

# Galaxy formation and dark matter: small scale problems and quantum effects on astrophysical scales

**Amr A. El-Zant**

Centre for Theoretical Physics, the British University in Egypt, Sherouk City, Cairo 11837, Egypt

E-mail: [amr.elzant@bue.edu.eg](mailto:amr.elzant@bue.edu.eg)

**Abstract.** Although non-baryonic dark matter seems essential in the context of the currently favoured cosmological model, the standard dark matter scenario is facing problems: experimental searches have failed to find the relevant particles, closing the mass-corsssection window of the 'WIMP miracle', and the model suffers from problems on (sub) galactic scales. The cold dark matter (CDM) invoked may turn out to be too cold and needs to be heated; so that its solution to the dearth of visible matter in the outer parts of galaxies is not accompanied by the problem of an excess of matter in their centres (along with other possibly related problems, such as the numerical excess of predicted satellites). After a heuristic introduction to some aspects of the rationales that lead to the CDM paradigm, I discuss the properties of self gravitating CDM structures (haloes) and the proposed reasons for their apparently 'universal profiles' (including new simulations attempting to explain aspects of their advent), the galactic-scale problems associated with them, and proposed solutions, focussing on baryonic solutions and the recently topical ultra-light axion particles as replacement for the standard weakly interacting massive particles (WIMPs). It is hoped that at least parts of this review would be helpful to a general physics audience interested in the problem of dark matter in an astrophysical context.

## 1. Introduction

This review is meant to give a taste of the field of galaxy formation in the context of the standard cold dark matter (CDM) scenario, sketch its successes and describe problems it faces on galactic scales and some of the proposed solutions. Extensive prior knowledge of the subjects of galaxy formation and physical cosmology is not assumed for a reasonable reading of the text. It is hoped that it may help introduce readers with a general physics background to current problems in the field.

The field of galaxy formation and evolution has evolved enormously in the past quarter century. Much of this progress came as a result of its merging with the at least equally rapidly developing field of physical cosmology. Cosmology, once relegated to an esoteric corner of scientific inquiry, largely concerned with examining the solutions of Einstein's field equations for proposed universes proposed it a priori, had been coming of age in the past decades, as the much touted era of 'precision cosmology' dawned.

This progress was primarily observationally driven. Of particular influence were increasingly precise observations of the cosmic microwave background (CMB) and the large scale distribution of galaxies. The CMB had already been discovered by Penzias & Wilson in 1965 [1]. The fact that this discovery was made by chance – and by radio engineers working for Bell Telephone

labs who failed to find an alternative explanation for the pervasive background noise they heard — testifies to the minor role cosmology played in serious physical or astronomical research. This becomes especially clear when one notes that the CMB was actually *predicted* to exist decades earlier by George Gamow [2].

Then several discoveries, perhaps primary among them the 1992 detection of CMB fluctuations in by the COBE satellite, heralded a new era for cosmology, one that is still developing into its golden years.

The CMB has the most perfect black body Planck spectrum ever observed. As such, it is a relic of a once reigning cosmic thermal equilibrium with a definite temperature that does not vary much from place to place. The Universe is currently evidently not in such a state, with extreme temperature variations that may range from billions of degrees to near absolute zero. Most of it, particularly the CMB itself has cooled to a temperature of near zero by virtue of the redshifting of its photons by cosmic expansion — about the only firm observational fact of physical cosmology up to the second half of the twentieth century.

The current Universe is also highly inhomogeneous, with matter clustered into galaxies and clusters and larger scale structures, and in stars and planets (and so on) on smaller scales. To reach this state from an initial 'cosmic soup', fluctuations had to exist in this soup. Under gravity these could in principle grow to seed the clustered universe we see. This was the major discover of COBE, for which the Nobel prize was given in 2016. These fluctuations were studied in ever more details through the subsequent decades; through the WMAP and PLANCK satellites, as well as a host of ground and balloon based experiments [3].

By pointing out instruments to different directions in the sky, the tiny differences — of order of one part per hundred thousand — could be measured, then expanded into spherical harmonics and their power spectrum of peaks studied. The power spectrum of the modes that lead to structure in the Universe could be inferred. The results suggested some stunning corollaries, already evident from other lines of astronomical observations such as supernova surveys and galaxy rotation curves: in order to fit these peaks one had to assume that the vast majority of the Universe consisted of hitherto unknown components. There is dark matter, which did not couple to CMB radiation and therefore could collapse early as the universe expanded and cooled. And there is dark energy, accelerating the expansion of the Universe and placing the peaks of the CMB power spectrum in exactly the right place (the pages [background.uchicago.edu/whu/](http://background.uchicago.edu/whu/) discuss these issues at several levels). By analysing the CMB spectrum in terms of angular mode power (via spherical harmonics) and fitting it to theoretical model predictions, it was possible to precisely derive parameters such as the the age and content, and the form of the spectrum of initial fluctuation on a range of scales, despite the missing pieces of dark matter and energy.

For the dark energy, the simplest explanation for it was Einstein's cosmological constant  $\Lambda$ . But its origin and energy scale (and whether it is really constant or not) remain deep enigmas. We will be mainly concerned here with the role of dark matter. After introducing some basic cosmological concepts, we quantitatively discuss the CDM model of structure formation, which has been dominant for decades and lead to many successful predictions regarding the large scale structure of the Universe as revealed by the CMB and redshift surveys [4]. We discuss why the weakly interacting massive particles (WIMPs) associated with it arise 'naturally' with the right abundance from an early cosmic thermal equilibrium, and how the parameter range of this 'WIMP miracle' is being closed off by empirical searches. We discuss the properties of collapsed cold dark matter structures, the haloes, and their apparently universal density and phase space distributions. We present new simulations attempting to explain aspects the still ill-understood origin of that universality, which requires invariance under merging. We then discuss problems facing this model from an astrophysical point of view and their possible solutions within CDM context, by invoking interactions with baryons, and from without, by changing the particle physics (or even gravitational) model. For it turns out that the cold dark matter may be too

cold, and thus may need heating, and that this can be done in several ways, some involving complex baryonic gas physics others involving fundamental physics.

## 2. The basics of cosmic expansion and collapse of smaller structures

For a spherical mass distribution, Newton's theorem says that the gravity originating from that distribution at any given point is due solely due to the mass shells from the radius of this point to the centre of the sphere [5]. For self gravitating gas with pressure  $P$  and mass density  $\rho$  one can then write for a spherical shell at radius  $r$  (and in Lagrangian coordinates moving with the shell)

$$\frac{d^2r}{dt^2} = \rho \frac{-GM(< r)}{r^2} - \frac{dP}{dr}. \quad (1)$$

If the pressure is negligible, this is simply the radial equation of motion a particle moving in the gravitational field of a mass  $M$  at  $r = 0$ . Its solution is the cycloid solution of a radial trajectory of the Kepler problem. If the pressure is negligible and the sphere homogeneous, with density  $\rho_0$ , then  $M = \frac{4\pi}{3}r^3\rho_0$ , the equation is that of a harmonic oscillator with characteristic period  $\sim 1/\sqrt{\rho_0}$ . This 'dynamical time' is characteristic time for a trajectory to cross a system with density "rho. It depends only on the density and not directly on the mass and spatial scales of the system. It is also of the order of the characteristic time over which would collapse under gravity (again characterized by the cycloid solution) if no pressure is present.

As a corollary of the above, in an initially homogeneous matter dominated universe, any perturbation collapses on the dynamical time unless some pressure (or other) force prevents this. If pressure forces are present, then there is a minimal scale that will not collapse. It can be estimated through the sound speed  $c^2 \sim P/\rho$  from  $c_s^2 \sim R^2G\rho$ , where  $R$  is the initial size of the system. For  $R$  less than the 'Jeans length'  $R_J$  the system collapses, as the speed of the collapse under gravity  $R\sqrt{G\rho}$  is greater than the sound speed associated with pressure transmission  $c_s$ .

Letting  $P = 0$  and integrating (1) for a shell inside which mass is conserved (valid also for a whole evolving sphere if shells do not cross) we can obtain the energy integral

$$\frac{1}{2} \left( \frac{dr}{dt} \right)^2 - \frac{GM(< r)}{r} = E. \quad (2)$$

The Friedmann equation describing the (Hubble) expansion rate of a marginal universe ( $E = 0$ ) can be obtained by simply setting  $r(t) = a(t)r_0$ , where  $a(t)$  is referred to as the scale factor. By assuming spatial homogeneity ( $\rho = \rho(t)$ ) and naively setting  $M = \frac{4\pi}{3}\rho r^3$  one finds:

$$\left( \frac{\dot{a}}{a} \right)^2 = H^2 = \frac{8\pi}{3}G\rho. \quad (3)$$

This Newtonian derivation was possible since, for a homogeneous and isotropic universe (i.e. one described by this equation), Newton's theorem (Birkhoff theorem in general relativity) implies that taking the evolution of any isolated sphere is sufficient to represent the evolution of the scale factor in time, since this evolution is unaffected by the mass distribution outside the sphere. It is also the same inside any sphere by virtue of global homogeneity.

A more careful Newtonian derivation can be found in the classic text by Binney & Tremaine [5], but it still leaves open questions as to what the meaning of the energy constant  $E$ , which we took to be zero, is — and conceptual issues, like what are all the spheres expanding into? A general relativistic derivation is possible by using the Friedmann Robertson Walker metric, which can be written in polar spherical coordinates as

$$ds^2 = -c^2 dt^2 + a(t)^2 \left[ \frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2) \right], \quad (4)$$

where ( $k = -1, 0$  or  $1$ ) determines the spatial curvature. This metric can then be used in conjunction with the Einstein field equations

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \quad (5)$$

where  $G$  is the Einstein tensor, dependent on the metric and its derivatives and  $T$  describes the matter, energy and momentum content of the universe. For an isotropic and homogeneous universe the only non-zero components are the time time one  $T_{00} = \rho c^2$  and those corresponding to isotropic pressure  $T_{ij} = a^2 P g_{ij}; i, j = 1, 3$  [2].

In this context, it turns out that the constant of integration of the Newtonian theory is  $2E = -kc^2/a^2$  and is thus related to the spatial curvature of the Universe. Also, that  $\rho$  in equation (3) corresponds not only to Newtonian mass density but also the radiation density ( $\rho_r = \frac{1}{3}\epsilon_r$ , or that of any other component (e.g, any vacuum energy density, including a cosmological constant as we have assumed in writing the Einstein equations that this can be absorbed into the tensor  $T$ )). Current observations suggest that  $k = 0$  and therefore the Friedmann equation derived above (with  $E = 0$ ) is relevant.

Note also that the expansion time of the universe  $\sim 1/H$ , described by (3) is  $\sim 1/\sqrt{G\rho}$ ; that is, the same as the collapse time (the dynamical time) of a matter perturbation of the same density. This has the following significant consequence. Suppose that the total energy density has matter and radiation components such that  $\rho = \rho_m + \rho_r$ . Radiation cannot collapse under gravity because radiation pressure is associated with a Jeans length of the order of the light horizon. Matter can in principle collapse on a time scale  $1/\sqrt{G\rho_m}$ ; but if  $\rho_r \gg \rho_m$ , the expansion time is much smaller than that, so any density fluctuation will be stretched out by the general Hubble expansion with no chance for collapse.

As we will see below, the energy budget of the universe is dominated by radiation before matter comes to dominate. According to the argument above, density perturbations cannot grow much during that radiation dominated era (or dark energy domination for that matter).

When matter dominates, a spherical overdensity will initially expand with the universe, but at a slower rate and increasing overdensity (that is density inside the sphere relative to the average background density). Eventually it can then turn around and recollapse. The resulting structure then 'virialises', and the kinetic energy of motion balances the gravitational potential energy such that, for typical particle speeds  $v$  and characteristic mass and size  $M, R$ ,  $M\langle v^2 \rangle \sim -GM/R^2$ . If the density perturbations are in the dark matter component the virialised structures are termed 'haloes' [5, 6].

### 3. The CMB and what it tells of matter radiation equality and recombination

The current CMB temperature is about 2.7 degrees Kelvin. But since the Universe is expanding, this radiation must have been compressed to a much higher temperature in the past. It must have also been strongly coupled to charged particles in a fully ionized hot universe. Today the energy density of the CMB is very small  $\rho_{\text{CMB}} = 4.2 \cdot 10^{-14} \text{ J m}^{-3}$ . This is much smaller than corresponding to the rest mass of a proton per meter cubed ( $= 1.6 \cdot 10^{-10} \text{ J m}^{-3}$ ). Nevertheless, there are many more photons than protons: for a Planck spectrum, the energy of a typical photon can be estimated from its frequency from  $h\nu \sim kT$ , which for the aforementioned energy density  $\rho_{\text{CMB}}$  implies the existence of about  $10^8$  CMB photons per meter cubed. This could be compared to the average number of protons in the Universe per  $\text{m}^3$ , which is of order one.

The density of a fixed mass of nonrelativistic matter inside a volume  $r^3$  proportional to  $1/r^3$ . If this volume is expanding with time, such that  $r(t) = r(t = 0)a(t)$ , the density of matter decreases as  $\sim 1/a^3$ . The energy density of radiation however decreases as  $1/a^4$ , since the number of photons decreases  $1/a^3$  and the energy of each photon decreases as  $\sim \nu \sim 1/\lambda \sim 1/a$ , due to the redshifting and stretching of wavelength  $\lambda$  of photons as the Universe expands. Given

this, there is an epoch in the past when radiation was the dominant form of energy, despite its negligible present contribution. In cosmology, this era is termed the 'radiation domination era' (see Kolb & Turner for more details and careful discussion of the thermal evolution of the Universe). As the ratio of the energy densities varies as  $\sim a$ , the numbers quoted above suggest matter and radiation energy density were equal at  $a_0/a_{\text{eq}} \sim 3800$ , where  $a_0$  is the current scale factor, and assuming the current mass of nonrelativistic matter to be exactly 1 proton per meter cubed. More careful estimates show that  $a_0/a_{\text{eq}}$  is about 3400.

As we saw in the previous section, matter density perturbations cannot grow much if the energy density is dominated by radiation (the so called Mezsarov effect; see also [6]). Therefore one has to wait till the Universe has expanded sufficiently, such that  $a > a_{\text{eq}}$ , before perturbations can grow. Furthermore, density perturbations of 'normal' matter — that is 'visible' matter that can electromagnetically couple to radiation — cannot grow till significantly later, when the Universe has sufficiently cooled, so that the primordial plasma has recombined into atoms and the radiation is free to travel through the neutral medium. The start of this epoch can be estimated as follows.

If there is approximately one photon per baryon, and considering for simplicity a purely hydrogen plasma, then the ionised fraction due to excitation by energetic photons is  $n_i \sim e^{\frac{-E_B}{kT}}$ , where  $E_B$  is the binding energy of Hydrogen  $E_B = 13.6\text{eV}$ . But there are as we saw at least  $10^8$  photons per proton, so a transition to a neutral fluid occurs when  $10^8 \sim e^{\frac{-E_B}{kT}} \lesssim 1$ . This corresponds to a temperature of about  $0.7\text{eV}$  according to this estimate. A more careful calculation shows this to be  $0.3\text{eV}$ , which corresponds to  $3600\text{K}$ . As  $T \sim a$ , for a current CMB temperature of  $2.7\text{K}$ , this corresponds to  $a_0/a_{\text{rec}} = 1300$ .

Using (3) in the matter dominated era one finds that  $a/a_0 = (t/t_0)^{2/3}$ , which implies that at the time of recombination the universe was about 300000 years old, if the present age is of order of 14 Gyr old. After recombination, the cosmic microwave background is only slightly affected by any other interactions; it can be observed as a snapshot of the Universe at this early age, unveiling a wealth of information.

The coupling of the CMB to the 'normal' baryonic matter component till  $a_0/a_{\text{eq}} \sim 1000$  has the consequence that the matter perturbations in the baryonic component would be late to develop. They are also washed out at scales smaller than those of galaxy clusters (Silk, or diffusion damping. [7]). This was known to already be problematic since Zeldovich proposed his 'top down pancake' structure formation scenario, given the lack of mechanism and time for fragmentation of the pancakes into galaxies. It is now known to be quite implausible, given the rich observations of galaxies and supermassive black holes already at distances corresponding to times of a few hundred million years since the start of the expansion [8, 9] (with no sign of cluster size objects that these could have fragmented from). Evidence for non-baryonic dark matter comes also from big bang nucleosynthesis, which limits the fraction of baryonic matter in the Universe, in order to match observations of the lighter element abundances [10].

#### 4. Cold dark matter and the WIMP miracle

In the phase of thermal equilibrium dominating the early universe, a particle is relativistic if its rest energy is small relative to the energy associated with the ambient radiation temperature:  $mc^2 < kT$ ; or, in natural units (where both  $m$  and  $T$  are measured in eV), simply  $m < T$ . Below this, kinetic equilibrium means that its speed is such that  $mv^2 \sim kT$ , as its velocity distribution is Maxwellian. Chemical equilibrium implies another Maxwellian, for the number density arising from detailed balance of production and annihilation in the thermal bath. In natural units this is

$$n_X \sim (m_X T)^{3/2} e^{-m_X/T}, \quad (6)$$

where the subscript  $X$  here refers to properties of the dark matter particle. The number of photons, as we saw, is very large, it is therefore practically conserved during the thermal equilibrium phase — and also, of course, when it essentially free streams following the decoupling from the primordial plasma as discussed above. Since the energy density of CMB photons is  $\sim T^4$ , and the energy of a single photon  $\sim T$ , we expect the number density of photons at temperature  $T$  to be of order of  $n_\gamma \sim T^3$  (full details can be found in [11]). Thus the number of dark matter particles per photon is

$$\frac{n_X}{n_\gamma} = (m_X/T)^{3/2} e^{-m_X/T}. \quad (7)$$

This number is nearly invariant once the DM particles have decoupled from the primordial thermal soup. In order to fit observations we need about 5 times more dark matter than baryons, currently the above ratio is therefore

$$\frac{n_X}{n_\gamma} \sim 5\eta \frac{m_P}{m_X}, \quad (8)$$

where assumed all baryons are simply Hydrogen and  $m_P$  is the proton mass (about 1GeV) and  $\eta \sim 10^{-9}$  is the baryon to photon ratio.

If the DM is assumed to be weakly interacting, decoupling would be expected to happen much earlier than the electromagnetic decoupling discussed above. Again, as a rule of thumb, this happens when the interaction rate becomes smaller than the Hubble expansion rate. The interaction rate for DM is  $\sim n_X \sigma v_X$ , where  $\sigma$  is the crosssection of (weak) interactions between DM particles; it is believed to be of the order of  $10^{-8} \text{GeV}^{-2}$  in the relevant non-relativistic regime. During the radiation era  $\rho \approx \rho_r \sim T^4$  and in natural units  $G = 1$  and the reduced Planck mass ( $M_{\text{Pl}} = 2.4 \cdot 10^{18} \text{GeV}$ ) takes its place in equation (3), which becomes  $H \sim T^2/M_{\text{Pl}}$ . Then using this and equation (8) one gets decoupling at

$$\frac{m_X}{T_{\text{dec}}} \approx 26.6 \left( \frac{\eta}{6.5 \cdot 10^{-10}} \frac{\sigma}{2 \cdot 10^{-8} \text{GeV}^{-2}} \right)^{2/3}, \quad (9)$$

where we have assumed typical numbers for  $\eta$  (from big bang nucleosynthesis) and the thermally averaged weak interaction crosssection. Inserting this into (7) gives the same conserved number of DM particles per photon expected from current measurements — i.e. as described by (8) — if  $m_p/m_X \sim 10^{-2}$ , which is of the order of the electroweak scale! This is essentially the so-called WIMP miracle. Note however how sensitive the chain of argument is to the chosen values of  $\eta$  and  $\sigma$ . Changing their product by just one order of magnitude, up or down, gives very different results.

In addition to 'naturally' transpiring with right abundance from the early thermal equilibrium, with no extra physical input other than the assumption of chemical equilibrium and a simple recipe for decoupling (more formal derivation using the Boltzmann equation can be found in [11]), they are slowly moving due to their assumed heaviness (since, again, kinetic equilibrium requires  $v \sim \sqrt{T}$ ). This is important for a bottom up structure formation scenario, so that smaller objects can form. To see this more clearly consider the situation during the radiation era, the dark matter cannot collapse yet into the 'haloes' that would later seed and parent the galaxies, it free streams. The comoving (that is measured on present distance scales) free streaming length at the end of the radiation era is then

$$\lambda_{\text{FS}} = \int_0^{t_{\text{eq}}} \frac{v(t)}{a(t)} dt. \quad (10)$$

Up to the time the particle is still relativistic its speed  $v \sim c$ . After this, if it is coupled to the primordial plasma, kinetic equilibrium implies  $v \sim \sqrt{T/m} \sim (ma(t))^{-1/2}$ , then it goes as  $\sim 1/a$  after it decouples.

Clearly, for a given interaction crosssection, the more massive the particle the smaller its free streaming length, and the initial condition for the  $v \sim 1/a$  stage is determined by the decoupling temperature. On the other hand, standard neutrinos decouple relativistically – they are 'hot dark matter (HDM) — their number density does not decay due to Maxwellian suppression in chemical equilibrium so their numbers are large, which means their mass must be small (combined mass  $< 50\text{eV}$ ) if they are not to contribute much more than the observed mass density of the Universe (including DM). The  $v \sim 1/a$  stage, if it exists, is set by the transition to non-relativistic evolution (at temperature  $kT \sim m_\nu c^2$ ). It is then easily shown — assuming  $a \sim t^{1/2}$ , as expected from solution of Friedmann equation above with  $\rho = \rho_r \sim 1/a^4$  and  $T \sim 1/a$  — that  $\lambda_{\text{FS}} \gtrsim 10\text{Mpc}$  [12]. This is larger than the scale of galaxy clusters. So, as baryonic matter, standard neutrinos are not good DM candidates.

If the dark matter is to seed galaxy formation from the smallest scales up, this must be smaller than the expected size of haloes of small galaxies, that is  $\lesssim \text{kpc}$ , which corresponds to  $m_d \gtrsim \text{keV}$  – or a millionth of a proton mass.

The models that correspond to the WIMP miracle predict masses that are much larger than this, the associated minimal length scales are actually of the order of the radius of the earth. The structure formation scenario based on such DM particles is called the cold dark matter scenario (due to the small speeds associated with the heavy DM particles). It predicts that structures form 'hierarchically' or bottom up; with small haloes collapsing first, then merging into successively larger structures if the initial power spectrum of potential fluctuations is flat, as predicted by sourcing these fluctuations from quantum fluctuations during an early inflationary era [13].

Already in 1978, it was predicted that the smaller scale structures will largely wash out, as they are stripped through the process of merging and a single 'parent' halo provides the main contribution to the mass. Disks can form inside such haloes from gas that settles in the ir centres by dissipating energy and conserving angular momentum while cooling [14].

The WIMP based CDM model then developed into a quite successful model of structure formation. With an added cosmological constant  $\Lambda\text{CDM}$  is faring particularly well when confronted with observations of the large scale galaxy distribution and lensing; its almost perfect fits to the CMB fluctuations, and a host of observations such as the baryon acoustic oscillations, render it still the most plausible and favoured paradigm of structure formation [15].

Nevertheless, two severe problems confront the WIMP based CDM scenario. The first is that the WIMPs were never found, despite extensive direct detection and accelerator based searches. Indeed, the particle physics theories that give rise to them, based on 'naturalness' of models such as minimal supersymmetry are being ruled out, as is the crosssection space corresponding to the WIMP miracle. Furthermore, and this is the subject of the rest of this presentation, the CDM scenario suffers from problems when it comes to the dynamics of galaxies on smaller scales.

## 5. CDM haloes and their discontents

### 5.1. The structure of haloes

The cold dark matter scenario is so successful in explaining structure formation in part because 'cold' particles can collapse on sufficiently small scales early on and merge hierarchically, in order to timely form the seeds of galaxies. It may turn out however that the CDM particles' coldness may also be the model's bane. Due to their small velocities, CDM overdensities begin their gravitational collapse when they effectively inhabits three dimensional sheets embedded in the a six dimensional phase space of particle positions and velocities. These structures are effectively

of infinite density, which is conserved by virtue of Liouville's theorem through the collapse (at least at the fine grained level). As the phase space density  $\rho \sim \rho_p \sim \rho/v^3$  — where  $\rho$  is as usual the spatial density and  $v$  the typical particle speed — this results in the possibility of regions of very high spatial densities and low velocities, which is what is indeed found in numerical simulations (see [16] for an overview of current simulation methods in cosmology; [17] for how these simulations lead to diverging phase space densities near the centres of haloes).

Decades ago, as the resolution limit of numerical simulations approached that required to resolve the centres of the collapsed structures, it had already become apparent that the central densities of the haloes seemed to increase up that limit with no saturation [18, 19]. Then it became apparent that the profiles were quasi-universal, in the sense that they could always be approximately fit by the same two parameter model

$$\rho(r) = \frac{\rho_s}{\frac{r}{r_s} \left(1 + \frac{r}{r_s}\right)^2}, \quad (11)$$

where  $r_s$  is a scale radius separating a  $1/r$  density 'cusp' and subsequent quasi-isothermal  $\rho \sim 1/r^2$  region (with nearly flat rotation curve) from the outer  $\rho \sim 1/r^3$  density falloff. The scaling density  $\rho_{s0}$  is approximately the RMS density inside  $r_s$  (more precisely  $\rho_s = \sqrt{\frac{8}{7}\langle\rho^2\rangle_{r_s}}$ ).

This discovery by Navarro, Frenk & White [20] generated a large body of work, not least aiming at understanding the origin of the apparent 'universality'. This is not an easy feat, especially given the absence of general phenomenological or statistical-mechanical guiding principles determining the dynamical equilibria of gravitational systems; their thermodynamics is generally not well defined, and no maximum entropy states, strictly speaking, exist for finite-mass systems without a physical boundary. Two body relaxation due to gravitational encounters between particles is far too slow to have much effect in determining DM halo equilibria, and in case would lead to a runaway thermodynamic instability characterised by a contracting core and expanding outer mass in any case [5], a result of 'gravothermal catastrophe' [21]. Violent relaxation (due to collective effects [22, 23]) does not complete, and it predicts isothermal systems, as opposed to the simulated halo velocity profiles, which actually have a temperature inversion, with the velocity dispersion decreasing towards the centre (and, as we will see below, distribution functions that are power laws in the energy instead of exponential as expected of a Boltzmann distribution). Indeed, it has been argued by the author that, far from violent relaxation being at the origin of the universal halo profiles, it is the early abortion of the process, due to the overdamping of the collective modes that drive the relaxation process, that leads to centrally concentrated profiles efficient at damping these modes [24].

Yet, the universal profiles are ubiquitous. Haloes identified at any mass scale and any time during the evolution of numerical simulations have them, at least approximately, with no sensitive dependence even on the details of the initial fluctuation spectrum. They are also unlikely to be an artifact of those simulations, given the extensive research by many different groups using various techniques and convergence studies. But despite intense research into their origin there is no consensus on the mechanism that leads to their ubiquity.

In principle the mechanism leading to the universal profiles could originate either through from the initial monolithic collapse or from the subsequent process of merging that haloes are subjected to a hierarchical structure formation simulation. Given that monolithic cold collapse does indeed lead to profiles of the sort parametrised by equation (11), it seems clear that these should at least be invariant under merging, in order to persist.

One mechanism whereby this can be achieved, at least for minor mergers between a large halo and smaller ones, is through dynamical friction. This is a process that is similar to regular frictional force born of random collisions and governed by the fluctuation dissipation relations. Except here the 'collisions' are gravitational encounters rather than actual impacts

as in Brownian motion for example. Indeed, in seminal studies, Chandrasekhar had reviewed both situations (laboratory and gravitational) and laid out much of the modern foundations of the subject [25, 26]. For the gravitation frictional force on a particle of mass  $M$  moving with velocity  $\{\mathbf{v}\}$  through a homogeneous medium of density  $\rho_0$  (made of much lighter particles) and Maxwellian velocity distribution with one dimensional dispersion  $\sigma$ , he deduced the dynamical friction formula

$$\frac{d\mathbf{v}}{dt} = -\frac{4\pi G^2 \ln \Lambda \rho_0 M}{v^3} \left[ \text{erf}(X) - \frac{2X}{\sqrt{\pi}} e^{-X^2} \right] \mathbf{v} \quad (12)$$

where  $X = \sigma/\sqrt{2}$  and  $\Lambda$  is the Coulomb logarithm, which is the ratio of the maximal and minimal impact parameters of stochastic gravitational encounters leading to the frictional force. The logarithmic divergence when the minimal gravitational 'collision' impact parameter goes to zero reflects the singularity in the potential, while the divergence at large cutoffs reflects the long range nature of the gravitational force.

One notes that for small  $v$  the acceleration described by equation (12) is simply proportional to  $v$ ; as with regular frictional force, associated with Brownian motion for example. For larger  $v$ , however, it decreases as  $1/v^2$ . This reflects the fact that, for larger relative speeds the deflections associated with gravitational encounters (as in plasmas with similar Coulomb form of the potential), decrease in strength. The associated rate of energy loss from the massive particle is in this case is given by

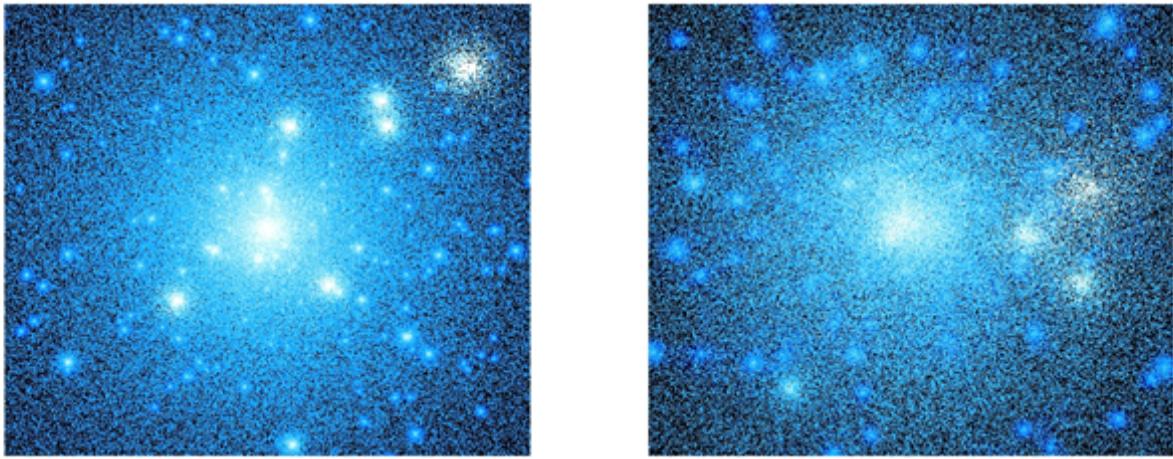
$$\frac{dE}{dt} \approx -\frac{4\pi G^2 \ln \Lambda \rho_0 M}{v}. \quad (13)$$

In work discussed below, such energy transfer was considered. The direction of this dynamical friction mediated energy transfer is from compact massive clumps — halo substructure of baryonic gas lumps or bound stellar clusters — to the background material making up the galaxy halo, initially following the NFW profile (of equation 11).

Using a Fokker-Planck formulation of the aforementioned process, and idealised numerical simulations involving solid massive objects, it was argued by the author that mass inflow, resulting from the loss of energy by massive substructure, was balanced by an outflow of the light background particles gaining energy, in such a way that the total profile (spatially and in phase space) remains invariant [27]. Again, this seems a necessary condition for the observed universality of the halo profiles, as cold collapse leads to these density distributions [28, 29, 30, 31], but for the profiles to persist in the context of hierarchical structure formation, the structures need to be invariant under the effect of the interaction of the clumpy CDM medium, including repeated accretion and merging into 'higher' (more massive) structures.

The results of the simulations shown in Fig. 1 and 2 may clarify further what is meant by this invariance. Here, as opposed to the simulations of [27], the clumps are not represented by solid particles, but are 'live' subhaloes made of CDM particles themselves. These simulations, though still idealised, are more realistic. We start with an NFW density profile for both main halo and subhalo distribution, the velocities are adjusted so that the system is initially in dynamical equilibrium (which is not exact, due to the existence of a few large clumps). The clumps internal density distribution is also of NFW type, as they represent the remaining relics from the process of merging. In line with what is typically found in cosmological simulations, the subhalo system is assumed to constitute a total of 20% the mass of the halo. This, after truncation of their NFW profile to the tidal radius (the point in their radii where their gravitational field is equal to that of the parent halo; e.g., [5] for discussion). The subhaloes follow a mass function derived from cosmological simulations. and the parameters of the simulations correspond to a Milky Way like galaxy.

The results shown Fig. 2 are averaged over three simulations (each involving  $2 \times 10^6$  particles and runs for 16 dynamical times  $\tau_{\text{D}}(r_s) = 1/$



**Figure 1.** Snapshots showing mass distribution of simulated dark matter haloes. Left hand panel: Initially the background mass distribution, assumed to be in the form of CDM, follows the NFW profile (11), and the subhaloes (the clumpy structures) are also distributed according to this density distribution; the velocities of the background particles and the centres of mass of subhaloes are chosen such that the system starts from dynamical equilibrium (up to fluctuations due to finite number of subhaloes); the mass function of the subhaloes is such that  $\frac{dn(M_{\text{sub}})}{dM_{\text{sub}}} \sim M_{\text{sub}}^{-9/5}$  [32]. Right hand panel: After 16 dynamical times, more massive subhaloes have spiraled to the centre due to dynamical friction, 'heating' the smooth background material in the process; stripping significantly reduces the total mass in the clumpy component.

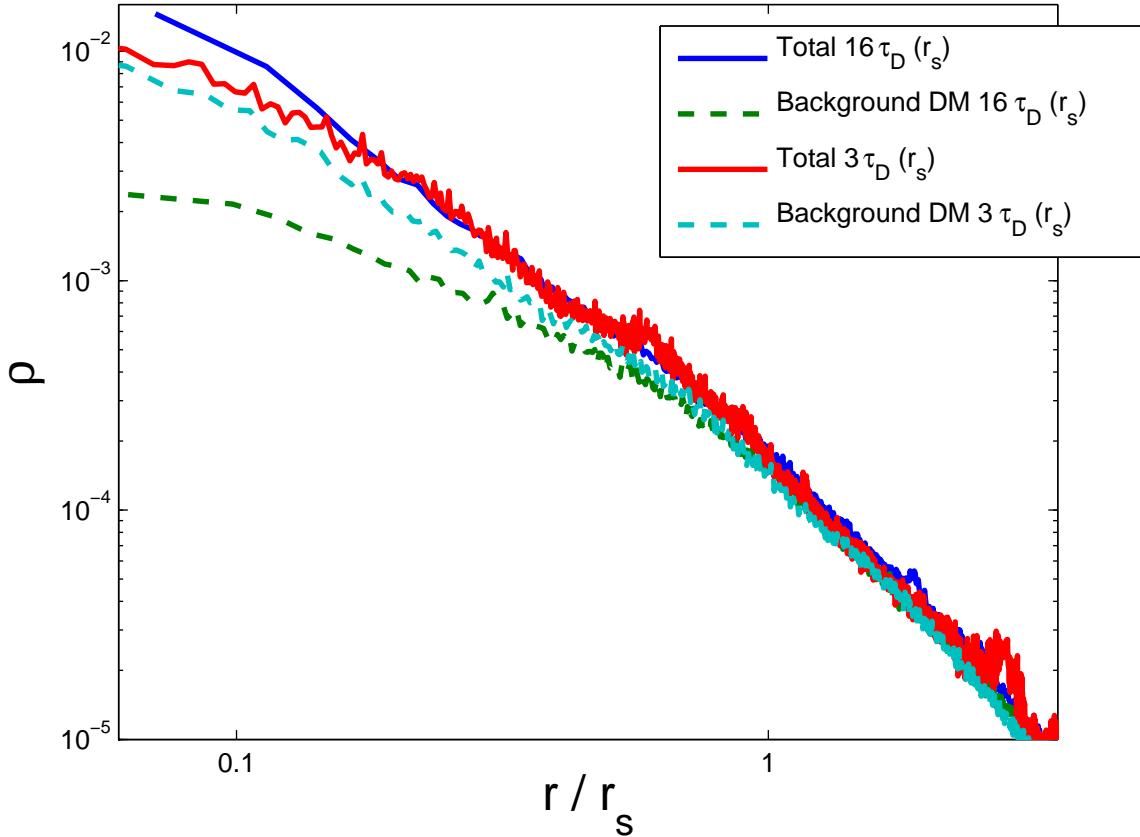
$\sqrt{G}\langle\rho_s\rangle$ , where  $\langle\rho_s\rangle = 3M(< r_s)/(4\pi r_s^3)$ . Evolution leads to two clear effects: first, there is the spiraling in of the heavier clumps into the centre; this is due to dynamical friction-mediated transfer of energy from clumps to background as discussed above. It is in addition accompanied by further stripping of the subhaloes, such that the mass fraction in clumps evidently decreases between the snapshots. As can be seen from Fig. 2 however, these processes do not lead to significant change in the total mass profile — even if they do change the mass distribution of each (clumpy and background) component significantly.

In [27], the effect of dynamical friction was quantitatively discussed in terms of a Fokker-Planck formulation, and it further argued that stripping would not change the total profile, because as long as the particles are stripped off the subhaloes with approximately the same energy as the larger halo itself (assuming internal energy of the halo is small compared to the energy resulting from motion in the parent halo potential), then the total mass distribution function remains invariant. The results shown here, — where stripping of the live haloes is automatically included in the computation — seem to confirm this.

### 5.2. The problems with simulated haloes

The main problem with haloes identified in cosmological simulations is that their central density seems to be too large; they contain too much matter in the centre. This 'cusp/core' crisis may be significant; as one of the main reasons that dark matter was introduced into astrophysics was to explain the flat outer rotation curves of galaxies. There was insufficient mass there and so it seemed that matter had to be added *via* extended haloes of dark matter that did not cool and shrink (given that they did not radiate electromagnetically).

But if such structures are too centrally concentrated, containing too much mass in the centre, then the solution to the problem of apparent lack of visible mass in the outer parts of galaxies entails a no less serious one of an excess of mass in the centre. To solve the outer dearth-of-mass



**Figure 2.** Evolution of density associated of the system shown in Fig. 1. After 3 dynamical times the system has relaxed to proper dynamical equilibrium (the initial equilibrium is not exact due to the existence of the clumps), but the effect of dynamical friction is still negligible. After  $16\tau_D(r_s)$ , dynamical friction has transferred significant energy from the clumps to the smooth background material, which is 'heated' out of the central region, a process which lowers its density there. The total profile remains largely invariant, despite the dynamical friction mediated coupling between clumps and background and the stripping of the former by the parent halo tidal field.

crisis, haloes with quite definite properties (resulting from the scaling inferred from simulations) are to be chosen and fitted to the outer rotation curve (with little wiggle room in choice of parameters), the ones that solve that crisis cannot at the same time be compatible with the measured mass distribution in the centres of galaxies.

In addition to the excess of central concentration, there appears to be a couple of additional, quite plausibly related, sub-galactic scale problems in the context of the WIMP based CDM scenario. First there are too many small subhaloes; for example in Fig. 1 — where the parent halo was populated with a number of subhaloes expected for one enveloping a Milky Way like galaxy — there are more than a hundred subhaloes, while the Milky Way hosts at most a few dozen identified satellite galaxies. Also, there is the 'too big to fail' problem, which involves the existence of massive satellite galaxies, whose counterparts, instead of being ubiquitous, are instead not found in simulations [33]. Given that smaller central densities of haloes can lead to efficient stripping and change in internal mass distribution of subhaloes, it seems quite plausible

that the solution of the cusp/core problem would automatically entail the resolution of these other two problems, through easing their dissolution into main halo material and the modifying the mass distribution of the bigger haloes that do remain [34].

Since its discovery in the mid nineteen nineties [35, 36], the cusp/core and allied problems have been subject to extensive investigation from both the theoretical and observational point of views. Observationally, most work has focussed on the case of dark matter dominated galaxies such low surface brightness and dwarf galaxies [37]. From a theoretical perspective, there have also been many attempts at a solution which could have wide repercussions for the physics of galaxy formation and beyond — as in changing the particle physics model for the DM. We discuss some of these below and refer the reader to recent reviews for further overview [38, 39].

## 6. Solutions to the small scale problems

With the exception of proposals invoking such effects as a cutoff in the primordial spectrum of perturbation — e.g, due to broken scale invariance during an early inflationary epoch [40] — solutions to the galactic scale problems of the CDM scenario come in two categories: effects born of the coupling of baryons to CDM during the epoch of galaxy formation; modifications to the particle physics models of the DM. Both involve, some way or other, ‘heating’ the cold dark matter, so as to move it from the centre of the halo and decrease its mass density there.

To see how this works, first note that the inner parts of CDM haloes density distribution (with central density  $\rho \sim 1/r$ ) correspond to a distribution function given (assuming isotropic density and velocity distribution) in terms of the energy  $E$  by [27]

$$f \sim E^{-5/2}, \quad (14)$$

where we assume that the zero point of the energy is taken as the centre, so that this energy is positive and the potential energy increasing with radius. This diverges as  $E \rightarrow 0$ , that is for particles near the near the centre of the system. The corresponding density is

$$\rho \sim \int_{E_0}^{\infty} f \sqrt{E - \phi} dE, \quad (15)$$

(where the  $\sim$  here refers to proportionality, as well as the pure power law approximation). If there is no cutoff in the distribution function ( $E_0 = 0$ ), then the potential  $\phi \sim r$ , and we recover  $\rho \sim 1/r$ . If a cutoff is introduced, representing the depletion of lowest energy states by the heating, a core results: depopulating the lower energy levels by ‘heating’ the CDM, so that  $E_0$  is non-zero, which leads to a ‘cored’ profile, with smaller density in the centre that is nearly constant with radius. To see this, note that for significant  $E_0$ , there will always be a radius inside which  $E_0 \gg \phi$ . Inside this region

$$\rho_{\text{core}} \sim E_0^{-3/2}. \quad (16)$$

for  $\phi \sim r$ , the extent of this core is  $\sim E_0$ .

We discuss baryonic solutions that invoke precisely this process in more detail below, here we first list the most popular modifications of the particle physics model that implicitly or explicitly invoke some form of ‘hotter’ dark matter:

- Warm dark matter (WDM) [41, 42, 43, 44, 45, 46, 47]. Here the DM is ‘preheated’. The simplest scenario corresponds to a thermal particle of mass near the free streaming limit (equation 10 implies a few keV). The mass is small so that its velocity is relatively large. WDM can eliminate the overproduction of small satellites but does not solve the cusp-core problem; the collapse is still cold unless the free streaming length is quite large, in which

case the galaxy does not form at all [44]. There are also strong constraints from Lyman- $\alpha$  alpha observation, which show the existence at intermediate redshifts (times) that would not exist in the most extreme WDM models. The loss of the 'WIMP miracle' (applicable to much more massive particles) also implies that WDM is difficult to produce with the right abundance.

- Self interacting dark matter (SIDM). Here, a scattering crosssection between the DM particles is invoked [48, 49, 50, 51, 52, 53, 54]. The heating of the DM in the central regions of haloes is then achieved through (heat) conduction from the outer parts. This eliminates the temperature inversion and leads to a core. Nevertheless, a certain amount of fine tuning seems necessary, since as already noted gravitational systems are thermodynamically unstable; further conduction leads to core collapse, born of gravothermal catastrophe, which is accompanied by an even more severe cusp/core crisis. There are constraints from the shapes of galaxy clusters (which become too round if much DM self-interaction occurs) and structure formation. Still, it has been argued that in practice SIDM can solve the cusp/core problem in dwarf galaxies without much fine tuning [54].
- Large scale quantum effects, the most popular models exhibiting these being associated with 'fuzzy dark matter' (FDM) made of ultra light axions [55, 56, 57, 58, 59, 60, 61, 62], which are predicted by string theories. In order to help solve the sub-galactic scale problems of CDM they would have to be so light so as to have a de Broglie wavelength of the order of the scales at which the problems with CDM appear. The 'heating' here is due to quantum fluctuations arising from the uncertainty principle. The inability to localise the particles naturally leads to a lower density core in the centres of the larger self-gravitating haloes and to the erasure of smaller structures. The wave nature of the particles however give rise to ubiquitous interference patterns that can give rise to density and potential fluctuations that may affect delicate structures such as tidal streams and disks and thus place constraints on particle masses. FDM models are discussed further below.

### 6.1. Baryonic solution

Ideas for baryonic solutions to the small scale problems associated with CDM-based structure formation can be divided into three (not necessarily independent) classes. The classes may involve:

- A single starburst of exploding young stars during the epoch of galaxy formation, which may remove the baryons out of the halo entirely, resulting in reduced gravitational field and expansion of the CDM [63].
- They may involve processes such as dynamical friction, to heat the cold DM out of the central parts, as baryonic clumps spiral in losing their energy and heating the ambient CDM during the process of galaxy assembly. This is similar to the mechanism discussed above for the invariance of the universal profiles, except that here the clumps are baryonic and not made of CDM subhaloes. Indeed, this mechanism was originally proposed by the author and collaborators [64, 65] before it was realised that it could also explain the persistence of the universal profiles. If the density distribution  $i$ (clumpy baryons plus CDM) is required to decrease then the baryons must originally be in a less centrally concentrated distribution than the CDM, or some of the baryons are eventually lost through feedback effects [64].
- Or they may invoke repeated input of energy, due to AGN or starburst-based energy feedback driving the gas and causing stochastic potential fluctuations that may heat the CDM [66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77]. This was proposed in an effort to remedy perceived shortcomings of the above mechanisms.

A single starburst may not be sufficient to expel enough mass, even in dwarf galaxies, so as to represent an efficient mechanism of central halo expansion. The dynamical friction mechanism, on the other hand requires that the baryonic clumps do not themselves suffer significant starbursts (if they are gaseous) that would cause them to disrupt on short timescales before they can have significant effect in heating the cusp.

Although the repeated feedback mechanism appears closer to the first in this list rather than the second, this is not necessarily the case, as we discuss below.

Since the original proposal [64], which relied on an idealised model and simulations, the idea of the coupling of CDM to baryon clumps, leading to the expansion of the former component, has been tested in more realistic simulations and semi-analytical studies. It is generally found to indeed ameliorate the problem of central CDM concentration significantly, though questions as to the nature and survival of the baryonic clumps remain [78, 79, 80, 81, 78, 82, 83, 84].

The original model of El-Zant et. al. [64] invoked a spherical system. However, a dissipative configuration, of the sort that may involve baryonic clumps of sufficient mass to heat the CDM halo cusp, would be more naturally expected in disk configuration. And indeed, since that original work, such clumps were found to be quite ubiquitous in early galactic disks through direct observations. And though their survival times are still debated, they can be shown in principle to couple sufficient to the CDM *via* dynamical friction as to heat the cusp [85, 86, 87].

Alternatively, or in addition, if one does not accept that individual baryonic clumps can survive long enough to heat the CDM, one may invoke the following situation: the clumps are not monolithic; instead, due to the pumping of energy from AGN or starburst, the gas is driven to a state akin to fully developed turbulence. It is therefore clumpy, but with the clumps continually disrupting and reforming; it then should develop density fluctuations with power spectrum akin to that of Kolmogorov's spectrum for velocity fluctuations in incomprehensible turbulence. The energy pumped into the CDM in this case is the starburst/AGN energy, converted into kinetic energy of a turbulent clumpy medium. It is then transferred to the CDM through stochastic fluctuations in density that lead to potential and force fluctuations that can heat the CDM. This here is the source of energy input, in contrast to the dynamical friction case, when orbital energy of individual clumps was directly deposited into the background CDM. Nevertheless, the situation is quite similar in both cases, in the sense that fluctuations due to a clumpy distribution in one medium lead to dissipation of energy from that mediated and into the smoother one made of the light particle. This is quite a different situation, physically and dynamically, from the case of a single starburst, where the gas escapes the galaxy entirely; as in the stochastic potential fluctuations mechanism the gas is driven and may suffer repeated inflow and outflows within a given radius, it never 'leaves'.

A model for the effects of such stochastic fluctuations, born of a fully turbulent compressible medium, was developed by the author and collaborators by assuming that they can be described by a random Gaussian process [88]. One first considers a medium that is homogeneous on large scales with average density  $\rho_0$  and density fluctuations  $\delta = \frac{\rho(\mathbf{r})}{\rho_0} - 1$ , inside a volume  $V$  that is large compared to the largest scale fluctuations. The potential and force fluctuations can be derived by solving the Poisson equation by Fourier expansion:

$$\Phi_{\mathbf{k}}(t) = -4\pi G \rho_0 \delta_{\mathbf{k}}(t) k^{-2}. \quad (17)$$

Thus the potential fluctuations are known once the power spectrum  $\mathcal{P}\langle\delta_k\rangle$  of density fluctuations is fixed. If the fluctuations are random Gaussian, this completely determines the stochastic process, and its force power spectrum. By virtues of large scale homogeneity, the the force correlation function can simply be derived by Fourier transforming the power spectrum of force fluctuation. For a system that is furthermore isotropic, this gives

$$\langle \mathbf{F}(0) \cdot \mathbf{F}(r) \rangle = \frac{1}{(2\pi)^3} \int \mathcal{P}_F(k) e^{i\mathbf{k} \cdot \mathbf{r}} d\mathbf{k} \quad (18)$$

$$= \frac{1}{(2\pi)^3} \int \mathcal{P}_F(k) \frac{\sin(kr)}{kr} 4\pi k^2 dk.$$

This spatial force correlation function can be transformed into a temporal one, along particle motion, by invoking the sweeping approximations widely employed in the field of turbulence. The resulting correlation function is one simply transformed with a typical speed  $v_r$  of the a test particles motion (representing a CDM halo particle) with respect to the fluctuating gaseous medium. To evaluate the effect of the fluctuation force described by the correlation function on the CDM particles, it can be inserted into a stochastic equation with solution of the form

$$\langle (\Delta v)^2 \rangle = 2 \int_0^T (T-t) \langle \mathbf{F}(0) \cdot \mathbf{F}(t) \rangle dt. \quad (19)$$

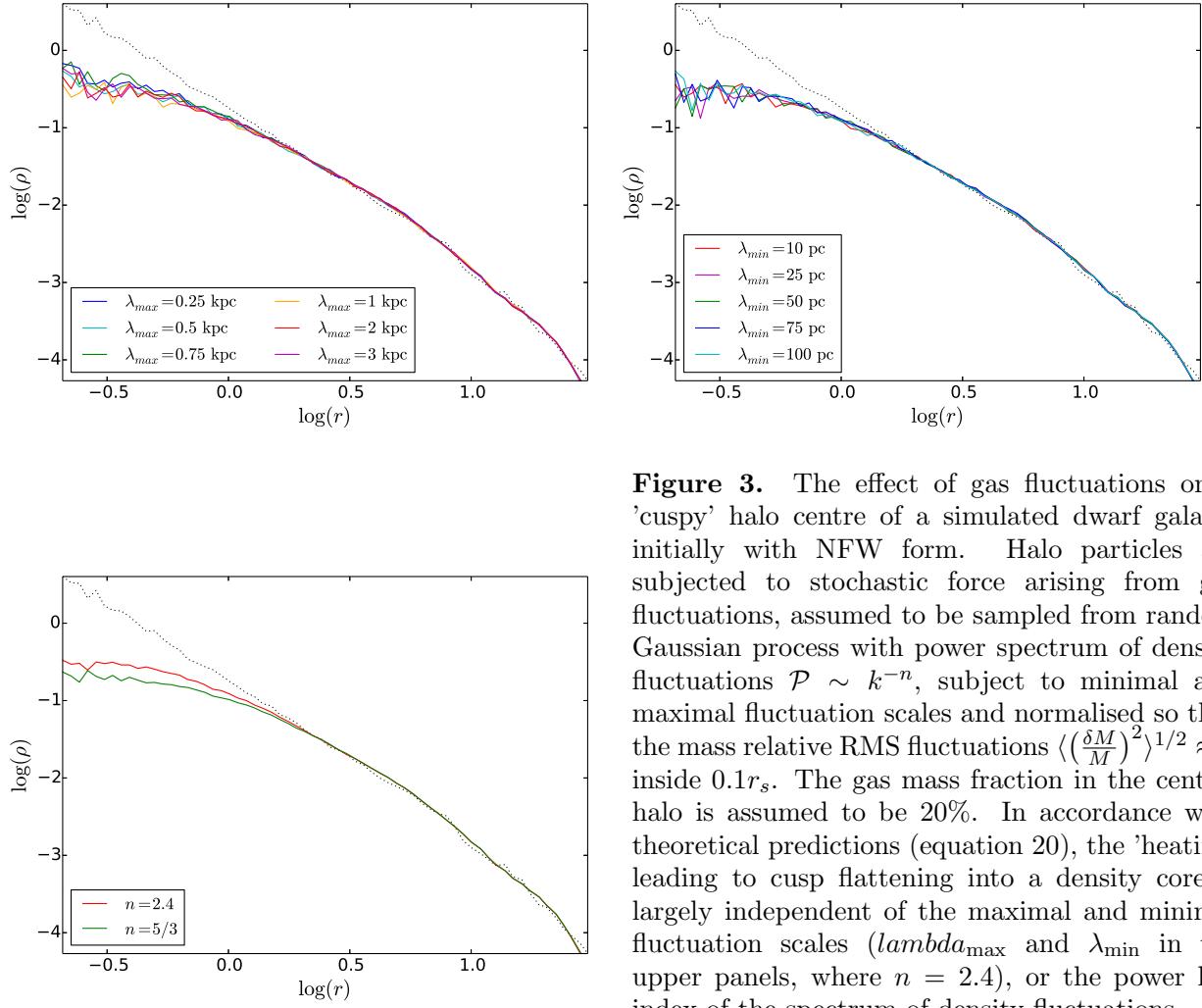
As usual (in two body relaxation studies of stellar dynamics for example), in the diffusion limit, a relaxation time can be defined by evaluating the time scale the velocity variance, described by the above equation, reaches the typical orbital velocity  $v$  of a CDM particle. If  $v \sim v_r$ , and assuming a power law density fluctuation power spectrum with of power law form ( $\mathcal{P}(k) \sim k^{-n}$ ), with upper and lower wave number cutoffs in wave number  $k$ , this is given by

$$t_{\text{relax}} = \frac{nv_r \langle v \rangle^2}{8\pi(G\rho_0)^2 \mathcal{P}(k_m)}, \quad (20)$$

where  $\mathcal{P}(k_m)$  is the value of the power spectrum at the minimal fluctuation wave number (maximal scale) and  $v$  is the typical speed of halo particles. This is  $\sim v_r$ , their speed relative to the fluctuations, as both the halo and large scale gas motions can be assumed to be in dynamical equilibrium with gravitational field of the halo. For typical halo velocities and gas mass fraction inside the central halo of the order of the universal baryon mass fraction ( $\sim 20\%$ ) and  $n \sim 1$ , this gives a timescale of the order of a Gyr, if the fluctuation levels are of the order of those observed in full high resolution hydrodynamical simulations with feedback, where parameters such as star formation thresholds are such that cusps are indeed turned into cores on similar timescales [74, 88] (note, in addition, that this timescale can be significantly shorted, due enhancement of the effect of fluctuations by collective modes of the halo, particularly non-radial one [88]).

This model was an attempt to explain process of core formation due to gravitational potential fluctuations from first principles; in simple dynamical terms sidestepping the very complex non-gravitational baryonic 'gastrophysics' — details of the heating and cooling processes in the driven, highly turbulent, clumpy gas, which is furthermore subject to ill-known physics related to star formation and its precise threshold, the feedback energy fraction retained by the gas and so on. To add to all the uncertainties from all this 'subgrid physics', there are numerical issues regarding the different methods of implementing the hydrodynamics and resolution levels etc... [16, 89].

In the context of the aforementioned model, on the other hand, the core formation process virtually depends only on the gas mass fraction and normalization of the power spectrum (which may in addition be correlated in practice through the star formation efficiency). This prediction is to be tested against full hydrodynamic gravity simulations, in order to understand whether the processes leading to core formation in some of them can indeed be characterised so simply and, in this case, also to understand precisely why some simulations do not produce such cores at all, or only under certain conditions, e.g. in connection to the star formation threshold, which should correlate with the fluctuation levels [90, 91, 92]). This work is ongoing, but simple idealised simulations already conducted already suggest that the core formation process is quite insensitive for example to the maximal and minimal fluctuation scales (3).



**Figure 3.** The effect of gas fluctuations on a 'cuspy' halo centre of a simulated dwarf galaxy, initially with NFW form. Halo particles are subjected to stochastic force arising from gas fluctuations, assumed to be sampled from random Gaussian process with power spectrum of density fluctuations  $\mathcal{P} \sim k^{-n}$ , subject to minimal and maximal fluctuation scales and normalised so that the mass relative RMS fluctuations  $\langle (\frac{\delta M}{M})^2 \rangle^{1/2} \approx 6$  inside  $0.1r_s$ . The gas mass fraction in the central halo is assumed to be 20%. In accordance with theoretical predictions (equation 20), the 'heating' leading to cusp flattening into a density core is largely independent of the maximal and minimal fluctuation scales ( $\lambda_{\max}$  and  $\lambda_{\min}$  in the upper panels, where  $n = 2.4$ ), or the power law index of the spectrum of density fluctuations.

### 6.2. Ultra-light axions, fuzzy DM and quantum effects on astrophysical scales

Fuzzy dark matter is made of axions with very small mass  $m_b$  (of order  $10^{-22}$ eV). Such particles can arise in string theories and can also play roles as drivers of early inflationary and late accelerated expansion phases [93]. The very small masses are associated with very large de Broglie wavelength. For typical dwarf galaxy halo parameters for example

$$\frac{\hbar}{m_b v} = 1.92 \left( \frac{10^{-22}\text{eV}}{m_b} \right) \left( \frac{10^{-22}\text{km s}^{-1}}{v} \right) \text{kpc.} \quad (21)$$

The 'heating' associated with such particles, when they are gravitationally confined to the central halo parts, is due to the spatial delocalisation arising from the uncertainty principle, which results in a maximal central (Bose condensate) density characterising the central 'soliton'. It may also be expected to result in washing out small scale structures and eliminating the excess of smaller haloes.

These expectations are indeed borne out in the few high resolution cosmological simulations conducted by solving the Schrodinger-Poisson system (e.g. [58])

$$i\hbar \frac{\partial}{\partial t} \psi(\mathbf{r}, t) = -\frac{\hbar^2}{2m_b} \nabla^2 \psi(\mathbf{r}, t) + V(\mathbf{r}, t), \quad (22)$$

$$\nabla^2 V(\mathbf{r}, t) = 4\pi G |\psi(\mathbf{r}, t)|^2, \quad (23)$$

where  $V$  is the self consistent gravitational potential. This system can be looked at as a mean field approximation of a many body quantum mechanical systems of the Hartree-Fock type. The wave function here however should be interpreted as a classical field, as it represents many particles in the same state and therefore a stable classical density not subject to the issue of the wave function collapse (a lucid discussion of such situations is given in Feynman's lectures [94]. The correspondence of this classical wave field to that of the many particle *classical dynamics* limit are studied in [95]).

Although the aforementioned simulations show that the small scale problems of the CDM scenario can be ameliorated by invoking FDM, whether haloes made of such particles can be in line with known galaxy scaling relation is not clear [96, 97]. There are also constraints on the Boson mass from Lyman- $\alpha$  and 21 cm observations of early structure formation [98, 99]. Furthermore, the large de Broglie wavelength and associated wave nature of FDM leads to interference patterns accompanied by density fluctuation, which like the fluctuations in the gas discussed above are associated with potential fluctuations. These in turn can have observable effects on baryonic components in galaxies and can put limits on the dark matter mass [61, 100, 101, 102].

Indeed, adopting the model for gas fluctuations described above, it was shown that the effect of such fluctuations in the ultra-faint dwarf galaxy Eridanus II can place strict constraints on  $m_b$  [101]. In general, it is possible to show that the methods employed in [88], to evaluate the effect on gas fluctuations on a test particle, can be adapted to estimate the effect of fluctuations on classical galactic components (such as stellar disks) embedded in FDM haloes (El-Zant et. al., in preparation). First, it is noted that the equivalent of equation (20) for white noise density fluctuation spectrum ( $n = 0$ ) is the standard two body relaxation time, defined as the time for a test particle, moving with relative speed  $v$  through a discrete set of 'field particles' of mass  $m$  each, to acquire a velocity variance of the order of its initial velocity

$$t_r = \frac{v^3}{8\pi G^2 \rho_0 m \ln \Lambda}, \quad (24)$$

where the Coulomb logarithm  $\Lambda$ , normally directly identified with the ratio of maximal and minimal parameters of individual two body encounters, is identified in the terms of the model described in [88] with the maximal and minimal fluctuation scales (see also [100]). By applying the same methods to an FDM system the same form is shown to arise, but with the mass of the background particles replaced by  $m_{\text{eff}}$  of the order of the FDM mass enclosed by the de Broglie wavelength (cf. [61, 100]). This reflects the fact that the fluctuations are not due to discreteness noise as in the classical particle case (which is tiny in the case of a halo made of  $\sim 10^{100}$  ultra light axions), but due the interference patterns associated with the large de Broglie wavelength.

Diffusion coefficients can be derived and, following the formulation found in Binney & Tremaine [5] (Section 7.4), it is then possible to derive that fluctuation lead to increase with vertical velocity dispersion in galactic disks  $\sigma_z$ , which for Milky Way parameters can be estimated as

$$\sigma_z = 2.65 \text{ km s}^{-1} \left( \frac{10^{-22} \text{ eV}}{m_b} \right)^{3/2} \left( \frac{8 \text{ kpc}}{r} \right)^2 \left( \frac{T}{10 \text{ Gyr}} \right)^{1/2} \ln \Lambda^{1/2}. \quad (25)$$

For a Coulomb logarithm typically of order ten, comparison with current data shows that  $m_b$  must be of order  $10^{-22} \text{ eV}$  or more to obtain vertical velocity dispersion in line with what is observed the Milky Way disk ( $\sim 30 \text{ kpc}$  in the solar neighbourhood; see [103] for recent results); which is just consistent with requirements of solving the small scale problems of galaxy formation.

## 7. Concluding remarks

The current standard cosmological model is extremely successful in explaining the parameters and the large scale structure of our Universe. Its contents can be precisely determined. Yet this reveals that most of its mass and energy content is missing. Up to the turn of the century, the 'standard' cosmological model was the SCDM; a critical, flat universe dominated by cold dark matter, which was expected from beyond standard model particle theories such as supersymmetry. It even seemed that the relevant particles could be ('WIMP-miraculously') produced with just the right amount needed for the matter dominated critical-density SCDM, by simply assuming thermal production in an early Universe already known to have been a thermal state, as reflected in the Planckian CMB.

The discovery of cosmic acceleration and the implied 'dark energy' complicated matters; if dark energy arose from a cosmological constant associated with vacuum energy, there seemed to be no natural, unique, predictions for it within dozens of orders of magnitude [104]. Still, the simplest assumption was to invoke a cosmological constant and see how it could fit the observations. This is how SCDM became  $\Lambda$ CDM. The model that fits observations of the CMB and large scale structure exceptionally well, despite its the vast missing components. For a time it was almost invariably referred to as the concordance model, in reference to the various observations that lead to the same parameters.

But then problems with the CDM part gradually also arose. The many especially designed direct detection experiments have begun to rule out the WIMP miracle parameter space, and no trace of the WIMPs or the supersymmetric theories they often arise from came up at the gigantic machine largely designed to look for these — the Large Hadron Collider at CERN. Finally, all sorts of problems started transpiring from the astrophysical side, namely at (sub) galactic scales. These were discussed here in some detail, with some possible solutions outlined. Basically, those problems may arise from an incomplete knowledge of the highly complex baryonic physics of galaxy formation, or from modifications of the particle physics model. Another possibility, not discussed here, could invoke major modifications to physical law, e.g. gravitational interaction [105, 106]. In addition, there are some apparent faultlines in the paradigm, the seriousness of which remains to be seen. There is for example early galaxy and supermassive black hole formation [8, 9], that may be too early even for a CDM-based structure formation scenario, a phenomenon that would seem to be in contradiction with the small scale problem discussed here. There is also apparent tension between CMB-inferred and locally measured values of the Hubble constant. If real, solutions to this problem could in principle also involve modifications to gravitational law, quite different from those discussed above [107].

The coming period is an exciting one for research in galaxy formation and evolution and physical cosmology: the GAIA satellite is set to map the positions and velocities of up to a billion stars, implicitly probing the details of the dark matter distribution in the process. Satellites such as EUCLID will provide a wealth of information on fundamental physics, particularly the nature of the dark energy and whether it is indeed a cosmological constant, or instead varies with time [108]. The James Web Space Telescope and a new generation of earth based telescopes will map further and deeper the galaxy population and its distribution. The Square Kilometer Array will map the largest cosmological scales accessible to observation, and probe back into the era of formation of the first galaxies and stars. A new generation of CMB experiments will measure spectral distortions in the Planck spectrum (in order to probe the primordial spectrum at smaller scales) and look for polarisation signals of primordial gravitational waves, which would open a window into the the state of the Universe a tiny fraction of a second since the start of its expansion. Probes of gravitational waves are also being developed, opening a new window of exploration of high energy phenomena in the post-LIGO era.

These are just a few of the experimental efforts that should tell us more, in the coming

period, on the nature of dark matter and energy, and whether the solution to current problems in the astro-particle physics of galaxy formation is relatively mundane or of a more revolutionary nature.

### Acknowledgments

This project was supported financially by the Science and Technology Development Fund (STDF), Egypt. Grant No. 25859.

### References

- [1] Penzias A A and Wilson R W 1965 *Astrophys. Jour.* **142** 419–421
- [2] Weinberg S 1972 *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*
- [3] Durrer R 2015 *Classical and Quantum Gravity* **32** 124007 (*Preprint* 1506.01907)
- [4] Frenk C S and White S D M 2012 *Annalen der Physik* **524** 507–534 (*Preprint* 1210.0544)
- [5] Binney J and Tremaine S 2008 *Galactic Dynamics: Second Edition* (Princeton University Press)
- [6] Peacock J A 1999 *Cosmological Physics*
- [7] Dodelson S 2003 *Modern cosmology*
- [8] Steinhardt C L, Capak P, Masters D and Speagle J S 2016 *Astrophys. Jour.* **824** 21 (*Preprint* 1506.01377)
- [9] Wise J H, Regan J A, O’Shea B W, Norman M L, Downes T P and Xu H 2019 *arXiv e-prints* arXiv:1901.07563 (*Preprint* 1901.07563)
- [10] Cyburt R H, Fields B D, Olive K A and Yeh T H 2016 *Reviews of Modern Physics* **88** 015004 (*Preprint* 1505.01076)
- [11] Kolb E W and Turner M S 1990 *The early universe*.
- [12] Mo H, van den Bosch F C and White S 2010 “Galaxy Formation and Evolution”
- [13] Peacock J A 1999 *Cosmological Physics*
- [14] White S D M and Rees M J 1978 *Mon. Not. Roy. Astron. Soc.* **183** 341–358
- [15] Planck Collaboration, Aghanim N, Akrami Y, Ashdown M, Aumont J, Baccigalupi C, Ballardini M, Banday A J, Barreiro R B, Bartolo N, Basak S, Battye R, Benabed K, Bernard J P, Bersanelli M, Bielewicz P, Bock J J, Bond J R, Borrill J, Bouchet F R, Boulanger F, Bucher M, Burigana C, Butler R C, Calabrese E, Cardoso J F, Carron J, Challinor A, Chiang H C, Chluba J, Colombo L P L, Combet C, Contreras D, Crill B P, Cuttaia F, de Bernardis P, de Zotti G, Delabrouille J, Delouis J M, Di Valentino E, Diego J M, Doré O, Douspis M, Ducout A, Dupac X, Dusini S, Efstathiou G, Elsner F, Enßlin T A, Eriksen H K, Fantaye Y, Farhang M, Fergusson J, Fernandez-Cobos R, Finelli F, Forastieri F, Frailis M, Franceschi E, Frolov A, Galeotta S, Galli S, Ganga K, Génova-Santos R T, Gerbino M, Ghosh T, González-Nuevo J, Górski K M, Gratton S, Gruppuso A, Gudmundsson J E, Hamann J, Handley W, Herranz D, Hivon E, Huang Z, Jaffe A H, Jones W C, Karakci A, Keihänen E, Keskitalo R, Kiiveri K, Kim J, Kisner T S, Knox L, Krachmalnicoff N, Kunz M, Kurki-Suonio H, Lagache G, Lamarre J M, Lasenby A, Lattanzi M, Lawrence C R, Le Jeune M, Lemos P, Lesgourgues J, Levrier F, Lewis A, Liguori M, Lilje P B, Lilley M, Lindholm V, López-Caniego M, Lubin P M, Ma Y Z, Macías-Pérez J F, Maggio G, Maino D, Mandolesi N, Mangilli A, Marcos-Caballero A, Maris M, Martin P G, Martinelli M, Martínez-González E, Matarrese S, Mauri N, McEwen J D, Meinhold P R, Melchiorri A, Mennella A, Migliaccio M, Millea M, Mitra S, Miville-Deschénes M A, Molinari D, Montier L, Morgante G, Moss A, Natoli P, Nørgaard-Nielsen H U, Pagano L, Paoletti D, Partridge B, Patanchon G, Peiris H V, Perrotta F, Pettorino V, Piacentini F, Polastri L, Polenta G, Puget J L, Rachen J P, Reinecke M, Remazeilles M, Renzi A, Rocha G, Rosset C, Roudier G, Rubiño-Martín J A, Ruiz-Granados B, Salvati L, Sandri M, Savelainen M, Scott D, Shellard E P S, Sirignano C, Sirri G, Spencer L D, Sunyaev R, Suur-Uski A S, Tauber J A, Tavagnacco D, Tenti M, Toffolatti L, Tomasi M, Trombetti T, Valenziano L, Valiviita J, Van Tent B, Vibert L, Vielva P, Villa F, Vittorio N, Wandelt B D, Wehus I K, White M, White S D M, Zacchei A and Zonca A 2018 *arXiv e-prints* (*Preprint* 1807.06209)
- [16] Springel V 2016 *Saas-Fee Advanced Course* **43** 251 (*Preprint* 1412.5187)
- [17] Ludlow A D, Navarro J F, White S D M, Boylan-Kolchin M, Springel V, Jenkins A and Frenk C S 2011 *Mon. Not. Roy. Ast. Soc.* **415** 3895–3902 (*Preprint* 1102.0002)
- [18] Dubinski J and Carlberg R G 1991 *Astrophys. Jour.* **378** 496–503
- [19] Warren M S, Quinn P J, Salmon J K and Zurek W H 1992 *Astrophys. Jour.* **399** 405–425
- [20] Navarro J F, Frenk C S and White S D M 1996 *Astrophys. Jour.* **462** 563 (*Preprint* astro-ph/9508025)
- [21] Lynden-Bell D and Wood R 1968 *Mon. Not. Roy. Ast. Soc.* **138** 495
- [22] Lynden-Bell D 1967 *Mon. Not. Roy. Ast. Soc.* **136** 101

[23] Saslaw W C 1987 *Gravitational Physics of Stellar and Galactic Systems*

[24] El Zant A A 2013 *Astrophys. Jour.* **779** 64 (Preprint 1311.0129)

[25] Chandrasekhar S 1943 *Astrophys. Jour.* **97** 255

[26] Chandrasekhar S 1943 *Reviews of Modern Physics* **15** 1–89

[27] El-Zant A A 2008 *Astrophys. Jour.* **681** 1058–1075

[28] Moore B, Quinn T, Governato F, Stadel J and Lake G 1999 *Mon. Not. Roy. Ast. Soc.* **310** 1147–1152 (Preprint astro-ph/9903164)

[29] Huss A, Jain B and Steinmetz M 1999 *Astrophys. Jour.* **517** 64–69 (Preprint astro-ph/9803117)

[30] Lu Y, Mo H J, Katz N and Weinberg M D 2006 *Mon. Not. Roy. Ast. Soc.* **368** 1931–1940 (Preprint astro-ph/0508624)

[31] Shapiro P R, Iliev I T, Martel H, Ahn K and Alvarez M A 2004 *arXiv Astrophysics e-prints* (Preprint astro-ph/0409173)

[32] Gao L, White S D M, Jenkins A, Stoehr F and Springel V 2004 *Mon. Not. Roy. Ast. Soc.* **355** 819–834 (Preprint astro-ph/0404589)

[33] Boylan-Kolchin M, Bullock J S and Kaplinghat M 2011 *Mon. Not. Roy. Ast. Soc.* **415** L40–L44 (Preprint 1103.0007)

[34] Ogiya G and Burkert A 2015 *Mon. Not. Roy. Ast. Soc.* **446** 2363–2369 (Preprint 1408.6444)

[35] Flores R A and Primack J R 1994 *Astrophys. Jour. Lett.* **427** L1–L4 (Preprint astro-ph/9402004)

[36] Moore B 1994 *Nature* **370** 629–631

[37] McGaugh S S and de Blok W J G 1998 *Astrophys. Jour.* **499** 41–65 (Preprint astro-ph/9801123)

[38] Del Popolo A and Le Delliou M 2017 *Galaxies* **5** 17 (Preprint 1606.07790)

[39] Bullock J S and Boylan-Kolchin M 2017 *Annual Review of Astronomy and Astrophysics* **55** 343–387 (Preprint 1707.04256)

[40] Kamionkowski M and Liddle A R 2000 *Phys. Rev. Lett.* **84** 4525–4528 (Preprint astro-ph/9911103)

[41] Colín P, Avila-Reese V and Valenzuela O 2000 *Astrophys. Jour.* **542** 622–630 (Preprint astro-ph/0004115)

[42] Bode P, Ostriker J P and Turok N 2001 *Astrophys. Jour.* **556** 93–107 (Preprint astro-ph/0010389)

[43] Schneider A, Smith R E, Macciò A V and Moore B 2012 *Mon. Not. Roy. Ast. Soc.* **424** 684–698 (Preprint 1112.0330)

[44] Macciò A V, Paduroiu S, Anderhalden D, Schneider A and Moore B 2012 *Mon. Not. Roy. Ast. Soc.* **424** 1105–1112 (Preprint 1202.1282)

[45] Shao S, Gao L, Theuns T and Frenk C S 2013 *Mon. Not. Roy. Ast. Soc.* **430** 2346–2357 (Preprint 1209.5563)

[46] Lovell M R, Frenk C S, Eke V R, Jenkins A, Gao L and Theuns T 2014 *Mon. Not. Roy. Ast. Soc.* **439** 300–317 (Preprint 1308.1399)

[47] El-Zant A, Khalil S and Sil A 2015 *Phys. Rev. D* **91** 035030 (Preprint 1308.0836)

[48] Spergel D N and Steinhardt P J 2000 *Physical Review Letters* **84** 3760 (Preprint astro-ph/9909386)

[49] Burkert A 2000 *Astrophys. Jour.l* **534** L143–L146 (Preprint astro-ph/0002409)

[50] Kochanek C S and White M 2000 *Astrophys. Jour.* **543** 514–520 (Preprint astro-ph/0003483)

[51] Miralda-Escudé J 2002 *Astrophys. Jour.* **564** 60–64 (Preprint astro-ph/0002050)

[52] Peter A H G, Rocha M, Bullock J S and Kaplinghat M 2013 *Mon. Not. Roy. Ast. Soc.* **430** 105–120 (Preprint 1208.3026)

[53] Zavala J, Vogelsberger M and Walker M G 2013 *Mon. Not. Roy. Ast. Soc.* **431** L20–L24 (Preprint 1211.6426)

[54] Elbert O D, Bullock J S, Garrison-Kimmel S, Rocha M, Oñorbe J and Peter A H G 2015 *Mon. Not. Roy. Ast. Soc.* **453** 29–37 (Preprint 1412.1477)

[55] Goodman J 2000 *New Astronomy* **5** 103–107 (Preprint astro-ph/0003018)

[56] Hu W, Barkana R and Gruzinov A 2000 *Physical Review Letters* **85** 1158 (Preprint astro-ph/0003365)

[57] Destri C, De Vega H J and Sanchez N G 2013 *New Astronomy* **22** 39–50 (Preprint 1204.3090)

[58] Schive H Y, Liao M H, Woo T P, Wong S K, Chiueh T, Broadhurst T and Hwang W Y P 2014 *Physical Review Letters* **113** 261302 (Preprint 1407.7762)

[59] Marsh D J E and Silk J 2014 *Mon. Not. Roy. Ast. Soc.* **437** 2652–2663 (Preprint 1307.1705)

[60] Chavanis P H, Lemou M and Méhats F 2015 *Phys. Rev. D* **91** 063531 (Preprint 1409.7838)

[61] Hui L, Ostriker J P, Tremaine S and Witten E 2017 *Phys. Rev. D* **95** 043541 (Preprint 1610.08297)

[62] Chavanis P H 2018 *arXiv e-prints* (Preprint 1810.08948)

[63] Navarro J F, Eke V R and Frenk C S 1996 *Mon. Not. Roy. Ast. Soc.* **283** L72–L78 (Preprint astro-ph/9610187)

[64] El-Zant A, Shlosman I and Hoffman Y 2001 *Astrophys. Jour.* **560** 636–643 (Preprint astro-ph/0103386)

[65] El-Zant A A, Hoffman Y, Primack J, Combes F and Shlosman I 2004 *Astrophys. Jour.l* **607** L75–L78 (Preprint astro-ph/0309412)

[66] Read J I and Gilmore G 2005 *Mon. Not. Roy. Ast. Soc.* **356** 107–124 (Preprint astro-ph/0409565)

[67] Mashchenko S, Couchman H M P and Wadsley J 2006 *Nature* **442** 539–542 (Preprint [astro-ph/0605672](#))

[68] Mashchenko S, Wadsley J and Couchman H M P 2008 *Science* **319** 174– (Preprint [0711.4803](#))

[69] Peirani S, Kay S and Silk J 2008 *Atron. & Astrophys.* **479** 123–129 (Preprint [astro-ph/0612468](#))

[70] Pontzen A and Governato F 2012 *Mon. Not. Roy. Ast. Soc.* **421** 3464–3471 (Preprint [1106.0499](#))

[71] Governato F, Zolotov A, Pontzen A, Christensen C, Oh S H, Brooks A M, Quinn T, Shen S and Wadsley J 2012 *Mon. Not. Roy. Ast. Soc.* **422** 1231–1240 (Preprint [1202.0554](#))

[72] Zolotov A, Brooks A M, Willman B, Governato F, Pontzen A, Christensen C, Dekel A, Quinn T, Shen S and Wadsley J 2012 *Astrophys. Jour.* **761** 71 (Preprint [1207.0007](#))

[73] Martizzi D, Teyssier R and Moore B 2013 *Mon. Not. Roy. Ast. Soc.* **432** 1947–1954 (Preprint [1211.2648](#))

[74] Teyssier R, Pontzen A, Dubois Y and Read J I 2013 *Mon. Not. Roy. Ast. Soc.* **429** 3068–3078 (Preprint [1206.4895](#))

[75] Pontzen A and Governato F 2014 *Nature* **506** 171–178 (Preprint [1402.1764](#))

[76] Madau P, Shen S and Governato F 2014 *Astrophys. Jour.l* **789** L17 (Preprint [1405.2577](#))

[77] Ogiya G and Mori M 2014 *Astrophys. Jour.* **793** 46 (Preprint [1206.5412](#))

[78] Jardel J R and Sellwood J A 2009 *Astrophys. Jour.* **691** 1300–1306 (Preprint [0808.3449](#))

[79] Romano-Díaz E, Shlosman I, Hoffman Y and Heller C 2008 *Astrophys. Jour.l* **685** L105–L108 (Preprint [0808.0195](#))

[80] Goerdt T, Moore B, Read J I and Stadel J 2010 *Astrophys. Jour.* **725** 1707–1716 (Preprint [0806.1951](#))

[81] Cole D R, Dehnen W and Wilkinson M I 2011 *Mon. Not. Roy. Ast. Soc.* **416** 1118–1134 (Preprint [1105.4050](#))

[82] Del Popolo A, Lima J A S, Fabris J C and Rodrigues D C 2014 *Jour. Cosmology Astropart. Phys.* **4** 021 (Preprint [1404.3674](#))

[83] Laporte C F P and White S D M 2015 *Mon. Not. Roy. Ast. Soc.* **451** 1177–1189 (Preprint [1409.1924](#))

[84] Del Popolo A, Pace F, Le Delliou M and Lee X 2018 *Phys. Rev. D* **98** 063517 (Preprint [1809.10609](#))

[85] Tonini C, Lapi A and Salucci P 2006 *Astrophys. Jour.* **649** 591–598 (Preprint [astro-ph/0603051](#))

[86] Inoue S and Saitoh T R 2011 *Mon. Not. Roy. Ast. Soc.* **418** 2527–2531 (Preprint [1108.0906](#))

[87] Nipoti C and Binney J 2015 *Mon. Not. Roy. Ast. Soc.* **446** 1820–1828 (Preprint [1410.6169](#))

[88] El-Zant A A, Freundlich J and Combes F 2016 *Mon. Not. Roy. Astron. Soc.* **461** 1745–1759 (Preprint [1603.00526](#))

[89] Somerville R S and Davé R 2015 *Annual Review of Astronomy and Astrophysics* **53** 51–113 (Preprint [1412.2712](#))

[90] Bose S, Frenk C S, Jenkins A, Fattahi A, Gomez F A, Grand R J J, Marinacci F, Navarro J F, Oman K A, Pakmor R, Schaye J, Simpson C M and Springel V 2018 *arXiv e-prints* (Preprint [1810.03635](#))

[91] Bose S, Frenk C S, Jenkins A, Fattahi A, Gomez F A, Grand R J J, Marinacci F, Navarro J F, Oman K A, Pakmor R, Schaye J, Simpson C M and Springel V 2018 *arXiv e-prints* (Preprint [1810.03635](#))

[92] Dutton A A, Macciò A V, Buck T, Dixon K L, Blank M and Obreja A 2018 *arXiv e-prints* arXiv:1811.10625 (Preprint [1811.10625](#))

[93] Marsh D J E 2016 *Phys. Rep.* **643** 1–79 (Preprint [1510.07633](#))

[94] Feynman R P 1964 *Feynman lectures on physics. Volume 3: Quantum mechanics*

[95] Mocz P, Lancaster L, Fialkov A, Becerra F and Chavanis P H 2018 *Phys. Rev. D* **97** 083519 (Preprint [1801.03507](#))

[96] Bar N, Blas D, Blum K and Sibiryakov S 2018 *Phys. Rev. D* **98** 083027 (Preprint [1805.00122](#))

[97] Deng H, Hertzberg M P, Namjoo M H and Masoumi A 2018 *Phys. Rev. D* **98** 023513 (Preprint [1804.05921](#))

[98] Kobayashi T, Murgia R, De Simone A, Iršič V and Viel M 2017 *Phys. Rev. D* **96** 123514 (Preprint [1708.00015](#))

[99] Nebrin O, Ghara R and Mellema G 2018 *arXiv e-prints* arXiv:1812.09760 (Preprint [1812.09760](#))

[100] Bar-Or B, Fouvry J B and Tremaine S 2019 *Astrophys. Jour.* **871** 28 (Preprint [1809.07673](#))

[101] Marsh D J E and Niemeyer J C 2018 *arXiv e-prints* arXiv:1810.08543 (Preprint [1810.08543](#))

[102] Church B V, Ostriker J P and Mocz P 2018 *arXiv e-prints* arXiv:1809.04744 (Preprint [1809.04744](#))

[103] Mackereth J T, Bovy J, Leung H W, Schiavon R P, Trick W H, Chaplin W J, Cunha K, Feuillet D K, Majewski S R, Martig M, Miglio A, Nidever D, Pinsonneault M H, Silva Aguirre V, Sobeck J, Tayar J and Zasowski G 2019 *arXiv e-prints* arXiv:1901.04502 (Preprint [1901.04502](#))

[104] Weinberg S 1989 *Reviews of Modern Physics* **61** 1–23

[105] Milgrom M 1983 *Astrophys. Jour.* **270** 365

[106] Verlinde E 2017 *SciPost Physics* **2** 016 (Preprint [1611.02269](#))

[107] El-Zant A, Hanafy W E and Elgammal S 2018 *arXiv e-prints* arXiv:1809.09390 (Preprint [1809.09390](#))

[108] Amendola L, Appleby S, Avgoustidis A, Bacon D, Baker T, Baldi M, Bartolo N, Blanchard A, Bonvin C, Borgani S, Branchini E, Burrage C, Camera S, Carbone C, Casarini L, Cropper M, de Rham C, Dietrich J P, Di Porto C, Durrer R, Ealet A, Ferreira P G, Finelli F, García-Bellido J, Giannantonio T, Guzzo L, Heavens A, Heisenberg L, Heymans C, Hoekstra H, Hollenstein L, Holmes R, Hwang Z, Jahnke K,

Kitching T D, Koivisto T, Kunz M, La Vacca G, Linder E, March M, Marra V, Martins C, Majerotto E, Markovic D, Marsh D, Marulli F, Massey R, Mellier Y, Montanari F, Mota D F, Nunes N J, Percival W, Pettorino V, Porciani C, Quercellini C, Read J, Rinaldi M, Sapone D, Sawicki I, Scaramella R, Skordis C, Simpson F, Taylor A, Thomas S, Trotta R, Verde L, Vernizzi F, Vollmer A, Wang Y, Weller J and Zlosnik T 2018 *Living Reviews in Relativity* **21** 2 (*Preprint* 1606.00180)