Cosmological Simulations of Multicomponent Cold Dark Matter

Mikhail V. Medvedev
Institute for Theory and Computation, Harvard University, 60 Garden Street, Cambridge, Massachusetts 02138, USA; Department of Physics and Astronomy, University of Kansas, Lawrence, Kansas 66045, USA; and ITP, NRC “Kurchatov Institute”, Moscow 123182, Russia
(Received 6 May 2013; revised manuscript received 13 May 2014; published 15 August 2014)

The nature of dark matter is unknown. A number of dark matter candidates are quantum flavor-mixed particles but this property has never been accounted for in cosmology. Here we explore this possibility from the first principles via extensive N-body cosmological simulations and demonstrate that the two-component dark matter model agrees with observational data at all scales. Substantial reduction of substructure and flattening of density profiles in the centers of dark matter halos found in simulations can simultaneously resolve several outstanding puzzles of modern cosmology. The model shares the “why now?” fine-tuning caveat pertinent to all self-interacting models. Predictions for direct and indirect detection dark matter experiments are made.

Introduction.—Dark matter (DM) constitutes about 80% of matter and 25% of the total energy density in the Universe but its nature remains completely unknown. The existence of DM requires revision of the present day physics. Most likely, DM is a hypothetical particle or particles beyond the standard model [1].

The current heuristic paradigm of the cold dark matter with a cosmological constant (ΛCDM) is remarkably successful at reproducing the large-scale structure of the Universe but appears to disagree with observations at small scales. First, simulations predict the overabundance of small mass (dwarf) halos as compared to the much lower number of the observed satellite galaxies in the Local Group [2–5] and in the field as inferred from the ALFALFA survey [6]. This problem was termed the “substructure problem” or “missing satellite problem.” Second, the cuspy \( \rho \propto r^{-1} \) DM density profiles found in ΛCDM simulations [7] disagree with the rotation curves of dwarf and low surface brightness galaxies, which indicate flattened or cored density profiles [8–14]. Observations of galaxy clusters also indicate the presence of cores [15]. Moreover, the largest ΛCDM subhalos in the Local Group-type environments are too dense in their centers to host any of the dwarf spheroidal galaxies around the Milky Way and Andromeda galaxies, and in the field [16–18]. These two, perhaps related, problems are known as the “core-cusp problem” and “too-big-to-fail problem,” respectively. Numerous attempts to reconcile the ΛCDM model with observations using baryonic processes made so far (modified star formation, tidal gas stripping, supernova feedback) are inconclusive [4,19–24]. This is because the latter problems require strong feedback and, hence, larger star formation, whereas the substructure problem requires just the opposite—the suppressed star formation. Contrary to early expectations, a mild DM paradigm shift to adopt warm dark matter (WDM) [25–27] also fails to resolve all these problems altogether [28–30] due to the similar constraints on the DM particle mass (but, see Refs. [31–39] for hybrid models).

The inability of conventional physics to resolve the aforementioned problems within the collisionless CDM paradigm can indicate that DM may exhibit nongravitational properties as well. The most natural alternative is to admit a large interaction cross section of DM with itself [40,41] but not with normal matter. Contrary to the early claims [42–44], such self-interacting dark matter (SIDM) was successful to explain the origin of cores without violating any constraints on the velocity-dependent cross section \( \sigma(v) \) [45–52]; however, it completely fails to solve the substructure problem [51]. Interestingly, SIDM can naturally explain the presence of supermassive black holes in red bulgeless galaxies [53] and their very early formation [54,55] via gravitational collapse of the central (collisional) parts of DM halos [56,57]—the process that is absent in the “vanilla CDM” paradigm. At last, the existence of a narrow plane of the Andromeda dwarf satellites [58], which has no explanation within collisionless CDM, can potentially be addressed in SIDM, because collisionality induces viscous drag on subhalos (whether it is enough is unknown).

However, an important possibility that some DM candidates are quantum-flavor-mixed particles, e.g., a neutralino, an axion, a sterile neutrino, has not been considered so far. In this Letter we demonstrate from first principles via N-body cosmological simulations that even the simplest model with two-component quantum-mixed DM with small mass degeneracy agrees with observational data at both large and small scales, thus possibly settling the above problems altogether. Moreover, it also agrees with the observational constraints on \( \sigma(v) \) set by SIDM models [47–52]. At last, the model makes predictions for and is testable with direct and indirect detection DM experiments.
Model.—First, we postulate that the dark matter particles are flavor-mixed. Generally, a mixed particle of flavor $\alpha$ is a superposition of several mass eigenstates $|f_\alpha\rangle = a_1|m_1\rangle + a_2|m_2\rangle + \cdots$, where $|f\rangle$ and $|m\rangle$ denote wave functions being flavor and mass eigenstates, and $a_1, a_2, \ldots$ are the elements of a unitary matrix. Here we consider the simplest DM model with two flavors and two mass eigenstates only [59,60], i.e., the two-component DM (2cDM) model. The masses of the mass eigenstates are $m_h$ and $m_l < m_h$, i.e., “heavy” and “light.” Generally, $|m\rangle$’s have different velocities [61,62] and propagate along different geodesics. Hence, they can be spatially separated by gravity during structure formation: the eigenstates with smaller speeds become trapped in a growing halo earlier than the faster ones. The DM halos are, thus, self-gravitating ensembles of non-overlapping heavy and light eigenstates.

Second, we postulate that DM particles can interact with each other nongravitationally with some velocity-dependent cross section, $\sigma(v)$, which is consistent with the existing SIDM constraints. It is customary in cosmology to parameterize it as $\sigma(v) = \sigma(v/v_0)^{-a}$, where $\sigma$ and $a$ are parameters and $v_0$ is a normalization constant. Previous studies and observational data allow for $a \geq 1$ [45,47,57], so $a = 1$ is used in the simulations reported here. This $1/v$ dependence is also natural for mass-eigenstate conversions [60]. Observations constrain the ratio $\sigma/m$, where $m$ is the DM particle mass, to be in the range $0.1 \lesssim \sigma/m \lesssim O(1)$ cm$^2$/g for the assumed normalization $v_0 = 100$ km/s [45,47,57].

The dynamics of nonrelativistic mixed particles is interesting and unusual. For instance, a collision of the mass eigenstate $|m_h\rangle$ with another particle can either be the elastic scattering $|m_h\rangle \to |m_h\rangle$ or the mass eigenstate conversion $|m_h\rangle \to |m_l\rangle$ (or simply the $m$-conversion $h \to l$), because of the nondiagonal elements of the flavor interaction matrix in the mass basis [59]. Let’s consider $h \to l$ off a static, $\delta$-localized flavor potential with $h$ being at rest, for simplicity. The energy conservation $m_h c^2 = m_l c^2 + m_1 c^2/2$ implies that $|m_l\rangle$ gets a velocity $v = c(2(m_h - m_l)/m_1)^{1/2}$ in a random direction. Our simulations indicate that the mass-degenerate case, $m_h = m_l = m$ and $\Delta m \equiv (m_h - m_l) \ll m$, fits observations the best. Thus we define the kick velocity parameter $v_k = c\sqrt{2\Delta m/m}$, which can be used in place of the $\Delta m/m$ parameter. If $v_k$ exceeds the escape velocity from a DM halo, $v_{esc}$, a part of the particle’s wave function—the resultant $l$ eigenstate—will escape, thus decreasing the particle’s probability to be in that halo and, hence, the halo mass. This irreversible escape of the flavor-mixed particles was called the “quantum evaporation” [59,60]. The evaporation ceases if $v_{esc} \gg v_k$.

Self-interactions of two mixed DM particles are more complex and involve all 16 combinations of mass-eigenstate pairs in the input and output channels; see [60] for the full quantum mechanical analysis. The $m$-conversions in which one or two heavy eigenstates are converted, $hh \to hl$, $hh \to ll$, and $hl \to ll$, can lead to the quantum evaporation. Because of the energy conservation, the kinetic energy increases by $\Delta m c^2$ in processes like $hh \to hl$ and twice as much in $hh \to ll$. The reverse processes $hl \to hh$, $ll \to hl$, and $ll \to hh$ can also occur if kinematically allowed, i.e., if the initial kinetic energy is large enough to produce a heavy eigenstate. Finally, the elastic scattering processes $ll \to ll$, $hl \to hl$, $hl \to lh$, and $hh \to hh$ can occur as well.

Complete evaporation of a halo is possible depending on the $m$-conversion cross sections, initial DM composition [60] and mixing angle $\theta$. For simulations, we chose one such case: the maximal mixing with equal initial numbers of $h$ and $l$ eigenstates. In general, the scattering and conversion cross sections depend on the flavor interaction strengths and $\theta$. The effect of $m$-conversions is the strongest for the maximal mixing, and 2cDM reduces to SIDM for $\theta \ll 1$; see [60] for details.

Implementation.—The physics of mixed-particle interactions was implemented in the publicly available cosmological TreePM/SPH code GADGET [63]. We simulated two types of DM particles representing $h$ and $l$ mass eigenstates; the total numbers of each type can change due to $m$-conversions. In the code, DM particles are interacting SPH particles but without hydro-force acceleration. To model particles’ binary interactions, we use the Monte Carlo technique with the “binary collision approximation” [42,44], which is reliable for weakly collisional systems. The algorithm is as follows. For each randomly chosen projectile particle $s_i$, a nearest neighbor is found; this is the target particle $t_j$. For each input channel $s_i t_j$ there are four output channels $s_j t_o$, namely: $hh, hl, lh,$ and $ll$. The probabilities of the four processes $s_i t_j \to s_j t_o$.

$$P_{s_i t_j \to s_j t_o} = \rho_{s_i} \sigma_{s_i t_j \to s_j t_o} |v_t - v_s| \Delta t \Theta(E_{s_i t_j})$$

are computed, where $\sigma_{s_i t_j \to s_j t_o} = \sigma(v)$ is the cross section, $v_t - v_s$ is the relative velocity of particles in the pair, $\rho_{s_i}$ is the density of target species computed by the SPH density routine, $\Delta t$ is the iteration time step, and $\Theta(E_{s_i t_j})$ is the Heaviside function which ensures that the process is kinematically allowed (i.e., negative final kinetic energy, $E_{s_j t_o} < 0$, means the process cannot occur). Whether an interaction occurs and through which channel is determined by random drawing in accordance with the computed probabilities. Kinematics of all the interactions is computed in the center of mass frame. If a scattering occurs, the particles are given random antiparallel velocities with magnitudes set by the energy-momentum conservation. If an $m$-conversion occurs, then (i) the type of one or both particles is changed, (ii) the magnitudes of the final velocities are computed with $\Delta m c^2$ given or taken, depending on the type of conversion, and (iii) these velocities are assigned to the particles in antiparallel directions. If no interaction occurs, the particle velocities and types remain intact. After this, the pair is marked inactive until the next time step. This process is repeated for all active particles at each time step.

Our 2cDM runs have $2 \times 400^3 = 128 \times 10^6$ SPH-DM particles (in 2cDM, the initial numbers of $h$ and $l$ particles...
are equal) in the box of $50 h^{-1}$ Mpc (comoving) with the force resolution scale of $3.5 h^{-1}$ kpc, and the reference \( \Lambda \)CDM run has $2 \times 640^3 \approx 524 \times 10^6$ particles and the force resolution of $2.2 h^{-1}$ kpc. Our box size was optimized to be large enough to be a representative sample the Universe volume, yet it provides reasonable resolution at small scales. All the runs are DM-only simulations with the standard cosmological parameters $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $\Omega_b = 0$ and $h = 0.7$. Initial conditions were generated using N-GenIC code with $\sigma_8 = 0.9$ and the initial redshift $z = 50$. The AHF code [64] was used to construct the halo mass function and maximum circular velocity function (MCVF), analyze halo density profiles, etc. Simulations of SIDM were done too. They fully confirm earlier studies, e.g., the inability to resolve the substructure problem; hence, these results are not reported here. Numerous runs were performed to explore the range of the 2cDM model parameters $\Delta m/m$ and $\sigma/m$, to compare with the reference CDM and SIDM models and to check for numerical convergence. Here we report the most important ones.

Results.—Simulations with the large mass difference (not presented here) grossly disagree with the observational data, so this case is not considered further. Hence, because of $\Delta m/m \ll 1$, the mass segregation of heavier species toward the halo center is negligible.

The DM maps in a zoomed-in region of 5 Mpc across at $z = 0$ for the 2cDM and \( \Lambda \)CDM models are presented in Fig. 1. One sees fewer subhalos in the 2cDM case. The parameters are $\Delta m/m = 10^{-8}$, which corresponds to $v_k = 50$ km/s, and $\sigma/m = 0.75$ cm$^2$/g at $v_0 \sim v_k$, which is fully consistent with observational constraints on the SIDM cross section [45–51,57]. For these values, the 2cDM MCVF matches the Local Group data the best, as shown in Fig. 2. This figure shows the number of halos with the maximum circular velocity above a certain value, $N(> V_{c,\text{max}})$ vs $V_{c,\text{max}}$, for 2cDM and \( \Lambda \)CDM; the data points are from [3,4]. The amount of substructure is volume dependent, so we appropriately rescaled the data points to reproduce the results of Refs. [3,4] using the MCVF from our \( \Lambda \)CDM simulation; the procedure is legitimate for a scale-free ergodic distribution of DM structure. However, no data rescaling is done for the substructure MCVFs of two individual Milky Way-like halos shown in the inset. In both cases, the agreement with 2cDM is much better than with \( \Lambda \)CDM.

The simulations show that $v_k$ uniquely determines the position of the break in the MCVF, $\nu_k^{\text{break}} = v_k$, whereas $\sigma/m$ determines the slope below the break. By comparing simulations with observational data, we determined $v_k$ (and, consequently, $\Delta m/m$) to be around $\sim 50–70$ km/s. Interestingly, a similar value of a characteristic velocity $\lesssim 100$ km/s was found in another independent analysis of survey data [5]. The “best” fit cross section is $\sigma/m \sim 0.75$ cm$^2$/g at $v_k$, but values a factor of 2 smaller or larger are acceptable too. The halo mass function exhibits the even sharper break at $M = 10^{10} M_\odot$. Thus, the overall suppression of the abundance of dwarf halos resolves the substructure problem.

Figure 3 shows 120 well-resolved halo density profiles for \( \Lambda \)CDM and 2cDM. The profiles are trustworthy everywhere because their inner parts were truncated according to the numerical binary collision criterion [65]. The \( \Lambda \)CDM profiles agree with the NFW profile. In contrast, the 2cDM inner profiles are shallower and less centrally concentrated. Although the stacked profiles are less noisy, the inset shows that individual profiles exhibit an order-of-magnitude variance in their central densities due to different formation histories. The lack of a common central density scale agrees with observations [28], which were originally argued against SIDM.

The softening of cusps is also seen from Fig. 4. Here, the effective power-law index is obtained by fitting the individual profiles with the function $\rho = \rho_0 r^{\alpha}(1 + r/r_c)^\beta$ and then evaluating $\alpha$ at $r = 7$ kpc/h. The distribution of the slopes ranges within $\alpha = -0.8... 1$ for CDM indicating a cusp and within $\alpha = -0.2... -0.6$ for 2cDM, which thus explains the core-cusp and, likely, too-big-to-fail problems. Importantly, the density profiles and core sizes of massive halos are mostly sensitive to $\sigma/m$, whereas $v_k$ plays little role, if any. The profiles of the halos with $M \lesssim 10^{10} M_\odot$...
may depend on \( v_k \), which should be explored with dedicated high-resolution simulations. We stress that SIDM runs show softened cusps but no substructure suppression—the MCVF is a scale-invariant power law—due to the lack of a physical parameter, such as \( v_k \) in 2CDM, which can set the break scale.

**Implications.**—(i) Cosmology with at least two flavor-mixed mass-degenerate \( \Delta m/\Delta m \sim 10^{-8} \) species can naturally resolve cosmological problems at small scales without invoking new or exotic physics. In contrast, single-species and/or nonmixed candidates, and nondegenerate multi-component models are disfavored.

(ii) 2CDM agrees with observations within a range of the velocity-dependent \( \sigma/m \) allowed for SIDM [45–51,57]. The constraints are tight: if \( \sigma \) is too small, then it is cosmologically uninteresting, if it is too large, then the cusps will be enhanced due to the gravothermal collapse of halos. This fine-tuning, rephrased as the “why now?” question, is a caveat of 2CDM. However, SIDM and dark energy or cosmological constant face the same problem.

(iii) Our model does not change the linear power spectrum, unlike WDM; all changes occur in the nonlinear stage. The quantum evaporation proceeds slowly over the Hubble time. We can speculate that the gas metal enriched by the stars in dwarf spheroidals should gradually become unbound from the weakening gravitational potential of the halos and enrich the intergalactic medium with metals, resembling the effect of supernovae or winds. Since not all small halos are evaporated by \( z = 0 \), the residual substructure can be responsible for the flux anomalies in gravitational lensing observations.

(iv) Our simulations can formally describe any multi-component DM where transformations of species are allowed. However, these models face a severe problem: Why have the heavy (e.g., “excited,” etc.) particles survived in the early Universe, but convert to lighter (or “ground state”) species now, when the density is much smaller? The flavor-mixed 2CDM model does not have this problem, because the \( m \)-conversion cross section in the flat spacetime is suppressed by \( (\Delta m/m)^4 \sim 10^{-32} \) over its current value [60] and becomes large only during the structure formation, when mass eigenstates separate.

(v) The 2CDM theory is testable with direct detection experiments. Indeed, DM is a collection of \( h \) and \( l \) eigenstates, which can convert into one another in interactions with normal matter in a detector. These conversions should result in the energy “mismatch” of \( \sim \Delta m c^2 \); i.e., the events will look like inelastic collisions: “exothermic” and “endothermic.” Particularly, the down-conversions \( h \rightarrow l \), which are always kinematically allowed, can look like exothermic interactions. In contrast, the \( l \rightarrow h \) up-conversions can occur only if the kinetic energy exceeds a threshold. Hence, the \( l \rightarrow h \) rate can exhibit a stronger annual modulation. Next, we can also speculate that if DAMA and CoGeNT anomalies are due to inelastic effects with \( \Delta m \sim keV \), then the DM mass is \( m \sim 10^5 \Delta m \sim 10^2 GeV \), which is close to that inferred from the GeV excess data [66,67]. Finally, we also suggest that the use of different targets (e.g., Ne, Ar) in the experiments may strongly affect the recoil signal strength because of possible different flavor couplings to the DM species, whose flavor composition is unknown.

(vi) 2CDM can be tested in indirect detection experiments. For instance, the direct DM annihilation into two photons results in a line triplet corresponding to the annihilations in the \( h + h \), \( h + l \), and \( l + l \) channels. Thus, the DM annihilation line can be a triplet at \( E = mc^2 \) split by \( \Delta E = \frac{1}{2} \Delta m c^2 \) and with different line strengths.

The author is grateful to Lars Hernquist, Avi Loeb, Ramesh Narayan, Lyman Page, Sergei Shandarin, and Mark Vogelsberger for discussions and suggestions. This work was supported in part by the Institute for Theory and Computation at Harvard University and by DOE Grant No. DE-FG02-07ER54940, NSF Grant No. AST-1209665 and XSEDE Grants No. AST110024 and No. AST110056. The simulations utilized XSEDE high-performance computing systems Trestles (SDSC) and Ranger (TACC).