The physical bases for the structures of collisionless dark matter halos

M. Sten Delos*
Max Planck Institute for Astrophysics, Karl-Schwarzschild-Straße 1, 85748 Garching, Germany

Bound, virialized halos of dark matter are the basic units of nonlinear structure in our Universe. The detailed structures of these objects are set by the cosmological initial conditions through collisionless gravitational dynamics. These notes present a review of this mechanism, focusing on halo density profiles.

I. BACKGROUND

Dark matter comprises about 5/6 of the matter in the Universe. Consequently, it drives the formation of cosmic structures. Dark matter forms halos of orbiting material that are bound together by their own gravity. Galaxies form at the centers of these halos, and halos are also expected to exist at scales too small to host galaxies. These notes discuss the structures of dark matter halos and the physical principles that set these structures.

In the standard paradigm of collisionless particle dark matter, two-body energy exchange occurs at such a low rate that dark matter systems cannot thermalize. Consequently, thermodynamic principles such as entropy maximization are not useful.[1] Instead, halo properties remain tightly linked to the details of the cosmological initial conditions.

The main halo property that I consider is the spherically averaged density profile, $\rho(r)$. Simulations of our concordance cosmological model have long suggested that density profiles have universal forms. The Navarro-Frenk-White form,

$$\rho(r) = \frac{\rho_\mathrm{s}}{r/r_\mathrm{s}(1+r/r_\mathrm{s})^2} \tag{1}$$

(Navarro *et al.* 1996, 1997), is a common parametrization. Its logarithmic slope $\mathrm{d}\log\rho/\mathrm{d}\log r$ drops from -1 when $r \ll r_\mathrm{s}$ to -3 when $r \gg r_\mathrm{s}$. Importantly, the only freedom in this functional form is an overall rescaling of the size (through the scale radius r_s) or the density (through the scale density ρ_s). A better-fitting parametrization is the Einasto form,

$$\rho(r) = \rho_\mathrm{s} \exp\left[-\frac{2}{\alpha}\left(\frac{r}{r_\mathrm{s}}\right)^\alpha\right] \tag{2}$$

(Einasto 1965). Its logarithmic slope drops continuously from 0 to $-\infty$ as radius increases, with $\mathrm{d}\log\rho/\mathrm{d}\log r = -2$ at $r = r_\mathrm{s}$. This function has an additional parameter α, but it is typically found that α can be fixed to a constant value in the range of 0.16 to 0.18 (e.g. Navarro *et al.* 2004, 2010, Wang *et al.* 2020, Diemer 2023).

The apparent universality of halo density profiles seems to suggest that despite the absence of collisional relaxation, some phenomenon (e.g. violent relaxation; Lynden-Bell 1967) still causes halos to relax toward a universal form. One of the conclusions of these notes will be that there is no basis for this idea. Halo structures are set by the initial conditions, and if they are universal, this must derive from universal features of the initial conditions that give rise to the halos.

The cosmological initial conditions are specified by the linear density field, which I discuss in section II. A simple, but surprisingly powerful, connection between the linear density field and nonlinear collapsed structures is supplied by the spherical collapse model, which I review in section III. In section IV, I show how the cosmological initial conditions determine halo accretion histories and density profiles, and particularly how profiles resembling equations (1) and (2) arise. Finally, the structure at the very center of a halo is set by the initial moment of halo formation, and I discuss the mechanism and its outcome in section V.

II. THE INITIAL DENSITY FIELD AND LINEAR THEORY

The cosmological initial conditions are specified by the field $\delta(\boldsymbol{x}, a)$ describing the fractional density contrast, i.e., $\delta \equiv (\rho - \bar{\rho})/\bar{\rho}$. Here, $\boldsymbol{x}$ is the comoving position coordinate and a is the cosmic expansion factor. While a general

* sten@mpa-garching.mpg.de

[1] Under Newtonian gravity, the configuration of locally maximal entropy is the isothermal sphere, for which the density scales with radius as $\rho \propto r^{-2}$. Indeed, gravitating systems in which two-body energy exchange is possible, such as halos of self-interacting dark matter or globular star clusters, settle into the isothermal sphere configuration after sufficiently long times (after the so-called core collapse), although for star clusters, few-body dynamics can complicate this conclusion.

collisionless system requires that the full distribution function $f(\boldsymbol{x}, \boldsymbol{v})$ be specified, particle velocities at early times are almost completely determined by the density field. Peculiar velocities associated with the mass distribution grow as $v \propto a^{1/2}$ during the matter dominated era, while other peculiar velocities, such as random thermal motion, decay due to cosmic expansion as $v \propto a^{-1}$. Thus, motion that is not associated with $\delta(\boldsymbol{x}, a)$ can be largely neglected.

When $|\delta| \ll 1$, the evolution $\delta(a)$ can be evaluated analytically. Neglecting terms of order δ^2 or higher, the continuity, Euler, and Poisson equations imply that

$$\frac{\partial^2 \delta}{\partial t^2} + 2H\frac{\partial \delta}{\partial t} = 4\pi G\bar{\rho}\delta, \tag{3}$$

where H is the Hubble rate. During the matter-dominated era, $H^2 = (8\pi/3)G\bar{\rho}$ and $a \propto t^{2/3}$, and it can be straightforwardly verified that the two solutions to equation (3) are $\delta \propto a$ and $\delta \propto a^{-3/2}$. Since the latter solution decays, density perturbations relevant to late-time structure grow as $\delta \propto a$.

Equation (3) describes the linear theory of cosmological perturbations, since again, terms quadratic or higher in δ are neglected. Linear theory is surprisingly versatile, having significant utility even when $|\delta| \ll 1$ is not satisfied. For this reason, it is often useful to redefine δ to be the density contrast extrapolated using linear theory, rather than the actual physical density contrast $(\rho - \bar{\rho})/\bar{\rho}$. The versatility of linear theory comes from its connection to spherical collapse models, which I discuss next.

III. SPHERICAL COLLAPSE

Spherical collapse models present a simplified, yet very powerful, connection between the cosmological initial conditions and the halos that arise at much later times. Consider a spherical shell initially of radius r_i that encloses an average density contrast $\delta_\mathrm{i} \ll 1$ at some early time a_i. Spherical collapse theory deals with the trajectory $r(t)$ of the material comprising this shell. If the Universe is flat and matter dominated, and the shell encloses an overdense region $\delta_\mathrm{i} > 0$, then it will initially expand with the Hubble flow but eventually turn around and collapse. Aside from spherical symmetry, the main assumption in a spherical collapse model is that no material crosses the shell under consideration. This implies that it encloses a constant mass

$$M = \frac{4\pi}{3}\bar{\rho}_0(1+\delta_\mathrm{i})\left(\frac{r_\mathrm{i}}{a_\mathrm{i}}\right)^3 = \frac{H_\mathrm{i}^2}{2G}(1+\delta_\mathrm{i})r_\mathrm{i}^3, \tag{4}$$

where $\bar{\rho}_0 = \bar{\rho}(a)a^3$ is the mean comoving density and $H_\mathrm{i} = H(a_\mathrm{i}) = \sqrt{(8\pi/3)G\bar{\rho}_0 a_\mathrm{i}^{-3}}$ is the initial Hubble rate.

I will assume matter domination for simplicity, although spherical collapse has been considered in other contexts as well. The initial velocity of the shell is specified by requiring that it match the linear theory prediction $\delta = (\delta_\mathrm{i}/a_\mathrm{i})a$, which implies $\mathrm{d}\delta/\mathrm{d}t = (\delta_\mathrm{i}/a_\mathrm{i})aH = H_\mathrm{i}\delta_\mathrm{i}$, where the latter equality is at the initial time. The density contrast within our spherical shell is

$$\delta_\mathrm{sc} = (1+\delta_\mathrm{i})\left(\frac{ar_\mathrm{i}}{a_\mathrm{i}r}\right)^3 - 1 \tag{5}$$

as a function of time t, $a = a(t)$, and $r = r(t)$. At the initial time, $\mathrm{d}\delta_\mathrm{sc}/\mathrm{d}t = (1+\delta_\mathrm{i})\left[3H_\mathrm{i} - 3v_\mathrm{i}/r_\mathrm{i}\right] \simeq 3[H_\mathrm{i} - v_\mathrm{i}/r_\mathrm{i}]$ if $\delta_\mathrm{i} \ll 1$. Matching this to the linear theory prediction yields

$$v_\mathrm{i} = H_\mathrm{i}r_\mathrm{i}(1 - \delta_\mathrm{i}/3). \tag{6}$$

The radius of the mass shell evolves straightforwardly under Newtonian gravity, $\mathrm{d}^2r/\mathrm{d}t^2 = -GM/r^2$, where M is given by equation (4). Its energy (per mass) is

$$E = \frac{v_\mathrm{i}^2}{2} - \frac{GM}{r_\mathrm{i}} = -\frac{5}{6}\delta_\mathrm{i}H_\mathrm{i}^2 r_\mathrm{i}^2; \tag{7}$$

the latter equality follows from substituting equations (4) and (6). The apocenter radius, often called the turnaround radius, is thus

$$r_\mathrm{ta} = -\frac{GM}{E} = \frac{3}{5}\frac{r_\mathrm{i}}{\delta_\mathrm{i}} \tag{8}$$

to linear order in δ_i. The turnaround time is half the Keplerian orbital period for a semimajor axis of $r_\mathrm{ta}/2$, i.e.,

$$t_\mathrm{ta} = T_\mathrm{orbit}/2 = \pi\sqrt{\frac{(r_\mathrm{ta}/2)^3}{GM}} = \frac{\pi}{2}\left(\frac{3}{5}\right)^{3/2}\delta_\mathrm{i}^{-3/2}H_\mathrm{i}^{-1} = \frac{3\pi}{4}\left(\frac{3}{5}\right)^{3/2}\delta_\mathrm{i}^{-3/2}t_\mathrm{i} \tag{9}$$

again to linear order in δ_i, where we have used $H_\mathrm{i} = 2/(3t_\mathrm{i})$.

The connection to linear theory develops as follows. By the time t_ta, the Universe has expanded by the factor $a_\mathrm{ta}/a_\mathrm{i} = (t_\mathrm{ta}/t_\mathrm{i})^{2/3}$. The linear-theory extrapolation of δ has grown by the same factor, so that the linear δ at turnaround is

$$\delta_\mathrm{ta} = (t_\mathrm{ta}/t_\mathrm{i})^{2/3}\delta_\mathrm{i} = \frac{3}{5}\left(\frac{3\pi}{4}\right)^{2/3} \simeq 1.062. \tag{10}$$

In terms of the linearly extrapolated $\delta(a)$, it follows from equation (8) that the turnaround radius is

$$r_\mathrm{ta} = \frac{3}{5}\frac{q}{\delta(a)/a}, \tag{11}$$

where $q = r_\mathrm{i}/a_\mathrm{i}$ is the comoving radius of the mass shell in the initial conditions. Note that $\delta(a)/a$ is constant. At turnaround, the shell encloses density

$$\frac{\rho_\mathrm{ta}}{\bar\rho} = \left(\frac{a_\mathrm{ta}r_\mathrm{i}}{a_\mathrm{i}r_\mathrm{ta}}\right)^3 = \left(\frac{3\pi}{4}\right)^2 \simeq 5.55 \tag{12}$$

(see equation 5).

When the shell collapses to the radius $r = r_\mathrm{ta}/2$, its total energy $-GM/r_\mathrm{ta}$ is half of its potential energy $-GM/r$. This is precisely the ratio predicted by the virial theorem, which is valid for stable systems of orbiting particles, such as halos. Consequently, it can be expected that the shell becomes part of a halo at around this time, although this consideration is not exact, since the virial theorem relates to the energy of the whole system and not individual shells. The time of this so-called virialization can be evaluated by integrating $(\mathrm{d}r/\mathrm{d}t)^{-1} = [2(E + GM/r)]^{-1/2} = [2(GM/r - GM/r_\mathrm{ta})]^{-1/2}$ from $r_\mathrm{ta}/2$ to r_ta, yielding (with the substitution $x = r/r_\mathrm{ta}$)

$$t_\mathrm{vir} = t_\mathrm{ta} + \sqrt{\frac{r_\mathrm{ta}^3}{2GM}}\int_{1/2}^1 \frac{\mathrm{d}x}{\sqrt{1/x - 1}} = \left(\frac{3}{2} + \frac{1}{\pi}\right)t_\mathrm{ta} = \left(\frac{3}{4} + \frac{9\pi}{8}\right)\left(\frac{3}{5}\right)^{3/2}\delta_\mathrm{i}^{-3/2}t_\mathrm{i} \tag{13}$$

since the integral evaluates to $(2+\pi)/4$ and its prefactor is just $(2/\pi)t_\mathrm{ta}$ (see equation 9). The linear density contrast at virialization is thus

$$\delta_\mathrm{vir} = \left(\frac{3}{2} + \frac{1}{\pi}\right)^{2/3}\delta_\mathrm{ta} = \frac{3}{5}\left(\frac{3}{4} + \frac{9\pi}{8}\right)^{2/3} \simeq 1.583, \tag{14}$$

and the density at virialization is

$$\frac{\rho_\mathrm{vir}}{\bar\rho} = 2^3\left(\frac{3}{2} + \frac{1}{\pi}\right)^2\frac{\rho_\mathrm{ta}}{\bar\rho} = \frac{9}{8}(2 + 3\pi)^2 \simeq 147, \tag{15}$$

where the factor of 2^3 arises because of the increase in the shell's enclosed density while the factor of $(3/2 + 1/\pi)^2$ is associated with the decrease in the average density, both in relation to the turnaround time. Thus, we expect that halos are around 150 times denser than the cosmological average.[2]

Finally, the shell collapses to radius 0 at the time $t_\mathrm{coll} = 2t_\mathrm{ta}$, assuming that no shell crossing happens before then. The linear δ at collapse thus takes the well known value

$$\delta_\mathrm{coll} = 2^{2/3}\delta_\mathrm{ta} = \frac{3}{5}\left(\frac{3\pi}{2}\right)^{2/3} \simeq 1.686. \tag{16}$$

Figure 1 summarizes the spherical collapse model.

Linear δ thresholds for halo formation, whether δ_coll, δ_vir, or ellipsoidal generalizations thereof, are extremely useful when evaluating halo populations. If the average linearly extrapolated density contrast within a spherical region of mass M is equal to the threshold at some time, then under the approximation of spherical collapse, the material comprises a halo of mass M at that time. This is the basis for excursion set halo mass functions (Bond *et al.* 1991), a rich topic that is beyond the scope of these notes. I will show next what spherical collapse says about individual halos.

---

[2] The standard virial overdensity is $\rho_\mathrm{vir}/\bar\rho(a_\mathrm{coll}) = 18\pi^2 \simeq 178$, the ratio between the shell's enclosed density at the virialization time to the density of the universe at the later collapse time (not the virialization time). I see no justification for comparing these quantities.

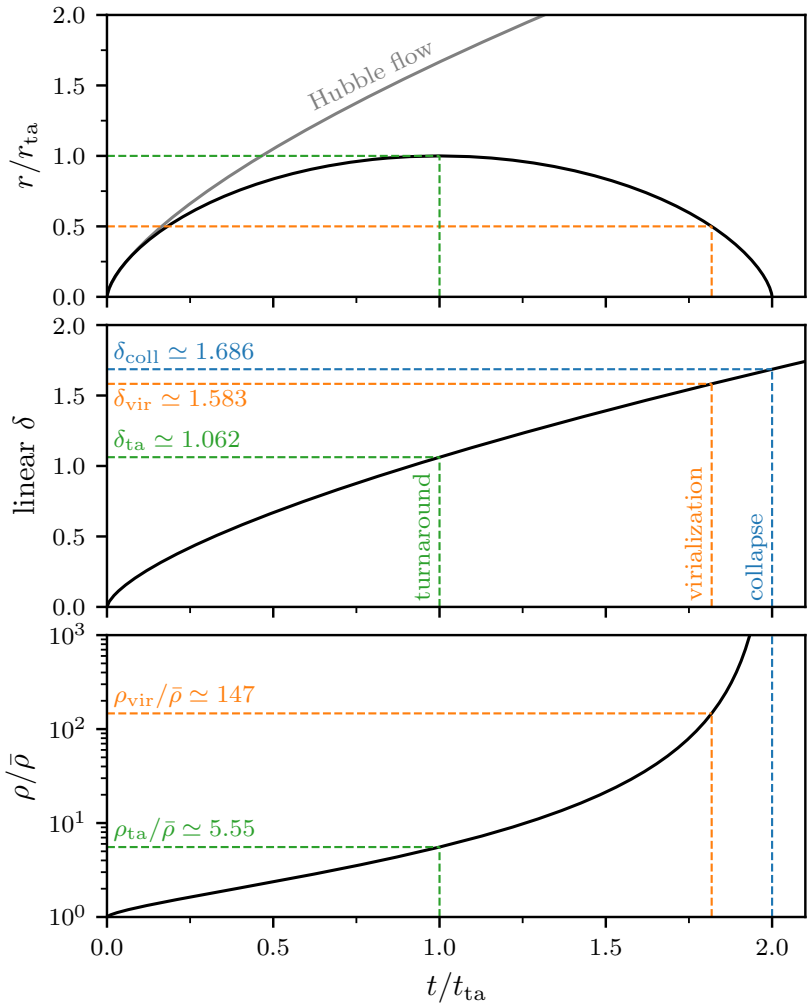

FIG. 1. Summary of the spherical collapse model in a flat matter-dominated universe. From top to bottom, the panels show a collapsing shell's radius, its linearly extrapolated density contrast, and its enclosed density. The dashed lines indicate turnaround (green), virialization (orange), and collapse (blue). The upper panel also shows the trajectory of the Hubble flow in gray.

IV. DENSITY PROFILES AND ACCRETION HISTORIES

Let us explore what spherical collapse can say about a halo's density profile, its accretion history, and the connection between the two. Instead of considering a single shell, let $\delta(q, a)$ be the average linear density contrast enclosed within the comoving radius q. We will imagine that every shell undergoes spherical collapse according to its respective δ. We continue to assume matter domination, so let us write for convenience

$$\tilde{\delta} \equiv \delta/a, \tag{17}$$

which is constant in time.

A. Mass accretion histories from spherical collapse

According to spherical collapse, a mass shell at the initial comoving radius q becomes part of the central halo when $\tilde{\delta}(q) = \delta_{\rm vir}/a$, i.e.

$$a_{\rm vir}(q) = \delta_{\rm vir}/\tilde{\delta}(q). \tag{18}$$

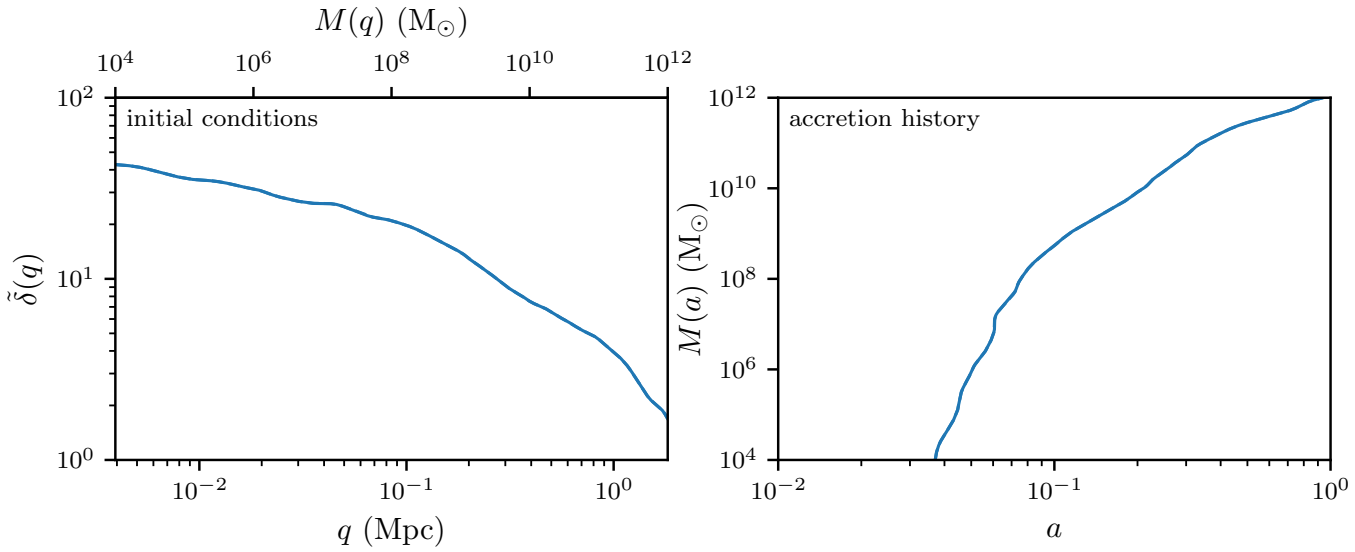

FIG. 2. Example initial conditions $\tilde{\delta}(q)$, on the left, and the resulting halo mass accretion history, on the right. The initial conditions are drawn from an Eisenstein and Hu (1998) power spectrum with Planck Collaboration (2020) cosmological parameters. The accretion rate Γ is the slope of the accretion history curve; it is high at early times and lower at late times.

But the mass enclosed within the comoving radius q in the initial conditions is

$$M(q) = \frac{4\pi}{3}\bar{\rho}_0 q^3 \tag{19}$$

(to zeroth order in δ). If the shell at q is part of the halo, then so are all shells initially below q, which would have accreted onto the halo earlier (at least if $\tilde{\delta}(q)$ decreases monotonically). Thus, the entire mass $M(q)$ belongs to the halo, so equations (18) and (19) together specify the mass accretion history $M(a)$ of the halo. Figure 2 shows an example $\tilde{\delta}(q)$, drawn from a realistic set of initial conditions for our Universe, along with the halo mass accretion history $M(a)$ that it predicts.

Of particular interest is the logarithmic halo accretion rate,

$$\Gamma(a) \equiv \frac{\mathrm{d}\log M}{\mathrm{d}\log a} = 3\left(\frac{\mathrm{d}\log a_\mathrm{vir}}{\mathrm{d}\log q}\right)^{-1} = -3\left(\frac{\mathrm{d}\log\tilde{\delta}}{\mathrm{d}\log q}\right)^{-1}. \tag{20}$$

The latter expressions follow from equations (18) and (19) and are to be evaluated at the q that satisfies $a = \delta_\mathrm{vir}/\tilde{\delta}(q)$. Evidently, at least under the spherical collapse approximation, a halo's accretion rate is directly connected to the radial gradient of the density contrast in the initial conditions.

B. Halo merging

If $\tilde{\delta}(q)$ decreases monotonically as a function of q, this implies that lower shells are attracted by a larger density excess and hence collapse earlier. This is the case in the example in figure 2. However, if $\tilde{\delta}(q)$ is increasing for any q, then there is shell crossing prior to collapse, as larger shells are attracted more strongly than smaller shells. Since $\bar{\delta}(q)$ is the average density contrast enclosed within q, for $\tilde{\delta}(q)$ to be increasing demands that the density contrast *at* q be quite high. For realistic initial conditions that are not spherically symmetric, this typically corresponds not to an overdense spherical shell of material but to high overdensity that is offset from $q = 0$. That overdensity would have formed its own halo, and its accretion corresponds to a halo merger.

Spherical collapse clearly breaks down in such a picture. However, we can still gain some insight by applying it (e.g. Angulo and White 2010). Figure 3 shows initial conditions that have increasing $\tilde{\delta}(q)$ at some q. The accretion rate that is naively predicted from equation (18) is multiple-valued. However, let us assume that if more distant material is predicted to collapse before nearer material, it pulls the nearer material with it. This is approximately expected in a halo merger, since the merging halo would accrete those surroundings. This assumption amounts to making $\tilde{\delta}(q)$

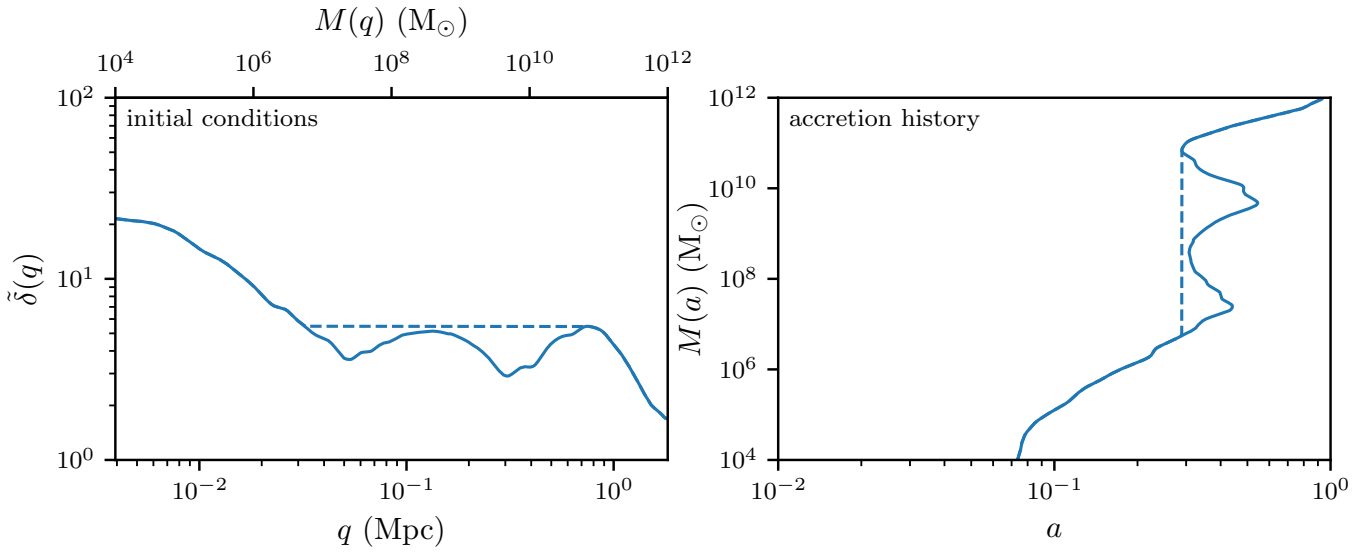

FIG. 3. Similar to figure 2 but showing an instance where $\tilde{\delta}(q)$ is not monotonically decreasing. Indeed, this is a more typical example of the spherically averaged initial density field about an arbitrary point. Equation (18) predicts that more distant material accretes before nearer material, but if we assume that the distant material pulls the nearer material with it, then the halo mass jumps discontinuously (dashed lines). We may interpret this jump as a merger with another halo.

monotonic, as indicated by the dashed lines in figure 3. The accretion history is now single-valued, and it exhibits a sharp jump associated with the halo merger.

If we plot $\tilde{\delta}(q)$ about arbitrary points in a random field, the results will usually look more like figure 3 than figure 2. This is a statement that halo mergers are common; if we pick an arbitrary dark matter particle, it likely undergoes one or more cycles of merging onto a much larger halo. However, if we are interested in a given halo's past (and not future) accretion history, we normally want to follow the most massive progenitor halo. This means we specifically select points in the initial conditions about which $\tilde{\delta}(q)$ is monotonically decreasing or nearly so. For example, if we take the jumps in $M(a)$ exemplified in figure 3 at face value, then we should select initial points for which the predicted $M(a)$ never jumps by more than a factor of 2. That way, our selected point represents a dark matter particle that is part of the halo's most massive progenitor at every time.

Minimization of jumps in $M(a)$ was indeed how I selected the initial point for figure 2. Moreover, this biased selection explains why $M(a)$ behaves as it does in figure 2. The accretion rate Γ (the logarithmic slope of $M(a)$) is high at early times, because we have specifically chosen a halo that accretes many other halos – as opposed, for example, to choosing one of those other halos. Γ is typically (but not necessarily) lower at late times for the simple reason that most halos have a lower accretion rate, and our biased selection does not bias this regime as significantly. This bias associated with the hierarchical clustering of halos predicts that halo mass accretion histories should typically resemble that in figure 2, with rapid accretion early on and slow accretion at more recent times. That is indeed a general trend (e.g. Ludlow *et al.* 2013).

C. Density profiles from spherical collapse

The spherical collapse model can also predict the internal structures of halos in the following way (e.g. Dalal *et al.* 2010, Delos *et al.* 2019). By equation (11), the turnaround radius for each shell is

$$r_{\rm ta}(q) = \frac{3}{5}\frac{q}{\tilde{\delta}(q)}. \tag{21}$$

To proceed from here to halo structures, we approximate that within the halo, each particle orbits stably with time-averaged radius proportional to its original turnaround radius. That is, a shell's time-averaged final radius is

$$r(q) = \beta r_{\rm ta}(q) = \frac{3\beta}{5}\frac{q}{\tilde{\delta}(q)} \tag{22}$$

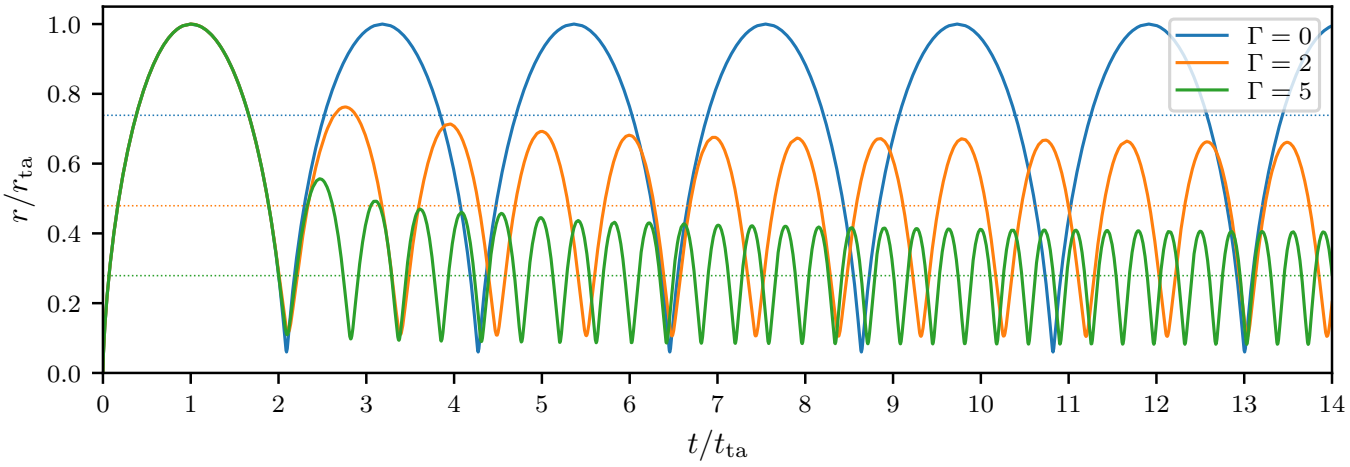

FIG. 4. Trajectories of spherical shells before ($t/t_{\rm ta} \lesssim 2$) and after ($t/t_{\rm ta} \gtrsim 2$) collapse, for halos with several different accretion rates $\Gamma = {\rm d}\log M/{\rm d}\log a$. More technically, the $\Gamma = 2$ and $\Gamma = 5$ cases are self-similar models (appendix B) with $\delta(q) \propto q^{-3/\Gamma}$, while the $\Gamma = 0$ case is Keplerian motion (the enclosed mass is constant). I inject a modest amount of angular momentum at turnaround in order to avoid pathological behavior that arises when all material is enforced to cross $r = 0$. This angular momentum is why the orbital period in the $\Gamma = 0$ case is not exactly $2t_{\rm ta}$. The general behavior is that the orbit contracts after collapse if $\Gamma > 0$, but it quickly stabilizes such that after a few orbits, the apocenter is static. The dotted lines mark for each case the time-averaged orbital radius after orbit stabilization.

for some number β. If there were negligible shell crossing, it can be shown straightforwardly that each shell's time-averaged orbital radius would be $r(q) = (3/4)r_{\rm ta}(q)$, i.e. $\beta = 3/4$. Equations (19) and (22) together predict a halo's mass profile $M(r)$.[3]

Figure 4 illustrates the ideas underlying equation (22). A mass shell initially follows the spherical collapse trajectory ($t/t_{\rm ta} \lesssim 2$). After collapse ($t/t_{\rm ta} \gtrsim 2$), it settles into a stable orbit with a constant time-averaged radius (dotted line). However, in general, the final orbit has an apocenter below $r_{\rm ta}$. This behavior arises due to shell crossing. During and after the expansion stage after collapse, a shell crosses infalling shells with larger q, so the enclosed mass M begins to grow. This causes the shell's orbit to contract, so in general we expect $\beta < 3/4$. As figure 4 demonstrates, the degree of contraction depends on the accretion rate Γ, because a higher Γ leads the enclosed mass to grow more rapidly. This means that β should in principle be a function of Γ (equation 20). Let us neglect this effect, however, since it significantly complicates the theory. I present a brief discussion of orbital contraction in appendix A.

As an example, let us consider the logarithmic slope of the mass profile $M(r)$ that results from these considerations. It is

$$\frac{{\rm d}\log M}{{\rm d}\log r} = \frac{{\rm d}\log M/{\rm d}\log q}{{\rm d}\log r/{\rm d}\log q} = \frac{3}{1 - {\rm d}\log\tilde{\delta}/{\rm d}\log q} = \frac{3}{1 + 3/\Gamma}, \tag{23}$$

where Γ, defined in equation (20), is to be evaluated when the shell at r accreted onto the halo, i.e. at the scale factor a related to r through equations (18) and (22). Slow accretion (low Γ) produces a compact structure, in which the mass does not grow strongly with radius, while fast accretion (high Γ) produces a more extended structure. The logarithmic slope of the density profile is

$$\begin{aligned}\frac{{\rm d}\log\rho}{{\rm d}\log r} &= \frac{{\rm d}\log M}{{\rm d}\log r} - 3 + \frac{\rm d}{{\rm d}\log r}\log\left(\frac{{\rm d}\log M}{{\rm d}\log r}\right) = \frac{-9}{\Gamma + 3} - \left(\frac{{\rm d}\log r}{{\rm d}\log q}\right)^{-1}\frac{{\rm d}\log(1 + 3/\Gamma)}{{\rm d}\log q} \\ &= \frac{-9}{\Gamma + 3} + \frac{3}{(1 + 3/\Gamma)^2}\frac{{\rm d}\Gamma}{{\rm d}\log a}.\end{aligned} \tag{24}$$

The second term indicates that the density slope is sensitive to the rate of change of the accretion rate Γ. However, orbital contraction (appendix A) introduces further sensitivity to the rate of change of Γ, so discussing it in the

[3] To be more precise, we should consider not just the average radius of a mass shell but the full distribution of radii that it covers. A shell contributes density toward all radii lying between its pericenter and apocenter, in the average. I neglect this effect, but see Dalal *et al.* (2010) for a discussion.

context of our simplified treatment would miss the full story. Let us focus instead on the first term in equation (24). Due to this term, a higher accretion rate Γ yields a shallower density slope, while a lower accretion rate yields a steeper density slope. This immediately gives at least qualitative justification for the universal density profiles seen in simulations (see section I), which invariably transition from a shallow logarithmic slope in the interior to a steep logarithmic slope farther out. As section IV B discussed, due to the biased selection of initial conditions associated with the choice to study halos that survived hierarchical clustering, the accretion rate Γ is generally high at early times and lower at late times. Consequently, the density slope is shallow at small radii, since particles at those radii accreted at early times when Γ was high, and steep at large radii, since particles at those radii accreted at late times when Γ was low.

The density profile itself is

$$\rho(r) = \frac{M}{4\pi r^3}\frac{\mathrm{d}\log M}{\mathrm{d}\log r} = \left(\frac{5}{3\beta}\right)^3 \frac{3}{1+3/\Gamma}\frac{M(q)}{4\pi q^3}\tilde{\delta}(q)^3 = \left(\frac{5}{3\beta}\right)^3 \frac{\delta_{\rm vir}^3}{1+3/\Gamma}\bar{\rho}_0 a^{-3} = \left(\frac{5}{3\beta}\right)^3 \frac{\delta_{\rm vir}^3}{1+3/\Gamma}\bar{\rho}(a), \qquad (25)$$

where $\bar{\rho}(a)$ is the cosmological mean density at a. This expression is to be evaluated at the time a when the shell at r accreted (per equations 18 and 22). It leads to a very natural interpretation: material accreted when the density of the Universe was ρ settles into regions of the halo where the density is $\mathcal{O}(10^2)\rho$ to $\mathcal{O}(10^3)\rho$, where the particular coefficient depends on the values of β and Γ. Also, the profile of the average enclosed density is the same as equation (25) but without the explicit sensitivity to Γ,

$$\frac{M(r)}{(4\pi/3)r^3} = \left(\frac{5}{3\beta}\right)^3 \delta_{\rm vir}^3\bar{\rho}(a), \qquad (26)$$

and a similar interpretation holds here. Ludlow *et al.* (2013) found that the halo mass accreted when the Universe was denser than ρ is equal, on average, to the halo mass in regions where the average enclosed density is larger than $\sim 800\rho$. Equation (26) supplies physical justification for this behavior.

Finally, it is worth noting that although the connection in equations (23-26) between density profiles and accretion histories was derived by relating both to the initial conditions, the profile-history connection may be more general than the relationship with the initial conditions that follows from spherical collapse. Spherical collapse may not be applicable to all of the complicated configurations that arise in a realistic initial density field (although it still manages to be remarkably successful). However, the only feature of spherical collapse that is relevant to the profile-history connection is that a particle settles within a halo at a radius that is proportional to the radius at which it accreted; we specified in equation (22) that a particle's final radius is proportional to $r_{\rm ta}$ and hence to the accretion radius $r_{\rm vir} = r_{\rm ta}/2$. This is likely to be true on average even in configurations that deviate greatly from spherical symmetry.

V. HALO FORMATION

We have explored the connection between a halo's density profile and its accretion history. However, there is a regime where this connection cannot be applied so straightforwardly: the initial formation stage where a halo condenses out of an initially smooth density field. I discuss this phenomenon next.

Initial variations in the density of the Universe are thought to be seeded by inflation. There is not necessarily any requirement that they be smooth. However, dark matter is typically expected to possess random motion. This motion may be thermal motion arising from a previous time at which the dark matter was kinetically coupled to the hot Standard Model plasma. It could also arise from the dark matter production mechanism, for example if it was produced with nonzero momentum by the decay of another species. This random motion would smooth the initial density field on a characteristic free-streaming scale, which is essentially the comoving distance that particles cover.

Note, however, that the random motion itself usually does not play a significant role in halo structures. Most of the smoothing occurs at very early times, deep in the radiation epoch, and its imprint is stretched over time due to cosmic expansion. The residual random velocities are tiny, in a relative sense, by the time that halos start to form.

A. Collapse of a smooth peak

Halos form at sites where the linear density field is locally maximal. Close to such a density peak, the enclosed density contrast can be expanded to lowest nontrivial order as

$$\delta(q,a) = \delta_{\rm pk}(a)[1-(q/q_{\rm pk})^2], \qquad (27)$$

where $\delta_{\rm pk}$ is the value of the density contrast at the peak. $q_{\rm pk}$ is connected to the curvature of the density field at the location of the peak, and it can be interpreted as a measure of the size of the peak. Under the approximations of spherical symmetry and no shell crossing prior to collapse, successive shells collapse when the linearly extrapolated enclosed density contrast reaches the critical value, $\delta_{\rm coll} \simeq 1.686$.[4] Using again the definition $\tilde{\delta}(q) \equiv \delta(q,a)/a$, the collapse time for the shell at q is

$$a_{\rm coll}(q) = \delta_{\rm coll}/\tilde{\delta}(q) = (\delta_{\rm coll}/\tilde{\delta}_{\rm pk})[1-(q/q_{\rm pk})^2]^{-1}. \tag{28}$$

As shorthand, let

$$a_{\rm coll} = \delta_{\rm coll}/\tilde{\delta}_{\rm pk} \tag{29}$$

be the scale factor of collapse for the center of the density peak.

Solving equation (28) for q, we find that the mass shell initially at

$$q(a) = \frac{4\pi}{3}\left(1-\frac{\delta_{\rm coll}}{\tilde{\delta}_{\rm pk}a}\right)^{1/2} q_{\rm pk} = \left(1-\frac{a_{\rm coll}}{a}\right)^{1/2} q_{\rm pk}, \tag{30}$$

collapses at a. The halo mass thus grows as

$$M(a) = \frac{4\pi}{3}q^3\bar{\rho}_0 = \frac{4\pi}{3}\left(1-\frac{a_{\rm coll}}{a}\right)^{3/2} q_{\rm pk}^3\bar{\rho}_0 \equiv \left(1-\frac{a_{\rm coll}}{a}\right)^{3/2} M_{\rm pk} \tag{31}$$

once $a > \delta_{\rm coll}/\tilde{\delta}_{\rm pk}$. Here, we define $M_{\rm pk} \equiv (4\pi/3)q_{\rm pk}^3\bar{\rho}_0$ as the characteristic mass of the peak. The accretion rate defined in equation (20) is

$$\Gamma = \frac{{\rm d}\log M}{{\rm d}\log a} = \frac{3}{2}\frac{a_{\rm coll}}{a-a_{\rm coll}}, \tag{32}$$

which is infinite at the moment of collapse, $a = a_{\rm coll}$, and drops rapidly. The rapidly changing logarithmic accretion rate means that the assumptions made in section IV C are not valid. In particular, the factor β by which mass shells contract after accretion must be much greater for earlier shells than for later ones. Therefore, the halo structure that arises from this process must be derived through a different argument.

B. Self-similarity of collapse

Let us focus on the moments after the initial collapse, when $\Delta a \equiv a - a_{\rm coll} \ll a_{\rm coll}$, or equivalently $\Delta t = t - t_{\rm coll} \ll t_{\rm coll}$, where $t = (2/3)[H(a)]^{-1}$ is the time and $t_{\rm coll} = (2/3)[H(a_{\rm coll})]^{-1}$ is the collapse time. We can rewrite equation (31) as

$$M(\Delta t) = M_{\rm pk}\left(\frac{\Delta a}{a_{\rm coll}}\right)^{3/2} = \left(\frac{2}{3}\right)^{3/2} M_{\rm pk}t_{\rm coll}^{-3/2}\Delta t^{3/2} \tag{33}$$

to lowest order in Δt, since $a \propto t^{2/3}$ implies $\Delta a/a = (2/3)\Delta t/t$. The prefactor in equation (33) is constant for any particular density peak. Once it is fixed, there is no longer any reference to cosmology or initial conditions in this equation. We can simply imagine an isolated Newtonian system that is accreting mass shells at this rate.

The initial velocities of the mass shells are also needed to fully specify the problem. These are connected to the shells' turnaround radii

$$r_{\rm ta} = \frac{3}{5}\frac{q}{\tilde{\delta}(q)} = \frac{3}{5}\frac{q_{\rm pk}a}{\delta_{\rm coll}}\left(1-\frac{a_{\rm coll}}{a}\right)^{1/2} = \frac{3}{5}\frac{a_{\rm coll}q_{\rm pk}}{\delta_{\rm coll}}\left(\frac{\Delta a}{a_{\rm coll}}\right)^{1/2} = \frac{3}{5}\left(\frac{2}{3}\right)^{1/2}\frac{a_{\rm coll}q_{\rm pk}}{\delta_{\rm coll}}t_{\rm coll}^{-1/2}\Delta t^{1/2}, \tag{34}$$

again to lowest order in Δa and Δt. However, I assert that due to the rapid halo growth, the particular turnaround radius of each mass shell is irrelevant, and its final position in the halo is dominantly set by orbital contraction. That

---

[4] I switch to considering $\delta_{\rm coll}$, instead of $\delta_{\rm vir}$ as in section IV, because the latter determines when the collapsing shell accretes onto an existing halo. In the context of halo formation, there is not an existing halo. The difference between $\delta_{\rm vir}$ and $\delta_{\rm coll}$ is minor, however.

is, we assume that a shell's relevant dynamics occur at radii $r \ll r_{\rm ta}$, so we can imagine instead that each shell falls in from infinity (e.g. White 2022). I will test this assumption.

Under these approximations, there is no preferred scale to the problem, so it is self-similar. The system's mass grows according to equation (33), but the length scale r of the system is not specified by the initial conditions (again assuming $r \ll r_{\rm ta}$) and must derive from scaling behaviors intrinsic to gravitational dynamics. Since the dynamical time scale for a system of density ρ is of order $(G\rho)^{-1/2}$ (e.g. Binney and Tremaine 2011), the density of the system must drop as

$$\rho(\Delta t) \propto G^{-1}\Delta t^{-2}, \tag{35}$$

where the proportionality constant is dimensionless. The size of the system must then grow as

$$r(\Delta t) \propto (M/\rho)^{1/3} \propto G^{1/3}M_{\rm pk}^{1/3}t_{\rm coll}^{-1/2}\Delta t^{7/6}, \tag{36}$$

where the proportionality constants are again dimensionless. Importantly, r grows more quickly than the turnaround radii $r_{\rm ta} \propto \Delta t^{1/2}$ of accreted shells. This means the assumption $r \ll r_{\rm ta}$ is valid in the early moments after the initial collapse, i.e., when Δt is small. Moreover, by equating equations (34) and (36), we see that $r \sim r_{\rm ta}$ is achieved when

$$\Delta t|_{r\sim r_{\rm ta}} \propto G^{-1/2}M_{\rm pk}^{-1/2}\left(a_{\rm coll}q_{\rm pk}\right)^{3/2} \propto G^{-1/2}\bar{\rho}_0^{-1/2}a_{\rm coll}^{3/2} \propto G^{-1/2}\bar{\rho}(a_{\rm coll})^{-1/2} \propto t_{\rm coll}, \tag{37}$$

since $t_{\rm coll} \propto [H(a_{\rm coll})]^{-1} \propto G^{-1/2}\rho(a_{\rm coll})^{-1/2}$. That is, the two approximations $r \ll r_{\rm ta}$ and $\Delta t \ll t_{\rm coll}$ that we made are equivalent.

C. The prompt cusp

We are now positioned to explore the nature of the structure that arises promptly from the collapse of a smooth density peak. By the time the $r \ll r_{\rm ta}$ and $\Delta t \ll t_{\rm coll}$ assumptions break down, equations (35) and (37) imply that the density of the collapsed object is

$$\rho_{\rm prompt} \propto \bar{\rho}(a_{\rm coll}), \tag{38}$$

i.e., it is some (possibly large) multiple of the density of the Universe at the initial collapse time. The mass of the collapsed object is just

$$M_{\rm prompt} \propto M_{\rm pk}, \tag{39}$$

some multiple of the characteristic mass of the initial peak.

The internal structure of the object can be determined as follows. Equations (35) and (36) imply

$$\rho(r) \propto r^{-12/7}. \tag{40}$$

Most directly interpreted, this expression only indicates how the density of the overall system scales with its radius as time goes on. However, due to the self-similarity of the system, no other scaling between ρ and r is possible. Equation (40) thus gives the density profile of the prompt structure: it is a density cusp with logarithmic slope $-12/7 \simeq -1.7$. Another perspective is that the apocenters of the orbits of shells that collapse at Δt must scale according to equation (36), while the mass that they enclose scales according to equation (33). This consideration also yields the density profile given by equation (40).

Simulation-based studies typically conclude that the density profile of the prompt cusp is closer to $\rho \propto r^{-3/2}$ (e.g. Delos *et al.* 2019, Angulo *et al.* 2017, Ogiya and Hahn 2018, Delos and White 2023), although some variation has been noted, and slopes as steep as -1.8 have been inferred (Colombi 2021). The main consideration missed in the argument in this section is spherical asymmetry, and it remains to be seen if this would yield a shallower density profile.

VI. SUMMARY

The density profiles of halos are set by the cosmological initial conditions. The general principle is that each accreted mass element quickly settles into an orbit with average radius proportional to the radius at which it was accreted. This orbit remains stable as halos accrete new material, which means that the interior structures are not

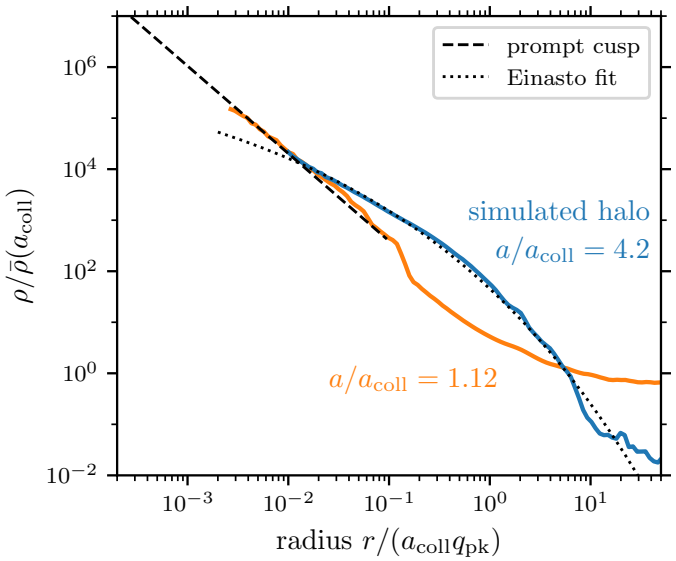

FIG. 5. Density profile of a simulated halo (halo W1 from Delos and White 2023), both shortly after the initial collapse (orange curve) and at a later time (blue curve). The dashed line scales as $\rho \propto r^{-12/7}$ and emphasizes the prompt cusp that arises at the moment of halo formation. A long time after halo formation, the density profile beyond the prompt cusp is fit well by the Einasto form (dotted line; $\alpha = 0.17$).

disrupted by new accretion. Thus as halos grow, they build their structures outward without significantly altering their established interiors.

Figure 5 shows the density profile of a halo in a cosmological simulation. At the center lies the $\rho \propto r^{-12/7}$ density cusp that arises promptly at the halo's formation ($a \sim a_{\rm coll}$). This profile is a consequence of the halo growth rate that results from a smooth density peak in the initial conditions, as discussed in section V. Close to the collapse time (orange), the prompt cusp is essentially the whole halo. At later times (blue), while the prompt cusp persists at the innermost radii, the Einasto form (equation 2) supplies a good fit to the density profile at intermediate and large radii. This portion of the profile is shallow in the interior and steep farther out, reflecting that the halo accreted rapidly early on and more slowly at late times, as discussed in section IV. Due to the biased selection of initial conditions associated with a halo that survived hierarchical clustering, the past accretion history of a general halo is expected to follow this trend.

Appendix A: Orbital contraction

The orbits of particles in a halo contract due to material that is accreted later. Dalal *et al.* (2010) and Delos *et al.* (2019) discuss models of orbital contraction that account for all future accretion. In this appendix, I present a simplified model of orbital contraction. It is based on the notion, illustrated in figure 4, that most of a shell's contraction occurs very quickly after its collapse, even before the next apocenter.

We can consider how much mass is gained within, say, a dynamical time interval

$$t_{\rm dyn} \equiv \sqrt{3\pi/(16G\rho)} \tag{A1}$$

for an object 147 times denser than the cosmological average. This is the time for an object to travel from the center to the edge, under the assumption of homogeneity. For $\rho = 147\bar{\rho}$, $t_{\rm dyn} \simeq 0.183/H$, implying a grows by a factor of about 1.2 during this time. More generally, time scales for orbital dynamics in halos correspond to a growing by some factor $x > 1$; the particular choice $x = 1.2$ is only approximate. But $M \propto a^\Gamma$, so when a grows by the factor x, the enclosed mass grows by a factor of x^Γ. Under the approximation of adiabatic contraction of a Keplerian orbit, the orbit scales as $r \propto 1/M \propto x^{-\Gamma}$, so that

$$\beta = 0.75x^{-\Gamma} \tag{A2}$$

again for some $x > 1$. While we are not in the regime where this approximation is valid, equation (A2) is as good an Ansatz as any.

The logarithmic slope of the mass profile $M(r)$ is then

$$\frac{\mathrm{d}\log M}{\mathrm{d}\log r} = \frac{\mathrm{d}\log M/\mathrm{d}\log q}{\mathrm{d}\log r/\mathrm{d}\log q} = \frac{3}{1+3/\Gamma+\mathrm{d}\log\beta/\mathrm{d}\log q} = \frac{3}{1+3/\Gamma-(\log x)\mathrm{d}\Gamma/\mathrm{d}\log q}$$
$$= \frac{3}{1+3/\Gamma+(\log x)(\mathrm{d}\log\tilde{\delta}/\mathrm{d}\log q)\mathrm{d}\Gamma/\mathrm{d}\log a} = \frac{3}{1+(3/\Gamma)[1-(\log x)\mathrm{d}\Gamma/\mathrm{d}\log a]}, \tag{A3}$$

where Γ is again to be evaluated when the shell at r accreted onto the halo. Compared to equation (23), the rate of change of the accretion rate Γ now enters. For example, if the accretion rate is decreasing ($\mathrm{d}\Gamma/\mathrm{d}\log a < 0$) at some time a, then later material is contracted less than earlier material. This means that more distant material is more dispersed, so the mass increase per radial distance, $\mathrm{d}\log M/\mathrm{d}\log r$, is smaller. The density profile for this model is

$$\rho(r) = \frac{[5\delta_{\rm vir}/(3\beta)]^3}{1+(3/\Gamma)[1-(\log x)\mathrm{d}\Gamma/\mathrm{d}\log a]}\bar{\rho}(a); \tag{A4}$$

compared to equation (25), there is additional sensitivity to the accretion rate through β (equation A2) and there is dependence on the rate of change of the accretion rate.

Appendix B: Self-similar models

Suppose that $\delta(q)$ is a power law,

$$\delta(q) \propto q^{-3\epsilon} \tag{B1}$$

for some $0 < \epsilon < 1$. It follows from the considerations in section IV that the halo mass grows as $M(a) \propto a^{1/\epsilon}$ (i.e., $\Gamma = 1/\epsilon$) and the density profile is

$$\rho(r) \propto r^{-9\epsilon/(1+3\epsilon)}. \tag{B2}$$

These models are self-similar: they have no characteristic scales. Consequently, they can be solved analytically (e.g. Fillmore and Goldreich 1984, Bertschinger 1985), because every particle follows the same trajectory, subject only to rescaling in length and time.

However, there is artificial behavior in the spherically symmetric case with radial orbits. Particles spend time $\mathrm{d}t = \mathrm{d}r/v_r$ in each radius range $\mathrm{d}r$, where v_r is the radial velocity. Far below their apocenters, particles have nearly constant v_r, so they spend the same time in each $\mathrm{d}r$ and thus contribute density $\rho \propto 1/(4\pi r^2\mathrm{d}r) \propto r^{-2}$. With each particle contributing a $\rho \propto r^{-2}$ density profile, it is not possible for the halo as a whole to develop a profile shallower than that. This behavior is artificial because a spherical configuration with radial orbits is unstable (e.g. Binney and Tremaine 2011). If angular momentum is added to the similarity solution (e.g. White 1994), or spherical symmetry dropped (e.g. Ryden 1993, Lithwick and Dalal 2011), then the density scaling in equation (B2) is confirmed.

J. F. Navarro, C. S. Frenk, and S. D. M. White, ApJ **462**, 563 (1996), arXiv:astro-ph/9508025 [astro-ph].
J. F. Navarro, C. S. Frenk, and S. D. M. White, ApJ **490**, 493 (1997), arXiv:astro-ph/9611107 [astro-ph].
J. Einasto, Trudy Astrofizicheskogo Instituta Alma-Ata **5**, 87 (1965).
J. F. Navarro, E. Hayashi, C. Power, A. R. Jenkins, C. S. Frenk, S. D. M. White, V. Springel, J. Stadel, and T. R. Quinn, MNRAS **349**, 1039 (2004), arXiv:astro-ph/0311231 [astro-ph].
J. F. Navarro, A. Ludlow, V. Springel, J. Wang, M. Vogelsberger, S. D. M. White, A. Jenkins, C. S. Frenk, and A. Helmi, MNRAS **402**, 21 (2010), arXiv:0810.1522 [astro-ph].
J. Wang, S. Bose, C. S. Frenk, L. Gao, A. Jenkins, V. Springel, and S. D. M. White, Nature **585**, 39 (2020), arXiv:1911.09720 [astro-ph.CO].
B. Diemer, MNRAS **519**, 3292 (2023), arXiv:2205.03420 [astro-ph.CO].
D. Lynden-Bell, MNRAS **136**, 101 (1967).
J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser, ApJ **379**, 440 (1991).
D. J. Eisenstein and W. Hu, ApJ **496**, 605 (1998), arXiv:astro-ph/9709112 [astro-ph].
Planck Collaboration, A&A **641**, A6 (2020), arXiv:1807.06209 [astro-ph.CO].
R. E. Angulo and S. D. M. White, MNRAS **401**, 1796 (2010), arXiv:0906.1730 [astro-ph.CO].
N. Dalal, Y. Lithwick, and M. Kuhlen, arXiv e-prints , arXiv:1010.2539 (2010), arXiv:1010.2539 [astro-ph.CO].

M. S. Delos, M. Bruff, and A. L. Erickcek, Phys. Rev. D **100**, 023523 (2019), arXiv:1905.05766 [astro-ph.CO].
A. D. Ludlow, J. F. Navarro, M. Boylan-Kolchin, P. E. Bett, R. E. Angulo, M. Li, S. D. M. White, C. Frenk, and V. Springel, MNRAS **432**, 1103 (2013), arXiv:1302.0288 [astro-ph.CO].
S. D. M. White, MNRAS **517**, L46 (2022), arXiv:2207.13565 [astro-ph.CO].
J. Binney and S. Tremaine, *Galactic dynamics*, Vol. 20 (Princeton university press, 2011).
R. E. Angulo, O. Hahn, A. D. Ludlow, and S. Bonoli, MNRAS **471**, 4687 (2017), arXiv:1604.03131 [astro-ph.CO].
G. Ogiya and O. Hahn, MNRAS **473**, 4339 (2018), arXiv:1707.07693 [astro-ph.CO].
M. S. Delos and S. D. M. White, MNRAS **518**, 3509 (2023), arXiv:2207.05082 [astro-ph.CO].
S. Colombi, A&A **647**, A66 (2021), arXiv:2012.04409 [astro-ph.CO].
J. A. Fillmore and P. Goldreich, ApJ **281**, 1 (1984).
E. Bertschinger, ApJS **58**, 39 (1985).
S. D. M. White, arXiv e-prints , astro-ph/9410043 (1994), arXiv:astro-ph/9410043 [astro-ph].
B. S. Ryden, ApJ **418**, 4 (1993).
Y. Lithwick and N. Dalal, ApJ **734**, 100 (2011), arXiv:1010.3723 [astro-ph.CO].