxies. According to the simulations, up to about 30% of the baryons are heated by the present epoch and remain diffuse. This fraction is an upper limit because the simulations lack adequate resolution, and moreover the amount of shock-heated gas is controversial [5]. Even more significantly, the theory of galaxy formation, as currently formulated, predicts that the WIM is metal-poor, in that those galaxies where most of the stellar mass resides, namely the massive galaxies, are energetically incapable of ejecting very much in the way of metal-enriched debris [6].

However, observations are confirming the existence of some WIM, in particular via detection of redshifted rest-frame UV OVI absorption towards quasars, extended soft x-ray emission near clusters [7], and OVII/OVIII x-ray absorption along lines of sight to AGN. The oxygen abundance exceeds $[\text{O}/\text{H}] > -1.5$ at $z \sim 2.5$ [8]. In practice, too few lines of sight have so far been probed to say a great deal about the WIM mass fraction.

In summary, something like 80 percent of the baryons at present have either been detected or are plausibly present with detection being imminent. One could conclude

$$\Omega_{b,observed} = 0.032 \pm 0.005.$$

Clearly the case for 10 - 20% of the local baryons being unaccounted for and dark is possible but far from convincing given the WIM uncertainties. If the WIM is indeed the dominant gas reservoir, there are strong implications for feedback from galaxy formation, in order to account for the observed enrichment of the WIM. Strong enrichment is indeed found for the intracluster medium, and this most likely is a consequence of early galaxy outflows. However the generation of these outflows is not understood.

One clearly needs to establish a more convincing case for the WIM before pursuing the impact of massive gas outflows on the early evolution of the typical field galaxy. Nevertheless, since the possible mass in unaccounted-for dark baryons is on the order of the baryon mass in stars, it is clear that such a result would profoundly affect our theories of galaxy formation and evolution. Hence demonstrating that these baryons are not present in the Milky Way is a useful exercise.

3. CONFIRMATION OF BARYONIC SHORTFALL

A detailed census of both the Milky Way and M31 confirms the lack of baryons in the amount predicted by primordial nucleosynthesis. The virial mass measured dynamically for the Milky Way from the HI rotation curve, dwarf galaxy orbits, and globular cluster peculiar velocities, amounts to $\sim 10^{12}M_{\odot}$. This is valid to a galactocentric radius of 100kpc. The baryon mass, including stars and gas is $(6-8) \times 10^{10}M_{\odot}$. However, the expected baryon fraction, both as observed at high redshift and in galaxy clusters, and especially as inferred from primordial nucleosynthesis and the CMB data, is about 17%. This is the initial baryon fraction when the Milky Way formed. A similar shortfall, amounting to a factor of about 2, is found for M31.

There are two possibilities for the "missing" baryons. Either they are present in the galaxy halo and as yet undetected, or they have been ejected via energetic outflows early in the history of the galaxy. Intensive searches for compact halo objects have been performed via gravitational microlensing of several million stars in the Magellanic Clouds. The EROS and MACHO experiments set the following limits, for more than 5 years of data: no more than 20 percent of the dark halo mass can be in objects in the mass range $\sim 10^{-8}M_{\odot}$ to $\sim 10M_{\odot}$, with a detection claimed by the MACHO experiment that saturates this limit for objects of mass $\sim 0.5M_{\odot}$. [10, 9]

The most plausible candidate for MACHOs of this mass are old halo white dwarfs. This requires a stellar initial mass function for the protogalaxy that forms the first stars with high efficiency in a narrow mass range ($4-8M_{\odot}$). While this seems implausible, it cannot be ruled out by theoretical arguments, one possible signature being that of occasional Type Ia supernovae. However old white dwarfs are still emitting light, albeit weakly, at visible wavelengths, and proper motion searches for faint candidates have imposed strong limits on the halo white dwarf mass fraction of between $\sim 2\%$ relative to the local dark matter density [11] and $\sim 0.2\%$ [12]. It seems reasonable to conclude that halo white dwarfs cannot account for more than a quarter of the unaccounted-for baryons, and this is most likely an overestimate.

One can imagine even less credible initial mass functions that would allow, say, ten percent of the dark halo to consist of primordial brown dwarfs, low mass primordial black holes, or even compact dense clouds of cold molecular gas. All of these possibilities have been studied as possible explanations for halo dark matter. Even if one's goal is only to account for halo baryonic dark matter, requiring even $10^{10}M_{\odot}$ to be in such a form stretches astrophysical credibility. But this cannot be ruled out.

A more plausible direction for investigation is that the "missing" baryons have been ejected from the galaxy, in the form of a vigorous, early galactic wind. Such a wind, if it occurs presently, could involve very little mass outflow. Observations indicate that at the present epoch, vigorous winds are exceedingly rare, and are seen only in low mass, star-bursting galaxies. In the early galaxy, however, the star formation rate was much higher, and the situation could have been quite different with regards to mass loss. Evidence for early winds comes indirectly from the highly enriched intracluster medium, whose mass exceeds that in the stellar component of cluster galaxies by a factor of several. The substantial amount of metals in the intracluster gas, and even the presence of magnetic fields, are most likely accounted for via ejection in early galactic winds.

At high redshift, the substantial population of the Lyman break galaxies (LBG) at $z \sim 3-4$ show broad linewidths displaced systematically to the blue by several hundred kilometres per second for the interstellar gas relative to the absorption lines of the stars [13]. Moreover, stacked spectral energy distributions of LBGs seen in projection near background quasars show evidence of a proximity effect, with a ~ 1 Mpc hole (comoving) inferred from the lack of Ly α and CIV absorption [14]. An energetic wind from galaxies with stellar mass similar to that of the Milky Way is inferred to have occurred, or at least, to provide the simplest explanation of these observations. Some of these galaxies most likely are massive, as their spatial clustering strongly favours their being the precursors of low redshift ellipticals [15].

The principal counterargument comes from wind simulations. While it is unanimously agreed that dwarf galaxies, with masses below $10^7-10^8M_{\odot}$, and escape velocities below 50kms^{-1} , are easily stripped of gas by supernova-driven winds, problems arise in driving winds from more massive galaxies. For disk galaxies, it is found that even for galaxies of mass $10^9-10^{10}M_{\odot}$, the supernovae ejecta stream out in a hot wind but most of the interstellar gas remains in the disk [16].

For forming galaxies, when the gas is more spherically distributed, ejection in a wind becomes inefficient for masses above about $10^{10}M_{\odot}$, according to the most recent multi-phase interstellar medium simulations [6]. These simulations adopt current supernovae rates and energetics per unit baryonic mass, along with a solar neighbourhood initial mass function, that is to say a rate of type II supernovae of 10^{51} energy input per $200M_{\odot}$ of gas that forms stars. This rate assumes a local fit to the initial mass function [17].

However in addition to the observational indications, semi-analytical galaxy formation theory requires a wind to have ejected approximately half of the baryons from even the most massive galaxies. Otherwise, one finds that

almost all of the gas that can cool within a Hubble time does cool and form stars, and the predicted luminosity function strongly disagrees with observations for luminosities above 2-3 times the galaxy characteristic luminosity, $L_* \sim 10^{10} M_\odot$ [18]. Related model malfunctions include unacceptably recent and inefficient star formation for distant massive galaxies as studied in deep surveys [19].

4. WHAT COULD BE WRONG WITH THE SIMULATIONS?

The numerical simulations of galactic outflows must cope with a variety of hydrodynamical and gravitational processes, including star formation, supernovae explosions, gas heating and cooling in a multi-phase interstellar medium, and gas escape from the galactic gravitational field. Hitherto, it has been necessary to severely approximate much of the relevant physics. For massive galaxies, winds are suppressed as the outflowing heated gas runs into surrounding, cold infalling gas, and most of the energy input is radiated away. Only about 2 percent of the initial supernovae energy is useful for expelling gas.

The situation may not be as bleak as depicted by the simulations. One omission due to lack of resolution is the effect of both Rayleigh-Taylor and Kelvin-Helmholtz instabilities. The former help the hot medium break out of the galaxy and enhance the wind efficiency. The latter enhances entrainment of the cold gas into the hot supernova-heated medium and can help account for the observed enrichment of the intergalactic medium. While the situation with regard to outflows may be alleviated in this fashion for low and intermediate mass galaxies, more drastic measures are required for massive galaxies. These may include any of the following: use of a top-heavy initial stellar function to enhance the supernova rate or appeal to an increased frequency of hypernovae at early epochs relative to supernovae, or finally, recourse to outflows from active galactic nuclei. Any or all of these may occur.

If indeed substantial mass loss via a wind occurs, then a plausible ansatz is that $\dot{M}_{outflow} \sim \dot{M}_*$ as observed in nearby starbursts, where the mass injection rate into the hot x-ray emitting diffuse gas is comparable to the star formation rate [20]. This means that about as much gas is ejected as is retained in stars. Such a conclusion is consistent with the observed baryon fraction in the Milky Way and M31, the two best-studied moderately massive galaxies. One can also understand the heavy element abundance observed both in the intercluster medium and in the warm ($T \sim 10^6$ K) intergalactic medium detected in OVI absorption. While the baryon fraction is probably not a major problem for consensus cosmology, I now turn to the issue of cold dark matter, and its relation to structure formation.

5. GALAXY FORMATION AND CDM: THE GOOD, THE BAD AND THE UGLY

There are some noteworthy success stories for cold dark matter (CDM). First and foremost is its success in predicting the initial candidates for structure formation that culminated in the discovery of the cosmic microwave background temperature fluctuations. The amplitude of the Sachs-Wolfe effect was predicted to within a factor of 2, under the assumption, inspired qualitatively by inflation, but quantitatively by the theory of structure formation via gravitational instability in the expanding universe, of adiabatic scale-invariant initial density fluctuations. A direct confrontation with this theory was first met with the detection and mapping of the acoustic peaks. These are the hallmarks of galaxy formation, first predicted some three decades previously, and demonstrate the imprint of the density fluctuation initial conditions on the last scattering surface of the CMB at $z \sim 1000$.

Another dramatic demonstration of the essential validity of CDM has come from the simulations of the large-scale structure of the universe. The initial conditions, including gaussianity, are specified, growth occurs by gravitational instability, and the sole requirements on dark matter are that it be weakly interacting and cold. Thus was born CDM, and the CDM scenario works so well that we cannot easily distinguish the artificial universe from the actual universe mapped via redshift surveys. More to the point, perhaps, is that the simulations are used to generate mock galaxy catalogues and maps that yield precise values of the cosmological parameters, in combination with the CMB maps.

Dark matter-dominated halos of galaxies are another generic success of CDM, as mapped out by rotation curves. However the detailed predicted properties of halos do not seem to be well-matched to observations. There is considerable scatter in the predictions of high resolution simulations for the structure of galaxy halos. Nevertheless, the predicted dark matter cusps ($\rho \propto r^{-\alpha}$ with $1 < \alpha < 1.5$) are not found in most low surface brightness dwarfs, nor is the predicted dark matter concentration ($C \equiv r_{200}/r_s \sim 5 - 10$, where r_{200} is the radius at density contrast 200 and r_s is the halo scale length) consistent with the dark matter distribution in barred galaxies, possibly including our own

galaxy, nor finally is the predicted number of satellites similar to the observed satellite frequency. In general, many observed halos seem to have softer cores, lower concentrations, and less clumpiness than predicted by the simulations.

However it has been argued that inclination, triaxiality and non-circular orbits make the dwarf situation unclear [22], quite apart from the fact that dwarf galaxy formation is not understood. Of course the same may be said for bars. The situation for early-type galaxies is at least as controversial. Indeed, for very round ellipticals, at least in projection, in low density environments and not especially luminous, studies of the distribution and kinematics of planetary nebulae suggest that mass traces light to ~ 5 effective radii [21]. However, the opposite conclusion is inferred for massive early-type galaxies, which display evidence for as much as a 50% contribution of dark matter within ~ 1 effective radius [23].

All of these issues have been debated. For example, reformation of bars by gas infall can avoid the problem of bar spin-down by dynamical friction, and astrophysical processes, discussed below, can render the dwarf satellites optically invisible. Hence it is difficult to be definitive about any possible contradiction between theory and observation. Certainly, on the baryonic front, the most accepted problem is the loss of angular momentum by the contracting and cooling baryons in the dark halo. The resulting disks are far too small. These various difficulties for galaxy formation theory have stimulated a variety of responses.

Resurrection via modifying fundamental physics

Suppose that one changes the nature of the dark matter. Increasing the scattering cross-section helps alleviate several of the problems, such as cusps and clumpiness. However the resulting dark halos are too spherical. Another approach modifies the law of gravity. Indeed, one may be able to dispense entirely with dark matter. These approaches seem rather drastic, however, and I believe that one should argue that all alternatives should be fully explored before tinkering with fundamental physics.

Resurrection via astrophysics

The obvious addition is stellar feedback. This can heat the baryons, and help reduce the loss of angular momentum. If the feedback is strong, mass loss is a likely outcome. The observed baryon fraction and the galaxy luminosity function for the most luminous galaxies both point to a possible loss of half the baryons during the galaxy formation process [24]. However to eject up to half the baryons may require more than normal stellar feedback, at least for galaxies comparable to, or more massive than, the Milky Way. One can appeal to a top-heavy IMF that would yield up to an order-of-magnitude more supernovae per unit mass of baryons, to an augmented fraction of hypernovae relative to supernovae, or to outflow generated by Eddington luminosity-limited accretion onto a supermassive black hole. Outflows may also be effective at reducing the dark matter concentration, at least for dwarfs [25].

Production of a soft core is best achieved for a massive galaxy by dynamical heating, as has been studied for the case of a rapidly rotating central baryonic bar [26], although a contrary view is expressed in [27]. Such bars are likely to be generic to galaxy formation via mergers, and if gaseous would leave little in the way of stellar tracers. Dynamical feedback also occurs via tidal evolution, and this can account for both the frequency and distribution of dwarf galaxies [28].

6. OBSERVING COLD DARK MATTER

The best way forward is to directly measure the halo properties by observing cold dark matter directly or indirectly. Direct detection is sensitive both to the local density of CDM and to its local phase space density. There is a candidate, motivated by supersymmetry, the LSP, usually considered to be massive with $m_{\chi} \sim 100\text{GeV}$, the SUSY breaking scale, and generically known as the neutralino or WIMP. However light LSPs, such as the axino, are also possible, and there is even a LSP with purely gravitational interactions, the gravitino. However, in general, the WIMP undergoes elastic interactions with ordinary matter and is therefore potentially detectable via laboratory experiments. Early universe freezeout yields a mass estimate; more specifically, the annihilation cross-section is inferred to be of order $\Omega_{\chi}^{-1} \sigma_{\text{weak}}$, and depends, via SUSY, on the WIMP mass. The corresponding elastic cross-section is model-dependent, but most models spans the range 10^{-10} to 10^{-6}pb for a relic abundance $\Omega_{\chi} h^2 \approx 0.1$.

Direct detection

Scattering of WIMP particles leads to nuclear recoils that can be measured by three different techniques: scintillation, phonon production, and ionization. The various experiments currently underway use different combinations of these techniques. Only one experiment, now running for 7 years, has reported a positive result, using NaI scintillation and a claimed detection of annual modulation, to yield a model-dependent detection of $m_\chi = 50(\pm 10)\text{GeV}$ with a cross-section of $7(\pm 1) \times 10^{-6}\text{pb}$. However other experiments, including Edelweiss, ZEPLIN and CDMS2, report a lower upper bound in the cross-section, with the more recent limit being $\sigma_\chi < 4 \times 10^{-7}\text{pb}$ at 60 GeV [29].

Indirect detection

Annihilations currently occur in the dark halo, although the annihilation time-scale $\sim (n_{h<\sigma v>_{ann}})^{-1} \sim 10^{26}(T_f/\text{GeV})^{3/2}\text{s}$, where T_f is the freeze-out temperature. The annihilation products are potentially observable in the form of high energy γ, e^+, \bar{p} and ν , and are enhanced by the effects of halo clumpiness. There are tentative indications of possible detections of e^+ and γ . A positron feature $\frac{e^+}{e^++e^-}$ is seen above 10GeV that cannot easily be attributed to secondary production of e^+ . A modest clumpiness boost is required for the measured flux to lie in the range allowed by annihilation models combined with cosmic ray diffusion [30]. Both the high galactic latitude gamma ray background and the unresolved diffuse gamma ray flux towards the galactic centre have relatively hard spectra that seems to be inconsistent with cosmic ray spallation and the ensuing π^0 decays. One possible explanation is in terms of population of hitherto unresolved discrete gamma ray sources, such as blazars in the extragalactic case or low mass x-ray binaries in the galactic case [33]. Similar boost factors, of 10-100, from dark matter clumpiness are required to that invoked for positron annihilation, if both the extragalactic and galactic diffuse gamma ray components have a WIMP annihilation origin.

A radical suggestion

The Integral SPI detector has measured a substantial diffuse flux of electron-positron annihilation line emission at 511 keV from throughout the galactic bulge. Some 10^{43} photons s^{-1} are generated over a region that extends up to 3 kpc from the galactic centre. There is no indication of any positron annihilation emission from any bulge source, such as might be connected with decays of Type II supernova-ejected radioactive ^{26}Al or $e^+ - e^-$ jets from x-ray binaries. This therefore has led to consideration of CDM annihilation as a possible explanation [31].

The principal novelty of such a hypothesis arises with the mass required for the annihilating particle. It must have a mass of ~ 10 MeV, as a much heavier particle would annihilate via pion production and produce an excessive flux of diffuse gamma rays from π^0 decays. From the measured flux and angular distribution, one immediately infers the required cross-section and radial profile, namely $\sigma_{ann} \sim 10^{-5}\text{pb}$ and $\rho_\chi \propto r^{-1/2}$. The profile is close to what is expected from CDM models, as inferred from rotation curve and microlensing modelling (actually, the derived CDM profiles are disputed for the Milky Way but a profile softer than NFW is inferred for barred galaxies and for LSB dwarfs). The required cross-section is very low, however, compared with the freeze-out value at $kT \sim m_\chi/20$, namely $\sigma_{ann} \approx (0.2/\Omega_\chi)\text{pb}$. One can reconcile the observed low annihilation cross-section required for the 511keV flux by assuming that the relativistic freeze-out limit is S-wave suppressed, so that $\sigma_{ann} \propto \left(\frac{m_\chi^2}{m_U^4}\right) \left(\frac{v}{c}\right)^2$

This naturally reduces the low temperature value of the halo annihilation cross-section relative to the freeze-out value by a factor $(v/c)^2 \sim 10^{-5}$.

There is a price to pay however for the low mass, namely the introduction of a new light gauge boson $m_U \propto m_\chi^{1/2}$, ordinarily comparable in mass to the Z boson if m_χ is at the SUSY breaking scale. A mediating $m_U \sim 0.1 - 1\text{GeV}$ could have observable consequences, for example with regard to the magnetic moment of the muon, and these are being investigated.

One should also eliminate possible astrophysical sources of the 511 keV line. The most promising of these is the population of low mass x-ray binaries, which have a bulge distribution and are known to occasionally have high energy jets and outflows. However there has hitherto been no association of 511 keV emission with any class of discrete sources.

An equally radical suggestion

Three atmospheric Cerenkov radiation telescopes have recently reported the detection of TeV photons from the Galactic Centre. HESS has the most significant detection. The supermassive black hole associated with the SagA* radio source is measured to have a mass of $3 \times 10^6 M_{\odot}$, and x-ray measurements indicate a low accretion rate. Hence a source of γ -rays powered by accretion is unlikely. One could appeal to a high energy cosmic ray accelerator associated with the central black hole. However the low observed accretion rate may (weakly) argue against this. An acceleration power in TeV electrons or EeV protons of 10^{36} - 10^{39} ergs s^{-1} , respectively, is required, where the bolometric luminosity is only 10^{36} ergs s^{-1} (or $10^{-8} L_{\text{Edd}}$).

An annihilation explanation requires WIMPs of mass at least 10-20 TeV. In this case, the observed hard spectrum is naturally explained [35, 36]. However there are difficulties that arise in reconciling the WMAP-constrained value of Ω_{χ} with the cross-section required to account for the HESS luminosity of $10^{35} s^{-1}$ above 200 GeV with half-width of 6 arc-minutes. To arrive at the required relic density for a 20 TeV neutralino mass, one has to fine-tune the particle physics annihilation channels via co-annihilations. The Ω_{χ} constraint prefers a cross-section around 1 pb. The natural value of the cross-section at 20 TeV tends to be lower than 1 pb, because of the unitarity scaling that sets in at large masses, and this results in WIMP overproduction: Ω_{χ} is too high. However, for a typical NFW profile, the inferred cross-section to account for the observed gamma ray flux at 10 TeV is about 10pb, and is even larger for a softer core. In this case, the inferred relic density is too low, only $\Omega_{\chi} \sim 0.03$.

To reconcile these conflicting requirements is not straightforward. The simplest option is to relax the relic density constraint. Suppose that the 20 TeV WIMPs are subdominant. One can now tolerate a larger cross-section. Particle physics fine-tuning is required via co-annihilations, but this is rarely an unsurmountable problem.

Although it appears to be very unnatural that the LSP mass would be any heavier than a few TeV, with a high degree of fine-tuning, co-annihilations can allow for much heavier LSPs. Even in this case, however, it would seem very unlikely that the LSP mass could be any heavier than 20 TeV, at least in the simplest classes of models. The following scenario might then apply. One would have two types of stable particle dark matter, as appropriate to N=2 SUSY [34]. The light particle ($m_{\chi} \sim 10\text{MeV}$) would be the principal dark matter component, and annihilate via e^+e^- to produce the 511 keV flux. The subdominant particle, with mass ~ 10 -20 TeV, would account for the HESS flux.

An alternative is the following. Suppose we settle for the lower cross-section as inferred from the relic WIMP density. Theory certainly has an easier time arriving at this goal. Then we need to boost the annihilation flux at the centre of the galaxy. It is unlikely we can appeal to the usual CDM clumpiness boost factor, because any clumps would be tidally disrupted. It is then appealing to reconsider the possibility of a spike of dark matter around the central SMBH within its zone of influence, a parsec or so. This occurs naturally for adiabatic formation of the SMBH, via the response of the CDM halo, and yields, in principle, an observable gamma ray signal from generic CDM annihilation models [32]. A spike formed in a pregalactic SMBH would survive infall of the SMBH by dynamical friction to the centre of the Milky Way galaxy. This works best if the SMBH forms by baryonic accretion rather than by black hole mergers, although only major mergers are potentially catastrophic for a spike [37]. The survival of a spike seems not unlikely because (a) there is no theoretical understanding of the "final parsec" problem of merging black holes, (b) minihalo mergers in hierarchical galaxy formation yield too few close-in SMBH candidates for successful mergers to prevail in the final system, and (c) forming the very massive SMBHs seen at $z \geq 6$ requires an accretion formation mechanism given the limited time available. The adiabatic spike, which has profile $\rho \propto r^{-\gamma}$ with $\gamma > \frac{3}{2}$, dominates accretion and would yield the HESS point-like source but be unobservable at INTEGRAL/SPI resolution.

THE FUTURE

Baryon dark matter will most likely be mapped out within five years. The intergalactic medium is the major repository where large uncertainty remains. The warm intergalactic medium can be studied via highly ionised oxygen, both in UV absorption and in x-ray emission. This most likely will require dedicated experiments that are being planned.

Of course to distribute the oxygen and other elements into the WIM/ICM requires a greatly improved understanding of galactic outflows. Considerable improvements will be needed in the accuracy and resolution of simulations of galactic outflows. Can the escape rate of gas be of the same order as the star formation rate in massive young galaxies? It will require improvement in the input physics of star formation as well as in the numerical sophistication of the codes before this question can be fully considered.

Advances on the non-baryonic matter front seem equally likely. Of course, here there is a big assumption, that the

elusive dark matter particle is a WIMP. Were it to be a light gravitino or an axion, almost all of the searches would be frustrated. Nevertheless there are more than a score of dedicated searches underway for direct and indirect detection of non-baryonic dark matter. These include searches for annihilation products, including positrons and antimatter (PAMELA, AMS2), high energy neutrinos from the sun (ANTARES, ICECUBE), and gamma rays (GLASST, HESS, VERITAS). It will be necessary with all of these searches to correlate complementary signals and corroborate astrophysical detections with accelerator evidence of existence of the relevant particle. Such evidence may be beyond the reach of the LHC, but a future linear collider should be able to provide the clean signature needed to identify the SUSY LSP, provided that the WIMP mass is below 1TeV. If the WIMP mass is greater, then ACT (gamma ray telescopes) may become the unique hope for detection. Other "smoking guns" include detection of gamma ray line emission and confirmation of annihilation signals associated with nearby dwarf galaxies and with the Galactic Centre, where primordial concentrations of dark matter should exist, by both spectral and spatial resolution.

ACKNOWLEDGMENTS

I thank my colleagues, including R. Bandyopadhyay, C. Boehm, P. Ferreira, D. Hooper, H. Mathis, J. Taylor and H. Zhao, for many discussions of relevant topics. I am also indebted to Professor Piet van der Kruit for hosting me as Blaauw Visiting Professor at the Kapteyn Institute in Groningen, where this review was completed.

REFERENCES

1. Bucher, M. et al., PRL, 93, 081301 (2004)
2. Lambert, D., in Mitchell Symposium on Observational Cosmology, astro-ph/0410418 (2004)
3. Fukugita, M. and Peebles, P., astro-ph/0406095 (2004)
4. Stocke, J., Shull, J. and Penton, S., in STScI Symposium, Planets to Cosmology, astro-ph/0407352 (2004)
5. Birnboim, Y. and Dekel, A., MNRAS, 345,344 (2003)
6. Springel, V. and Hernquist, L., MNRAS, 339, 289 (2003)
7. Zappacosta, L. et al., A&A, 394, 7 (2002)
8. Simcoe, R., Sargent, W. and Rauch, M., ApJ, 606, 92 (2004)
9. Afonso, C. et al., A&A, 404, 145 (2003)
10. Alcock, C. et al., ApJ, 542, 281 (2000)
11. Creze, M. et al., A&A, in press, astro-ph/0403543 (2004)
12. Spagna et al., A., A&A, in press, astro-ph/0410215 (2004)
13. Steidel, C. et al, Astrophys.J. 604, 534 (2004)
14. Adelberger, K. et al, ApJ, 584, 45 (2003)
15. Adelberger, K. et al, ApJ, in press, astro-ph/0410165 (2004)
16. Mac Low, M. and Ferrara, A., ApJ, 513, 142 (1999)
17. Kroupa, P., Science, 295, 82 (2002)
18. Benson, A. et al., MNRAS, 351, 121 (2003)
19. Thomas, D. et al., ApJ, in press, astro-ph/0410209 (2004)
20. Summers, L. et al, MNRAS, 351,1 (2004)
21. Romanowsky, A. et al., Science, 301, 1696 (2003)
22. Hayashi, E. et al., ApJL, submitted, astro-ph/0408132 (2004)
23. Treu, T. and Koopmans, L., ApJ, in press, astro-ph/0401373, (2004.)
24. Silk, J., MNRAS, 343, 249 (2003)
25. Read, J. and Gilmore, G., MNRAS, in press, astro-ph/0409565 (2004)
26. Holley-Bockelmann, K., Weinberg, M. and Neal Katz, K., MNRAS, submitted, astro-ph/0306374 (2004)
27. Sellwood, J., ApJ, 587, 638 (2003)
28. Kravtsov, A., Gnedin, O. and Klypin, A., ApJ, 609, 482 (2004)
29. CDMS Collaboration, PRL, in press, astro-ph/0405033 (2004)
30. Baltz, E. et al, PRD, 65, 065311 (2002)
31. Boehm, C. et al, PRL, 92, 1301 (2004)
32. Bertone, G., Sigl, G. and Silk, J., MNRAS, 337, 98 (2002)
33. Ullio, P. et al., PRD, 66, 123502 (2004)
34. Boehm, C., Fayet, P. and Silk, J., PRD, 69, 101302 (2004)
35. Bergstrom, L. et al., preprint astro-ph/0410359 (2004)
36. Horns, D., Phys. Lett. B, submitted, astro-ph/0408192 (2004)
37. Ullio, P., Zhao, H. and Kamionkowski, M., PRD, 64, 3504 (2001)