Observational Evidence for Self-Interacting Cold Dark Matter

David N. Spergel and Paul J. Steinhardt
Princeton University, Princeton, New Jersey 08544
(Received 20 September 1999)

Cosmological models with cold dark matter composed of weakly interacting particles predict overly dense cores in the centers of galaxies and clusters and an overly large number of halos within the Local Group compared to actual observations. We propose that the conflict can be resolved if the cold dark matter particles are self-interacting with a large scattering cross section but negligible annihilation or dissipation. In this scenario, astronomical observations may enable us to study dark matter properties that are inaccessible in the laboratory.

PACS numbers: 95.35.+d, 98.35.Gi, 98.62.Ai, 98.62.Gq

Flat cosmological models with a mixture of ordinary baryonic matter, cold matter, and cosmological constant (or quintessence) and a nearly scale-invariant, adiabatic spectrum of density fluctuations are consistent with standard inflationary cosmology and provide an excellent fit to current observations on large scales ($\gg$1 Mpc) [1]. However, an array of observations on galactic and subgalactic scales ($\leq$ few Mpc) appears to conflict with the structure formation predicted by analytical calculations and numerical simulations. The predictions are based on the standard view of cold dark matter as consisting of particles with weak self-interactions, as well as weak interactions with ordinary matter.

A generic prediction for weakly self-interacting dark matter, independent of other details of the cosmological model, is that cold dark matter forms triaxial halos with dense cores and significant dense substructures within the halo. Yet, lensing observations of clusters [2] reveal central regions (roughly galactic scale) with nearly spherical low density cores. Dwarf irregular galaxies appear to have low density cores [3–6] with much shallower profiles than predicted in numerical simulations [7,8]. The persistence of bars in high surface brightness galaxies implies that galaxies like our own Milky Way also have low density cores [9]. Observations of the Local Group reveal less than 100 galaxies [10], while numerical simulations [11,12] and analytical theory [13,14] predict that there should be roughly 1000 discrete dark matter halos within the Local Group.

In this paper, we propose that the inconsistencies with the standard picture may be alleviated if the cold dark matter is self-interacting with a large scattering cross section but negligible annihilation or dissipation. The key feature is that the mean free path should be in the range 1 kpc to 1 Mpc at the solar radius, where the dark matter density is about 0.4 GeV/cm^3. The large scattering cross section may be due to strong, short-range interactions, similar to neutron-neutron scattering at low energies, or weak interactions mediated by the exchange of light particles (although not so light as to produce a long-range force). Depending on the interaction and the mean free path, the requisite mass for the dark matter is in the range 1 MeV to 10 GeV. For the purposes of our proposal, only two-body scattering effects are important so either repulsive or attractive interactions are possible. Exchanged particles should be massive enough that they are not radiated by the scattering of dark matter particles in the halo.

We are led to consider self-interactions because ordinary astrophysical processes are unlikely to resolve the problems with standard, weakly interacting dark matter. Consider the dwarf galaxy problem. One might suppose that supernova explosions [15] could cause the galactic core density to be made smoother, but, while the explosions suppress star formation in dwarf galaxies, numerical simulations [16] find that star bursts in dwarfs are very inefficient at removing gas or matter from the core. One might also consider whether the apparent overabundance of halos found in simulations can be explained if the low velocity halos form primarily low surface brightness galaxies [17], which are difficult to find. However, while low brightness galaxy surveys suggest a steeper luminosity function outside of groups [18], even these surveys do not find enough small galaxies to eliminate the discrepancy between theory and observations. If star formation in dwarfs is sufficiently suppressed [19], then they should have been detected as gas clouds in the local group [20] or external systems. HI surveys do not find large numbers of small isolated gas clouds [21]. Even if any of the processes were successful in reducing the number of visible dwarfs, the dense small halos would still persist. When these halos fall onto galactic disks, they will heat the stellar disks and destroy them [12,22,23]. These dense halos will also settle to the centers of the central halo and produce a high density core in galaxies and clusters. Since the baryon fraction in the centers of low surface brightness galaxies is low [17], hydrodynamic processes are not likely to alter their dark matter profiles [3,4].

The success of the cold dark matter model on large scales suggests that a modification of the dark matter properties may be the best approach for resolving the problems on small scales. If the dark matter is not cold, but warm (moderately relativistic), this alleviates some of these discrepancies [24]. However, the remarkably good agreement between standard cold dark matter (CDM) models and the

 0031-9007/00/84(17)/3760(4)$15.00

observed power spectrum of Lyman α absorbers [25] likely rules out warm dark matter candidates.

We propose that a better resolution is dark matter that is cold, nondissipative, but self-interacting. There are stringent constraints on the interactions between dark matter and ordinary matter [26–28] and on long-range forces between dark matter particles [29]. However, as long as the dark matter annihilation cross section is much smaller than the scattering cross section, there are relatively few constraints on short-range dark matter self-interactions. Carlson, Machacek, and Hall [30,31] suggested a self-interacting dark matter model in which the dark matter particle is warm rather than cold. Their model assumed that the dark matter plus ordinary matter sum to the critical density predicted by inflationary cosmology. Their purpose was to reduce the power on 10 Mpc scales in the dark matter mass spectrum as required if the normalization of the spectrum is to agree with fluctuations measured by the COBE satellite. Subsequently, de Laix *et al.* [32] pointed out that the alteration cannot simultaneously fit the IRAS power spectrum and the observed properties of galaxies. Our proposal does not suffer from this problem because we assume a cosmological model with a low matter density and a cosmological constant (or quintessence) which satisfies the COBE constraint without self-interactions. Our proposed self-interactions do not change structure on the 10 Mpc scale but only on the 1 kpc scale. Consequently, our model satisfies the constraints raised by de Laix *et al.*

To be more specific, we suggest that the dark matter particles should have a mean free path between $\sim$1 kpc to 1 Mpc at the solar radius in a typical galaxy (mean density 0.4 GeV/cm^3), for reasons to be explained below. For a particle of mass m_x, this implies an elastic scattering cross section of

$$\sigma_{XX} = 8.1 \times 10^{-25}\ \mathrm{cm}^2 \left(\frac{m_x}{\mathrm{GeV}}\right)\left(\frac{\lambda}{1\ \mathrm{Mpc}}\right)^{-1}, \qquad (1)$$

intriguingly similar to that of an ordinary hadron. (In this paper, we consider the case of dark matter particles scattering only from themselves but, in a forthcoming paper, we consider the possibility that dark matter is a stable, neutral hadron.) If the dark matter particles scatter through strong interactions similar to low-energy neutron-neutron scattering, then the cross section is $\sigma = 4\pi a^2$, where a is the scattering length. For neutrons, the scattering length is more than 100 times its Compton wavelength. Using the estimate $a \approx 100 f m_x^{-1}$, we obtain

$$m_x = 4\left(\frac{\lambda}{1\ \mathrm{Mpc}}\right)^{1/3} f^{2/3}\ \mathrm{GeV}. \qquad (2)$$

Alternatively, the self-interaction may be weak but longer range, as in the case of the exchange of a light intermediate vector boson of mass m_y, in which case the cross section is $\sigma \approx \alpha_y m_x^2/m_y^4$. The mass of the vector boson must be large enough that there is no dissipation when dark matter particles scatter; this requires that $m_y >$ 450 eV $(m_x/1\ \mathrm{GeV})\,[v/(200\ \mathrm{km/s})]^2$, where v is the typical velocity of dark matter particles in the halo. This mass scale for m_y corresponds to a force that is short range compared to the dark matter interparticle spacing (about 1 cm in the halo). Hence, we need consider only two-body interactions in our analysis. If $m_y = g m_x$ and $\alpha_y = O(1)$, then the maximum dark matter mass is

$$m_x < 80\left(\frac{\lambda}{1\ \mathrm{Mpc}}\right)^{1/3} g^{-4/3}\ \mathrm{MeV}. \qquad (3)$$

Beyond what is expressed in the relations above, there is no significant constraint on how light the dark matter particles can be.

The strong self-interaction might occur if the dark matter consists of particles with a conserved global charge (such as a hidden baryon number) interacting through a hidden gauge group (e.g., hidden color). If the gauge group is unbroken, then the particles experience strong interactions which can be nondissipative but the particle number is conserved. M-theory and superstrings, for example, suggest the possibility that dark matter fields reside on domain walls with gauge fields separated from ordinary matter by an extra (small) dimension [33,34]. Similar scenarios can be constructed in purely four-dimensional supergravity models. Note that, if the sum of the hidden baryon number and the ordinary sector baryon number is zero, then $\Omega_x = (m_x/m_{\mathrm{proton}})\Omega_b \simeq 0.19\,(m_x/41\ \mathrm{GeV})$ (using $\Omega_b h^2 = 0.02$ and $h = 0.65$). The particles we suggest include light versions of Q-balls [35].

How does the mean free path of the dark matter particles affect astrophysics? Since interactions alter only the evolution of cold dark matter when the density inhomogeneities are large, the cosmic microwave background and large-scale power spectrum measurements are not sensitive to the self-interactions. So long as the dark matter is cold ($T_x/m_x < \Phi$, where Φ is the depth of the gravitational potential), the dark matter will collapse to form a bound halo regardless of its collisional properties. If the dark matter mean path were much longer than $\sim$1 Mpc, the typical dark matter particle would not experience any interactions as it moves through a halo. In this regime, the usual, triaxial cold dark matter halo with dense core forms through gravitational collapse. On the other hand, if the dark matter mean free path is much smaller than 1 kpc, then the dark matter behaves as a collisional gas and this alters the halo evolution significantly. The dark matter will shock: this will heat up the low entropy material that would usually collapse to form a core and produce a shallower density profile. Since collisions tend to make the dark matter velocity distribution isotropic, the halo cannot be triaxial and will be elliptical only if flattened by significant rotation. Since dark halos form with little angular momentum, if the dark matter is not dissipative, then all halos will be nearly spherical. X-ray observations of clusters [36] reveal that most halos are moderately ellipsoidal. This implies that the collision time scale for dark matter near the half-mass radius of clusters must be longer than the Hubble time: one of the strongest constraints on this model. Studies suggest that polar ring galaxies [37] are only mildly triaxial and oblate with the equatorial plane of the dark halo nearly coinciding with that of the stellar body. If the dark matter has an isotropic distribution function and the baryons form a disk, then the dark matter will form a slightly flattened halo.

In our scenario, we consider a mean free path in the intermediate regime, larger than 1 kpc, but smaller than $\sim$1 Mpc. Particles in this range have $1-10^3$ interactions per Hubble time in the local neighborhood, which is overdense by 10^6 relative to the mean density of the universe. At the virial radius of a typical galactic halo, which is overdense by $\sim$200 relative to the mean density of the universe, the typical particle has less than 1 collision per Hubble time. Thus, near the virial radius, halos can have anisotropic velocity ellipsoids and will be triaxial. However, in the inner halo of galaxies, dark matter is collisional. In-falling dark matter is scattered before reaching the center of the galaxy so that the orbit distribution is isotropic rather than radial. These collisions increase the entropy of the dark matter phase space distribution and lead to a dark matter halo profile with a shallower density profile. The characteristic scale for the core would correspond to an "optical depth" of 1, the "photosphere" of the dark matter.

When a dwarf halo with a low velocity dispersion falls into a larger, high velocity dispersion halo, the high velocity particles will scatter off the low velocity particles. After the collision, neither particle is likely to be bound to the dwarf. As dark matter is slowly removed, the dwarf halo expands, this also makes the halo more vulnerable to tidal stripping and shock heating. This process will slowly evaporate all substructure in the larger halo, particularly, in the centers of galaxies, groups, and clusters. However, near the half-mass radius, the collision time is longer than the Hubble time, so substructure will be destroyed only in the inner portions of halos. This dark matter evaporation will protect galaxy disks from dynamical heating by collisions with dwarfs in the halo. Dwarfs with high central densities evaporate more slowly as particles at optical depth greater than $\sim$1 are shielded by other particles from collisions. Intriguingly, all of our galaxy's dwarf companions have very high phase space densities [10]. As the central density and particle velocities increase, not all collisions will lead to particles being deflected out of the dwarf halo. A small fraction of the particles will be rescattered within the spheroidal core of the dwarf and a fraction of the momentum absorbed. [The phase space for capture is suppressed by $(\sigma/v)^3$, where σ is the velocity dispersion in the dwarf core and v is the typical particle velocity in the halo.] This could produce a ram pressure drag that can slow the dwarf halo and cause it to spiral into the cores of larger halos. Numerical studies of this process are required to determine whether the observed fraction of denser dwarfs are likely to survive after a Hubble time.

The halos of large (e.g., L_*) galaxies moving through groups and clusters are less prone to destruction. When a cluster dark matter particle strikes a galaxy dark matter particle, the probability that the recoiling particles will escape from the galaxy is significantly less than unity. For cross sections near the smaller end of our suggested range, most collisions take place within the half-mass radius, where the escape velocity from a galaxy halo is comparable to the characteristic recoil velocity. Thus, many collisions will not lead to recoil energies large enough to escape. At the smaller end of our cross-section range, the probability of a typical particle experiencing a collision during a Hubble time approaches unity only for galaxies that fall deep into the cluster core. Towards the larger end of our suggested range, the galaxies are opaque to dark matter and the collision products are also likely to experience multiple collisions within a massive galaxy halo. Because of these multiple collisions, the collision products are unlikely to escape from the galaxy. This effect is more important in massive halos as the optical depth through a galaxy scales as the one-third power of its mass. Thus, for large galaxies, the background cluster or group will primarily heat the entire dark halo rather than evaporate dark matter from it.

The presence of collisions will lead to energy transport within the dark matter halo, which eventually leads to core collapse [38]. We can obtain an estimate of the core collapse time from Quinlan's [39] Fokker-Planck simulations of the evolution of an isolated cluster of interacting particles with a central density profile of r^{-1} and an outer profile of r^{-3}. These models have a temperature inversion in the core and undergo two stages of core collapse. During the first stage, the inner region expands as heat is transported inwards. After 0.1 half-mass relaxation times, the inner 1% of the mass has moved out in radius by a factor of 2. After roughly 3 half-mass relaxation times, the whole system collapses as heat is transported from the virial radius outwards. We suspect that this second stage is delayed in our cosmological context as the in-fall of new material is constantly adding heat near the dark halo photosphere. If we model our galaxy as starting with a density profile corresponding to the Navarro-Frenk-White fit to CDM models [7] with $V_{200} = 225\ \mathrm{km\,s^{-1}}$, $\Omega_0 = 0.3$, $H_0 = 65\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$ and dimensionless concentration parameter $c = 8$, then its half-mass radius is 220 kpc. Since the density at the half-mass radius is 300 times smaller than the local dark matter density, the particle mean free path at that radius is between 0.3–300 Mpc, and, at that radius, the particle is in the weakly interacting regime. Thus, for our galaxy, the core collapse time (roughly 3 half-mass relaxation times) is between 4.5–6000 Gyr, or perhaps significantly longer if the collapse stage is delayed by the in-fall of new material. Hence, for most of our range of parameters, the collapse time for our galaxy exceeds the lifetime of the universe, yet there is sufficient number of interactions to lower the dark matter density in the inner 5 kpc of our galaxy. As the particle mean free path approaches the lower bound (0.3 Mpc) or the upper bound (300 Mpc) at the half-mass radius, our estimates suggest that one or the other condition is not satisfied, but more accurate methods are needed to determine the precise range.

To summarize, our estimated range of σ/m for the dark matter is between 0.45–450 cm^2/g or, equivalently, $8 \times 10^{-(25-22)}$ cm^2/GeV. Numerical calculations are essential for checking our approximations and refining our estimates. Even without numerical simulations, we can already make a number of predictions for the properties of galaxies in a self-interacting dark matter cosmology: (1) The centers of halos are spherical; (2) dark matter halos will have cores; (3) there are few dwarf galaxies in groups but dwarfs persist in lower density environments; and (4) the halos of dwarf galaxies and galaxy halos in clusters will have radii smaller than the gravitational tidal radius (due to collisional stripping). Intriguingly, current observations appear to be consistent with all of these predictions.

We thank R. Dave, J. Dalcanton, G. Dvali, J. Goodman, E. Kolb, J. March-Russell, J. Miralda-Escude, J. Ostriker, J. Peebles, J. Silk, S. Tremaine, M. Turner, and N. Turok for discussions. We are grateful to the West Anglia Great Northern Railway whose train delay provided the opportunity for initial discussions. This work was initiated at the Newton Institute for Mathematical Sciences. D. N. S. acknowledges the NASA Theory program and the NASA MAP Satellite program for support, and P. J. S. was supported by the U.S. Department of Energy Grant No. DE-FG02-91ER40671.

Note added.—Burkert [40] has constructed an N-body code that simulates self-interactions and obtains results consistent with these estimates based on the Fokker-Planck approximation.

[1] N. Bahcall, J. P. Ostriker, S. Perlmutter, and P. J. Steinhardt, Science **284**, 1481 (1999).
[2] J. A. Tyson, G. P. Kochanski, and I. P. Dell'antonio, Astrophys. J. **498**, L107 (1998).
[3] B. Moore, Nature (London) **370**, 629 (1994).
[4] R. A. Flores and J. A. Primack, Astrophys. J. **427**, L1 (1994).
[5] W. J. G. De Blok and S. S. McGaugh, Mon. Not. R. Astron. Soc. **290**, 533 (1997).
[6] J. Dalcanton and R. Bernstein, in *Galaxy Dynamics: from the Early Universe to the Present,* edited by F. Combes, G. A. Mamon, and V. Charmandaris, ASP Conference Series (Astronomical Society of the Pacific, San Francisco, CA, 1999), p. 23.
[7] J. F. Navarro, C. S. Frenk, and S. D. M. White, Astrophys. J. **462**, 563 (1996).
[8] B. Moore, F. Governato, T. Quinn, J. Stadel, and G. Lake, Astrophys. J. **499**, L5 (1998).
[9] V. P. Debattista and J. A. Sellwood, Astrophys. J. **493**, L5 (1998).
[10] M. Mateo, Annu. Rev. Astron. Astrophys. **36**, 435 (1998).
[11] A. A. Klypin, A. V. Kravtsov, O. Valenzuela, and F. Prada, Astrophys. J. **522**, 82 (1999).
[12] B. Moore, S. Ghigna, F. Governato, G. Lake, T. Quinn, J. Stadel, and P. Tozzi, Astrophys. J. **524**, L19 (1999).
[13] W. H. Press and P. Schechter, Astrophys. J. **187**, 425 (1974).
[14] G. Kauffmann, S. D. M. White, and B. Guiderdoni, Mon. Not. R. Astron. Soc. **264**, 201 (1993).
[15] A. Dekel and J. Silk, Astrophys. J. **303**, 39 (1986).
[16] M. M. MacLow and A. Ferrara, Astrophys. J. **513**, 142 (1999).
[17] J. Dalcanton, D. N. Spergel, and F. J. Summers, Astrophys. J. **482**, 659 (1997).
[18] J. Dalcanton, D. N. Spergel, J. E. Gunn, M. Schmidt, and D. P. Schnedider, Astron. J. **114**, 635 (1997).
[19] J. V. Kepner, A. Babul, and D. N. Spergel, Astrophys. J. **487**, 61 (1997).
[20] L. Blitz, D. N. Spergel, P. J. Teuben, D. Hartmann, and W. B. Burton, Astrophys. J. **514**, 818 (1999).
[21] M. Zwann, F. Briggs, D. Sprayberry, and E. Sorar, Astrophys. J. **490**, 173 (1997).
[22] G. Toth and J. P. Ostriker, Astrophys. J. **389**, 5 (1992).
[23] M. Weinberg, Mon. Not. R. Astron. Soc. **299**, 499 (1998).
[24] R. Schaeffer and J. Silk, Astrophys. J. **332**, 1 (1988).
[25] R. A. C. Croft, D. H. Weinberg, M. Pettini, L. Hernquist, and N. Katz, Astrophys. J. **520**, 1 (1999).
[26] R. Bernabei *et al.,* Phys. Rev. Lett. **83**, 4918 (1999).
[27] C. W. Stubbs, Phys. Rev. Lett. **70**, 119 (1993).
[28] G. Starkman, A. Gould, R. Esmailzadeh, and S. Dimopoulos, Phys. Rev. D **41**, 3594 (1990).
[29] B.-A. Gradwohl and J. A. Frieman, Astrophys. J. **398**, 407 (1992).
[30] E. D. Carlson, M. E. Machacek, and L. J. Hall, Astrophys. J. **398**, 43 (1992).
[31] M. E. Machacek, Astrophys. J. **431**, 41 (1994).
[32] A. A. de Laix, R. J. Scherrer, and R. K. Schaeffer, Astrophys. J. **452**, 495 (1995).
[33] G. Dvali, Phys. Lett. B **459**, 489 (1999).
[34] G. Dvali and G. Gabadadze, Phys. Lett. B **460**, 47 (1999).
[35] A. Kusenko and M. Shaposhnikov, Phys. Lett. B **418**, 46 (1998).
[36] J. J. Mohr, A. E. Evrard, D. G. Fabricant, and M. J. Geller, Astrophys. J. **447**, 8 (1995).
[37] P. D. Sackett, in *Galaxy Dynamics*, edited by D. Merritt, J. A. Sellwood, and M. Valluri (Astronomical Society of the Pacific, San Francisco, CA, 1999).
[38] J. Goodman, Astrophys. J. **313**, 576 (1987).
[39] G. Quinlan, New Astron. **1**, 255 (1996).
[40] A. Burkert, astro-ph/0002409.