APPENDIX A: THE DEKEL-ZHAO PROFILE

The Dekel-Zhao halo profile (Dekel et al. 2017; Freundlich et al. 2020b),[1] following a more general mathematical analysis by Zhao (1996), is a functional form for dark-matter haloes with two free shape parameters, a concentration c and an inner slope α, allowing the central region to range continuously from a steep cusp to a flat core. It has been found to fit dark-matter haloes in cosmological hydro simulations better than other two-parameter profiles such as the Einasto and the generalized-NFW profiles with a flexible inner slope. A unique feature is that it has analytic expressions not only for the density and mass-velocity profiles but also for the potential and kinetic energy profiles (as well as for gravitational lensing properties). Freundlich et al. (2020b) also provide the typical profile parameters as a function of mass.

The profile of *mean* density within a sphere of radius r is given by

$$\bar{\rho}(r) = \frac{\bar{\rho}_\mathrm{c}}{x^\alpha\,(1+x^{1/2})^{2(3-\alpha)}}\,, \quad x = \frac{r}{R_\mathrm{v}}\,c\,, \tag{A1}$$

$$\bar{\rho}_\mathrm{c} = c^3\mu(c,\alpha)\,\bar{\rho}_\mathrm{v}, \quad \bar{\rho}_\mathrm{v} = \frac{M_\mathrm{v}}{(4\pi/3)R_\mathrm{v}^3}\,, \tag{A2}$$

$$\mu(c,\alpha) = c^{\alpha-3}\,(1+c^{1/2})^{2(3-\alpha)}\,. \tag{A3}$$

For completeness, the local density profile is

$$\rho(r) = \frac{(1-\alpha/3)\,\bar{\rho}_\mathrm{c}}{x^\alpha\,(1+x^{1/2})^{2(3.5-\alpha)}}\,. \tag{A4}$$

The associated mass profile is

$$\frac{M(x)}{M_\mathrm{v}} = \frac{1}{c^3\bar{\rho}_\mathrm{v}}x^3\bar{\rho}(x) = \frac{\mu}{\bar{\rho}_\mathrm{c}}\,x^3\,\bar{\rho}(x)\,. \tag{A5}$$

The log slope of the mass profile is

$$\nu(r) = \frac{3-\alpha}{1+x^{1/2}}\,. \tag{A6}$$

The negative log slope of the density profile is

$$s(r) = \frac{\alpha+3.5x^{1/2}}{1+x^{1/2}}\,. \tag{A7}$$

In order to obtain the mass fraction $f = M(<r)/M_\mathrm{v}$ within a sphere of radius r, we express it using eq. (A5) as

$$f(r) = \mu\,\bar{\rho}_\mathrm{c}^{-1}\,x^3\,\bar{\rho}(x)\,. \tag{A8}$$

Combining eq. (A8) and eq. (A1), we get

$$f(r) = \mu\,x^{3-\alpha}\,(1+x^{1/2})^{-2(3-\alpha)}\,, \tag{A9}$$

from which we obtain

$$x = [(f/\mu)^{-1/[2(3-\alpha)]} - 1]^{-2}\,. \tag{A10}$$

[1] Available for implementation in https://github.com/JonathanFreundlich/Dekel_profile .

Inserting eq. (A10) in eq. (A8), we obtain an equation for $f(r)$,

$$(f/\mu)\,[(f/\mu)^{-1/[2(3-\alpha)]} - 1]^6 = \bar{\rho}(r_\mathrm{t})/\bar{\rho}_\mathrm{c}\,. \tag{A11}$$

We recall from equations 11-14 of Freundlich et al. (2020b) that a more physical pair of shape parameters, that refer to $\rho(r)$ rather to $\bar{\rho}(r)$, may be (c_2, s_1). The concentration c_2 refers to the virial radius with respect to the radius where the log density slope is -2,

$$c_2 = c\left(\frac{1.5}{2-\alpha}\right)^2\,, \tag{A12}$$

valid for $\alpha<2$. The inner slope s_1 is minus the log slope of the local density profile $\rho(r)$ at a given radius r_1 (say $r_1=0.01R_\mathrm{v}$),

$$s_1 = \frac{\alpha+3.5x_1^{1/2}}{1+x_1^{1/2}}\,, \quad x_1 = \frac{r_1}{R_\mathrm{v}}\,c\,. \tag{A13}$$

For completeness, the inverse relations are

$$c = \left(\frac{s_1-2}{(3.5-s_1)(r_1/R_\mathrm{v})^{1/2} - 1.5c_2^{-1/2}}\right)^2\,, \tag{A14}$$

$$\alpha = \frac{1.5s_1 - 2(3.5-s_1)x_{2,1}^{1/2}}{1.5-(3.5-s_1)x_{2,1}^{1/2}}\,, \quad x_{2,1} = \frac{r_1}{R_\mathrm{v}}c_2\,. \tag{A15}$$

We note that a valid DZ solution is not guaranteed for any arbitrary pair of values (c_2, s_1), e.g., there is no solution where the denominator in either eq. (A14) or eq. (A15) vanishes.

A best fit for an NFW profile with a given concentration is obtained, e.g., by minimizing residuals in uniformly spaced log radii in the range $\log(r/R_\mathrm{v}) = (-2, 0)$. For $c_\mathrm{NFW} = 5$ we obtain for the best-fit DZ parameters $(c,\alpha) = (7.126, 0.2156)$ or $(c_2, s_1) = (5.035, 0.9076)$. A slightly better fit near $0.01R_\mathrm{v}$ can be obtained with $(c_2, s_1) \simeq (4.8, 1.0)$, but this is at the expense of slightly larger deviations at large radii.

The gravitational potential as given in eq. 19 of Freundlich et al. (2020b) is

$$U(r) = -V_\mathrm{v}^2\left[1 + 2c\mu\left(\frac{\chi_c^{\tilde{\alpha}} - \chi^{\tilde{\alpha}}}{\tilde{\alpha}} - \frac{\chi_c^{\tilde{\alpha}+1} - \chi^{\tilde{\alpha}+1}}{\tilde{\alpha}+1}\right)\right]\,, \tag{A16}$$

$$\chi = \frac{x^{1/2}}{1+x^{1/2}}, \quad \chi_c = \frac{c^{1/2}}{1+c^{1/2}}, \quad \tilde{\alpha} = 2(2-\alpha)\,. \tag{A17}$$

The velocity dispersion that stems from the Jeans equation (Freundlich et al. 2020b, eq. 22), providing the kinetic energy per unit mass, is

$$\sigma_r^2(r) = 2c\mu V_\mathrm{v}^2\frac{\rho_\mathrm{c}}{\rho(r)}\Big[\mathcal{B}(4-4\alpha,\,9,\,\zeta)\Big]_\chi^{\chi_c}\,, \tag{A18}$$

where $\mathcal{B}(a,b,x) = \int_0^x t^{a-1}(1-t)^{b-1}dt$ is the incomplete beta function, the brackets denote the difference of the enclosed function between 1 and χ, i.e., $[f(\zeta)]_\chi^{\chi_c} \equiv f(\chi_c) - f(\chi)$, and $\rho_c = (1-\alpha/3)\bar{\rho}_\mathrm{c}$. The

definition of the incomplete beta function has been extended to negative parameters since the bracketed term is well-defined. Equations B1 and B3 of Freundlich et al. (2020b) are the equivalent expressions in terms of finite series.

Adding an additional point mass $M_{\rm b}$ at the halo center adds the velocity dispersion term (Freundlich et al. 2020b, eq. C15)

$$\sigma^2_{M_{\rm b}}(r) = 2c\frac{GM_{\rm b}}{R_{\rm v}}\frac{\rho_c}{\rho(r)}\Big[\mathcal{B}(-2-2\alpha,\,9,\,\zeta)\Big]^1_\chi\,. \quad \text{(A19)}$$

For comparison, the NFW density profile, with an inner cusp of negative log density slope $\alpha=1$ and a free NFW concentration parameter $c_{\rm h}$, is

$$\rho(r) = \frac{\rho_0}{x\,(1+x)^2}\,, \quad x = \frac{r}{R_{\rm v}}c_{\rm h}\,. \quad \text{(A20)}$$

The mass profile is

$$\frac{M(r)}{M_{\rm v}} = \frac{A(x)}{A(c_{\rm h})}\,, \quad \frac{\bar\rho(r)}{\bar\rho_{\rm v}} = \frac{A(x)}{A(c_{\rm h})}\frac{c_{\rm h}^3}{x^3}\,, \quad \text{(A21)}$$

where

$$A(x) = \ln(1+x) - \frac{x}{1+x}\,. \quad \text{(A22)}$$

The profile of the log slope of the mass profile is

$$\nu(r) = \frac{x^2}{(1+x)^2\,A(x)}\,. \quad \text{(A23)}$$

APPENDIX B: THE VELA SIMULATIONS

The VELA suite consists of hydro-cosmological simulations zooming-in on 34 moderately massive galaxies, presented in more detail in Ceverino et al. (2014) and Zolotov et al. (2015). This suite has been used to study central issues in the evolution of galaxies at high redshifts, including, e.g., compaction to blue nuggets and the trigger of quenching (Zolotov et al. 2015; Tacchella et al. 2016b,a), evolution of global shape (Ceverino, Primack & Dekel 2015; Tomassetti et al. 2016), violent disc instability (Mandelker et al. 2014, 2017; Inoue et al. 2016), the SFR-density relation by supernova feedback (Dekel et al. 2019), post-compaction formation of discs and rings (Dekel et al. 2020a,b), OVI in the CGM (Roca-Fàbrega et al. 2018; Strawn et al. 2020), and angular momentum and galaxy size (Jiang et al. 2019). Additional analysis of the same suite of simulations are discussed in Moody et al. (2014); Snyder et al. (2015). This appendix provides an overview of the relevant features of these simulations.

The VELA simulations make use of the Adaptive Refinement Tree (ART) code (Kravtsov, Klypin & Khokhlov 1997; Kravtsov 2003; Ceverino & Klypin 2009), which follows the evolution of a gravitating N-body system and the Eulerian gas dynamics using an adaptive mesh refinement. The maximum spatial resolution is $17-35\,$pc at all times. The code incorporates subgrid recipes for physical process that are relevant for galaxy formation, such as gas cooling by atomic hydrogen and helium, metal and molecular hydrogen cooling, photoionization heating by the UV background with partial self-shielding, star formation, stellar mass loss, metal enrichment of the ISM and stellar feedback. Supernovae and stellar winds are implemented by local injection of thermal energy as described in Ceverino & Klypin (2009); Ceverino, Dekel & Bournaud (2010) and Ceverino et al. (2012). Radiation-pressure stellar feedback is implemented at a moderate level, following Dekel et al. (2013), as described in Ceverino et al. (2014).

Cooling and heating rates are based on the CLOUDY code (Ferland et al. 1998). A uniform UV background based on the redshift-dependent Haardt & Madau (1996) model is assumed, except at gas densities higher than $0.1\,{\rm cm}^{-3}$, where partial self-shielding allows dense gas to cool down to ~ 300K. The assumed equation of state is that of an ideal mono-atomic gas. Artificial fragmentation on the cell size is prevented by introducing a pressure floor, which ensures that the Jeans scale is resolved by at least 7 cells (see Ceverino, Dekel & Bournaud 2010). Star particles form in timesteps of $5\,$Myr in cells where the gas density exceeds $1\ {\rm cm}^{-3}$ and the temperatures is below 10^4K. The code implements a stochastic star formation where a star particle with a mass of 42% of the gas mass forms with a probability $P = (\rho_{\rm g}/10^3\,{\rm cm}^{-3})^{1/2}$ but not higher than 0.2.

Thermal feedback that mimics the energy release from stellar winds and supernova explosions is incorporated as a constant heating rate over the 40 Myr following star formation. A velocity kick of $\sim 10\,{\rm km\,s}^{-1}$ is applied to 30 % of the newly formed stellar particles – this enables SN explosions in lower density regions where the cooling may not overcome the heating without implementing an artificial shutdown of cooling (Ceverino & Klypin 2009). The code also incorporates the later effects of Type Ia supernova and stellar mass loss, and it follows the metal enrichment of the ISM. Radiation pressure is incorporated through the addition of a non-thermal pressure term to the total gas pressure in regions where ionizing photons from massive stars are produced and may be trapped. This ionizing radiation injects momentum in the cells neighbouring massive star particles younger than $5\,$Myr, and whose column density exceeds $10^{21}\,{\rm cm}^{-2}$, isotropically pressurizing the star-forming regions (see more details in Agertz et al. 2013; Ceverino et al. 2014).

The initial conditions for the simulations are based on DM haloes that were drawn from dissipationless N-body simulations at lower resolution in cosmological boxes of $15-60\,$Mpc. The ΛCDM cosmological model was assumed with the WMAP5 values of the cosmological parameters, $\Omega_{\rm m}=0.27$, $\Omega_\Lambda=0.73$, $\Omega_{\rm b}=0.045$, $h=0.7$ and $\sigma_8=0.82$ (Komatsu et al. 2009). Each halo was selected to have a given virial mass at $z=1$ and no ongoing major merger at $z=1$. This latter criterion eliminated less than 10 % of the haloes, those that tend to be in a dense, proto-cluster environment at $z\sim 1$. The virial masses at $z=1$ were chosen to be in the range $M_{\rm v}=2\times10^{11}-2\times10^{12}\ M_\odot$, about a median of $4.6\times10^{11}\ M_\odot$. If left in isolation, the median mass at $z=0$ was intended to be $\sim 10^{12}\ M_\odot$.

The VELA cosmological simulations are state-of-the-art in terms of high-resolution adaptive mesh refinement hydrodynamics and the treatment of key physical processes at the subgrid level. In particular, they trace the cosmological streams that feed galaxies at high redshift, including mergers and smooth flows, and they resolve the violent disc instability that governs high-z disc evolution and bulge formation (Ceverino, Dekel & Bournaud 2010; Ceverino et al. 2012, 2015; Mandelker et al. 2014). To mention a few limitations, like in other simulations, the treatments of star formation and feedback processes are rather simplified. The code may assume a realistic SFR efficiency per free fall time on the grid scale but it does not follow in detail the formation of molecules and the effect of metallicity on SFR. The feedback is treated in a crude way, where the resolution does not allow the capture of the Sedov-Taylor phase of supernova bubbles. The radiative stellar feedback assumed no infrared trapping, in the spirit of low trapping advocated by Dekel & Krumholz (2013) based on Krumholz & Thompson (2013), which makes the radiative feedback weaker than in other simulations that assume more significant trapping (Murray, Quataert & Thompson 2010; Hopkins, Quataert & Murray 2012). AGN feedback, and feedback associated with cosmic rays and magnetic fields, are not yet implemented. Nevertheless, as shown in Ceverino et al. (2014), the star-formation rates, gas fractions, and stellar-to-halo mass ratio are all in the ballpark of the estimates deduced from observations.

The virial and stellar properties of the galaxies are listed for example in Table 1 of Dekel et al. (2020b). The virial mass $M_{\rm v}$ is the total mass within a sphere of radius $R_{\rm v}$ that encompasses an overdensity of $\Delta(z) = [18\pi^2 - 82\Omega_\Lambda(z) - 39\Omega_\Lambda(z)^2]/\Omega_{\rm m}(z)$, where $\Omega_\Lambda(z)$ and $\Omega_{\rm m}(z)$ are the cosmological parameters at z (Bryan & Norman 1998; Dekel & Birnboim 2006). The stellar mass $M_{\rm s}$ is the instantaneous mass in stars within a radius of $0.2R_{\rm v}$, accounting for past stellar mass loss. We start the analysis at the cosmological time corresponding to expansion factor $a = 0.125$ (redshift $z = 7$), and most galaxies reach $a = 0.50$ ($z = 1$). Each galaxy is analyzed at output times separated by a constant interval in a, $\Delta a = 0.01$, corresponding at $z = 2$ to ~ 100 Myr (roughly half an orbital time at the disc edge). The sample consists of totally ~ 1000 snapshots in the redshift range $z = 7 - 0.8$ from 35 galaxies that at $z = 2$ span the stellar mass range $(0.2 - 6.4) \times 10^{11} M_\odot$. The half-mass sizes $R_{\rm e}$ range $R_{\rm e} \simeq 0.4 - 3.2\,{\rm kpc}$ at $z = 2$. The determination of the centre of the galaxy is outlined in detail in appendix B of Mandelker et al. (2014). Briefly, starting form the most bound star, the centre is refined iteratively by calculating the centre of mass of stellar particles in spheres of decreasing radii down to 130 pc or when the number of stellar particles in the sphere drops below 20.

We identify the major event of wet compaction to a blue nugget for each galaxy. This is the one that leads to a significant central gas depletion and SFR quenching, and marks the transition from dark-matter to baryon dominance within $R_{\rm e}$. Following Zolotov et al. (2015) and Tacchella et al. (2016a), the most physical way to identify the compaction and blue nugget is by the steep rise of gas density (and SFR) within the inner 1 kpc to the highest peak, as long as it is followed by a significant, long-term decline in central gas mass density (and SFR). The onset of compaction can be identified as the start of the steep rise of central gas density prior to the blue-nugget peak. An alternative identification is using the shoulder of the stellar mass density within 1 kpc where its rise due to the starburst associated with the compaction turns into a plateau of maximum long-term compactness slightly after the blue-nugget peak of gas density. This is a more practical way to identify blue nuggets in observations (e.g. Barro et al. 2017).

APPENDIX C: SATGEN - A SEMI-ANALYTIC SATELLITE GENERATOR

The semi-analytic model for satellite galaxies SatGen is presented in Jiang et al. (2020).[2] It can generate statistical samples of satellite populations for a host halo of desired mass, redshift, and cosmological parameters. The model combines DM halo merger trees, empirical relations for the galaxy-halo connection, and simple analytical prescriptions for tidal effects, dynamical friction, and ram pressure stripping (if the satellites contains gas). SatGen emulates cosmological zoom-in simulations in certain aspects. Satellites can reside in subhaloes of desired density profiles, with cores or cusps, depending on the subhalo response to baryonic physics that are formulated from hydro-simulations or physical modeling. The host potential can be composed of a DM halo and baryonic components, such as a disc and a bulge, each described by a density profile that allows analytic integration of the satellite orbits. The subhalo profile and the stellar mass and structure of a satellite evolves due to tidal heating and tidal mass loss, which depend on its initial structure. SatGen complements simulations by propagating the effect of halo response found in simulated central galaxies to satellites (which are typically not properly resolved in simulations). It outperforms simulations by capturing the halo-to-halo variance of satellite statistics and overcoming artificial disruption due to insufficient resolution (van den Bosch & Ogiya 2018; Green, van den Bosch & Jiang 2021). Certain features of SatGen that are relevant for our current study are elaborated on below.

SatGen generates halo merger trees using the algorithm of Parkinson, Cole & Helly (2008) as re-calibrated by Benson (2017). Merger trees are constructed using the time-stepping advocated in Appendix A of Parkinson, Cole & Helly (2008), which corresponds to $\Delta z \simeq 0.001$, but for book keeping the temporal resolution is down-sampled to timesteps of $\Delta t = 0.1 t_{\rm dyn}(z)$, where $t_{\rm dyn} = \sqrt{3\pi/[16G\,\Delta\,\rho_{\rm crit}(z)]}$ is the instantaneous virial time of DM haloes. In the EdS regime, approximately

[2] Available for implementation in https://github.com/shergreen/SatGen .

valid at $z>1$, $\Delta\simeq 200$ and the mean universal density approaches the critical cosmological density.

The structure of the host potential is determined in the following way. First, the virial mass of the system $M_{\rm v}(t)$ is given by following the main progenitor along the main branch of the merger tree. The stellar mass $M_{\rm s}(t)$ is assigned according to the abundance matching relations of Rodríguez-Puebla et al. (2017). Second, we determine the DZ profile of the halo, including the effect of baryonic, as follows. The concentration parameter in a DM-only scenario, $c_{2,\rm DMO}(M_{\rm v},t)$, is obtained from the empirical relation of Zhao et al. (2009). We then consider the halo response to baryons following Freundlich et al. (2020a), which provides empirically the ratio of the baryon-affected concentration and the DM-only concentration, $c_2/c_{2,\rm DMO}$, as a function of the stellar-to-halo-mass ratio $M_{\rm s}/M_{\rm v}$, and the inner logarithmic slope of the system $s_1=\mathrm{d}\ln\rho/\mathrm{d}\ln\mathrm{r}$ at $r=0.01R_{\rm v}$. Finally, we compute the DZ-profile parameters (c,α) using c_2 and s_1. This procedure applies to both the host halo and the progenitors of satellites prior to infall.

The orbits of incoming satellites are initialized as follows. We consider the infall locations to be isotropically distributed on the virial sphere of the host halo, for which we randomly draw an azimuthal angle ϕ) from $[0,2\pi]$ and a cosine polar angle ($\cos\theta$) from $[0,1]$. We assume that the orbital energy is the same as that of a circular orbit of of velocity $V_{\rm v}(t)$ at radius $R_{\rm v}(t)$, and randomly assign a circularity ϵ from a distribution, $\mathrm{dP}/\mathrm{d}\epsilon=\pi\sin(\pi\epsilon)/2$, which approximates the ϵ distribution of infalling satellites measured in cosmological simulations (Wetzel 2011; van den Bosch 2017).

We follow the orbits by treating satellites as point masses. At each timestep, SatGen solves the equations of motion

$$\ddot{\boldsymbol{r}}=-\nabla\Phi\ +a_{\rm DF}\,, \tag{C1}$$

where $\boldsymbol{r}$ is the position vector, Φ is the gravitational potential, and $a_{\rm DF}$ is the acceleration due to dynamical friction, modeled using the Chandrasekhar (1943) formula as given in eq. 12 and eq. 13

We model the tidal mass loss using

$$\dot{m}=-A\frac{m(>\ell_{\rm t})}{t_{\rm dyn}(r)}, \tag{C2}$$

where we have introduced a fudge parameter A as the stripping efficiency to encapsulate uncertainties in the definition of the tidal radius. That is, the timescale on which stripping occurs is the local dynamical time $t_{\rm dyn}(r)=\sqrt{3\pi/16G\bar{\rho}(r)}$ divided by A, with $\bar{\rho}(r)$ the average density of the host system within radius r. We use $A=0.55$ following the calibration by Green, van den Bosch & Jiang (2021) from simulations. The mass loss over a timestep Δt is then given by $\Delta m=\dot{m}\,\Delta t$.

To keep track of DF heating, we register the work done by DF on a satellite at each step, or equivalently the orbital energy change at each step,

$$\Delta W(t+\Delta t)=E(t)-E(t+\Delta t)\,, \tag{C3}$$

Note that the orbital energy E at time $t+\Delta t$ includes the contribution from the stripped mass Δm, which is assumed to be on the same orbit of the satellite that it used to belong to.

The structural evolution of satellites in response to tidal mass loss, heating, and re-virialization, is modeled using the empirical tidal tracks from simulations (Penarrubia et al. 2010). Note that the tidal track is conditioned on the initial structure of the satellites, which is important for capturing the difference in DF heating due to a compact satellite versus a diffuse one. In the current study we do not explicitly include baryons within the satellites.

Jiang et al. (2020) used the model to study satellites of Milky-Way sized hosts, making it emulate simulations of bursty or smooth star formation and experimenting with a disc potential in the host halo. They found that the model reproduces the observed satellite statistics in the Milky Way and M31 reasonably well. Different physical recipes make a difference in satellite abundance and spatial distribution at the 25% level, not large enough to be distinguished by current observations given the halo-to-halo variance. The MW/M31 disc depletes satellites by $\sim$20% and has only a subtle effect of diversifying the internal structure of satellites, which may be important for alleviating certain small-scale problems. We do not explicitly include in the current study a central baryonic component.

APPENDIX D: N-BODY SIMULATIONS OF CUSP HEATING AND ITS RELAXATION

To test the impact of cusp heating by satellites and the following relaxation, we run idealized N-body simulations, each with a host halo and a single merging satellite. At the beginning of the simulation, the two haloes are set in equilibrium, spherically symmetric and with an isotropic velocity dispersion. In this case, the phase-space density of the particles is determined by the particle specific energy E and its radial distance from the halo center r. The halo density profile can be written as

$$\rho(r)=m_{\rm p}\int f(E)d^3\mathbf{v}=4\pi\sqrt{2}m_{\rm p}\int_0^{\Psi}\sqrt{\Psi-\mathcal{E}}f(\mathcal{E})d\mathcal{E}, \tag{D1}$$

where $\Psi=\Phi_0-\Phi$, with Φ the gravitational potential and Φ_0 its value at the boundary of the system, which we set at $4R_{\rm v}$. The energy per unit mass is $\mathcal{E}=\Psi-(1/2)v^2$. Given the density profile, the gravitational potential can be derived from the Poisson equation. For a realistic stationary halo, Ψ is a monotonically decreasing function of r, so ρ can be written as a function of Ψ. Taking the derivative of both sides of eq. (D1) with respect to Ψ, one gets

$$\frac{d\rho}{d\Psi}=\sqrt{8}\pi m_{\rm p}\int_0^{\Psi}\frac{f(\mathcal{E})}{\sqrt{\Psi-\mathcal{E}}}\,d\mathcal{E}, \tag{D2}$$

where $m_{\rm p}$ is the particle mass. The above equation can be solved to give the Eddington's inversion formula (Ed-

dington 1916)

$$f(\mathcal{E}) = \frac{1}{\sqrt{8}\pi^2 m_{\rm p}} \frac{\rm d}{{\rm d}\mathcal{E}} \int_0^{\mathcal{E}} \frac{{\rm d}\Psi}{\sqrt{\mathcal{E}-\Psi}} \frac{{\rm d}\rho}{{\rm d}\Psi}. \qquad \text{(D3)}$$

The particle positions and velocities are randomly drawn from the density profile, eq. (D1) and the velocity (energy) distribution, eq. (D3). For the host halo, an NFW profile is used with a sharp truncation at $4\,R_{\rm v}$. For the satellites, the density profile, either NFW profile (compact satellite) or Burkert profile (diffuse satellite), is truncated exponentially at the virial radius following Kazantzidis, Zentner & Kravtsov (2006) to roughly account for tidal truncation.

After generating the stationary halos, the satellite is put at the apocenter of an orbit with specified orbital parameters, e.g., the circularity and total energy. The system is then evolved using the public N-body SPH code GADGET-2 (Springel & Hernquist 2005).

The centers of the host and satellite are identified by searching for the most-bound particle, the particle that has the most negative total energy $E_i = \Phi(r_i) + (1/2)\,v_i^2$ within the corresponding halo. For the satellite, the bound mass is computed at each snapshot using an iterative un-binding algorithm (van den Bosch & Ogiya 2018).

APPENDIX E: CUSPCORE - AN ANALYTIC MODEL FOR DM RESPONSE TO OUTFLOWS

Freundlich et al. (2020a)[3] presents a simple analytic model for the response of a dissipationless spherical system to an instantaneous mass change at its center. It has been applied there to the formation of flat cores in low-mass dark-matter haloes and the origin of ultra-diffuse galaxies (UDGs) from outflow episodes driven by supernova feedback, but it is applicable for any rapid changes in the central mass. Here we use it for the dark-matter response to AGN-driven central gas ejection. This model generalizes an earlier simplified analysis of an isolated shell (Dutton et al. 2016) into a system with continuous density, velocity and potential profiles.

The DM response is divided into two steps: an instantaneous change of potential at constant velocities due to a given rapid mass loss (or mass gain), followed by energy-conserving relaxation to a new Jeans equilibrium. The halo profile is modeled by the two-parameter Dekel-Zhao profile described in §A, using the analytic expressions for the associated potential and kinetic energies at equilibrium. The way energy conservation is applied in the second stage of this model is not formally justified in the case of shell crossing, so its validity as an approximation should be based on testing against simulations. In Freundlich et al. (2020a), the model has been tested against NIHAO cosmological zoom-in simulations, where it successfully predicts the evolution of the inner DM profile between successive snapshots in about 75% of the cases, failing mainly in merger situations when the system strongly deviates from Jeans equilibrium.

The energy per unit mass of a shell at radius $r_{\rm i}$ in the initial halo at Jeans equilibrium is the sum

$$E_{\rm i}(r_i) = U(r_{\rm i}; p_{\rm i}) + K(r_{\rm i}; p_{\rm i})\,, \qquad \text{(E1)}$$

where $U(r_{\rm i}; p_{\rm i})$ and $K(r_{\rm i}; p_{\rm i})$ are functional forms for the potential and kinetic energies per unit mass, which depend on the parameters $p_{\rm i}$ that characterize the initial halo density profile. We use the DZ profile with the parameters c and α, for which the potential $U(r_{\rm i}; p_{\rm i})$ is given by eq. (A16), and the kinetic energy $K(r_{\rm i}; p_{\rm i})$ derives from eq. (A19), stemming from Jeans equilibrium. For the two energies we may consider an additional baryonic component, characterized by additional parameters. In the temporary state immediately after the instantaneous mass change by m (where $m<0$ for an outflow and $m>0$ for an inflow), the energy becomes

$$E_{\rm t}(r_{\rm i}) = U(r_{\rm i}; p_{\rm i}) - \frac{Gm}{r_{\rm i}} + K(r_{\rm i}; p_{\rm i})\,. \qquad \text{(E2)}$$

After relaxation to the final Jeans equilibrium state of the halo, whose profile is described by the parameters $p_{\rm f}$, the shell encompassing a given mass has moved to a final radius $r_{\rm f}$ and its energy is

$$E_{\rm f}(r_{\rm f}) = U(r_{\rm f}; p_{\rm f}) - \frac{Gm}{r_{\rm f}} + K(r_{\rm f}; p_{\rm f}, m)\,, \qquad \text{(E3)}$$

where the kinetic energy is again set by the Jeans equation but it also depends on the mass change m. The radius $r_{\rm f}$ is itself a function of the final parameters $p_{\rm f}$, given that the enclosed mass is constant, $M(r_{\rm f}; p_{\rm f}) = M(r_{\rm i}; p_{\rm i})$. The assumed energy conservation during the relaxation phase corresponds to $E_{\rm f}(r_{\rm f}) = E_{\rm t}(r_{\rm i})$, which is solved numerically to obtain the final halo parameters $p_{\rm f}$. In practice, we minimize the difference $E_{\rm f}(r_{\rm f}) - E_{\rm t}(r_{\rm i})$ for hundred shells equally spaced in $\log(r/R_{\rm v})$ from -2 to 0 (thus giving more weight to central regions than linearly-spaced shells). The assumed energy conservation per shell that encompasses a given mass is not formally justified in the case of shell crossing, and our use of it is based on the success of this model in reproducing the results of simulations. We refer to Freundlich et al. (2020a) for more details, and to Freundlich, Dekel & Jiang (2019) for a brief presentation of the model.

APPENDIX F: A SATELLITE WITH CENTRAL BARYONS

Complementing the main text, Fig. F1 is the analog of Fig. 4, showing the result of a SatGen run with our fiducial NFW host halo, but with the satellites following the more compact DZ-profile fits to the VELA simulated galaxies using the *total* mass including the baryons rather than the dark matter alone. We learn that for the more compact satellites, as expected, the penetrating mass to the host cusp is higher, and the energy deposited in the cusp is higher accordingly. However, the difference is rather small, with $m/m_{\rm v} \simeq 0.5$ compared to

[3] Available for application in https://github.com/Jonathanfreundlich/CuspCore .

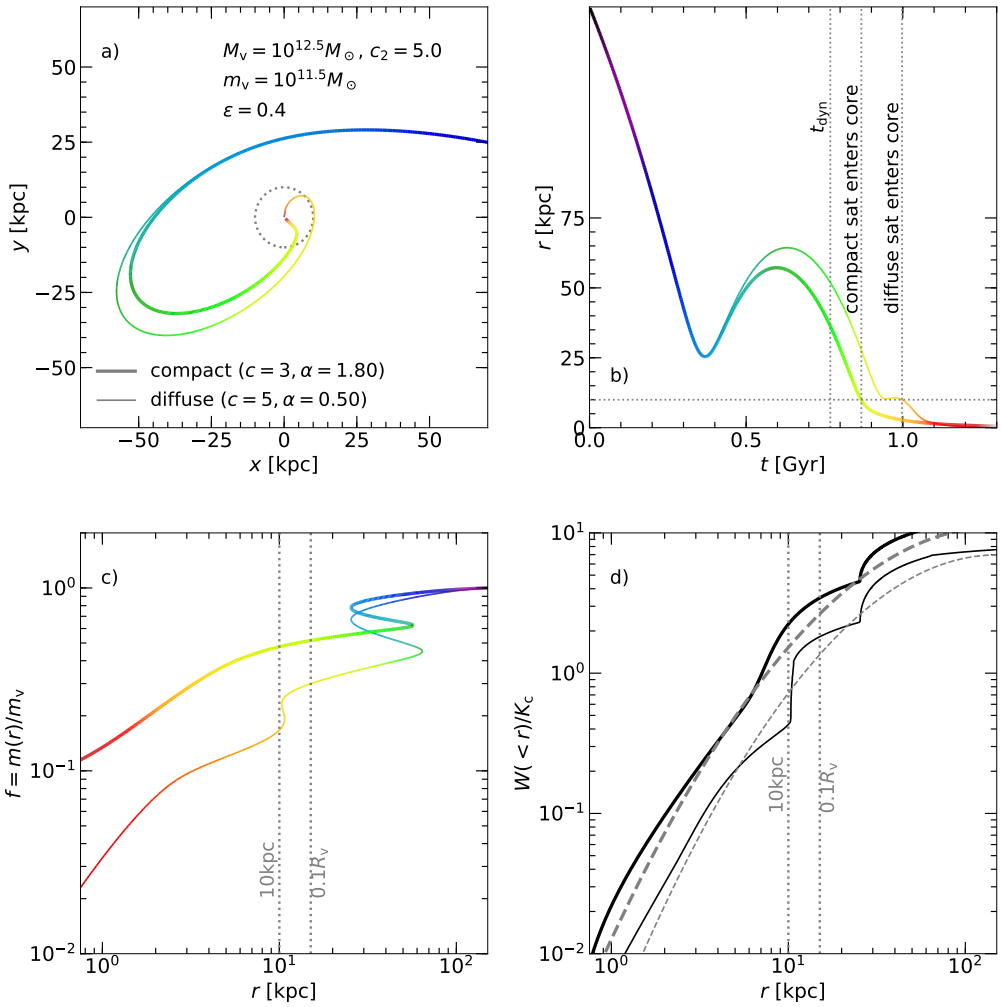

Figure F1. Semi-analytic SatGen simulations of single satellites, diffuse and compact, similar to Fig. 4, but where the satellite profiles are the best DZ fits to the *total* mass in the VELA pre and post compaction galaxies from Fig. 1, including the baryons, with DZ parameters $(c, \alpha) = (5, 0.5)$ and $(3, 1.8)$, respectively.

0.4, and $W_c/K_c \simeq 2.5$ compared to 2, for the compact satellites.

APPENDIX G: A STEEP-CUSP HOST HALO

Here we show the same results that have been shown in the main text for an NFW host halo with a moderately steep cusp, but for a steep-cusp host of a DZ profile with $s_1 = 1.5$ and $c_2 = 5$.

Figure G1, same as Fig. 3, shows the toy model predictions as a function of the satellite compactness. Figure G2, same as Fig. 4, shows the results of a SatGen run with a single satellite. Figure G3, same as Fig. 5, refers to SatGen runs with a cosmological sequence of satellites. Figure G4 and Fig. G5, same as Fig. 9 and Fig. 10 show the results of CuspCore for a single satellite and a cosmological sequence of satellites, respectively. The results for the steep-cusp host are discussed in comparison to the results for the NFW host in the main text.

APPENDIX H: OUTFLOW WITH SATELLITE MASS ADDED TO HOST

APPENDIX I: TOY-MODEL S_1 AND C_2

Figure I1 is the analog of Fig. 3, showing the toy-model estimates for the satellite mass in the host cusp and the energy deposited there by dynamical friction as a function of the satellite initial profile, but here for the more accessible parameters (c_2, s_1) instead of the natural DZ parameters (c, α). The conclusion is the same as in Fig. 3.

I1 Orbit circularity as a function of redshift and halo mass

One may elaborate on the redshift and mass dependence of the satellite orbit circularity as one of the factors in the tendency of DF heating to be more effective at higher masses and higher redshifts.

The orbit is characterized at R_v by two parameters, e.g., energy and angular momentum, or the veloc-

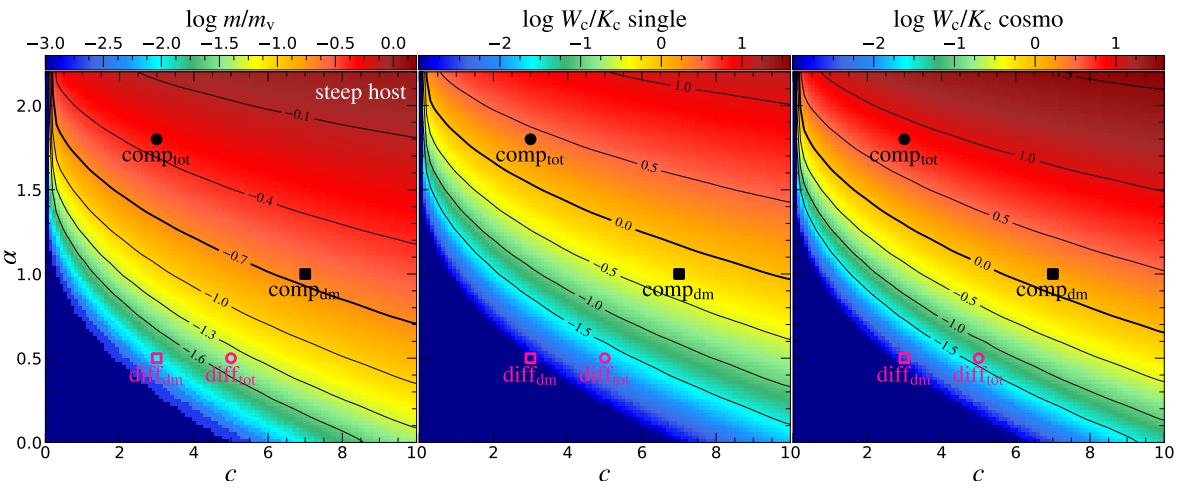

Figure G1. Same as Fig. 3 but for the fiducial steep-cusp host, with $s_1 = 1.5$ and $c_2 = 5$ in the DZ profile. Shown are the toy-model estimates for satellite penetration and energy deposited in the host cusp by dynamical friction, as a function of the satellite compactness via the Dekel-Zhao profile parameters of concentration and inner slope (c, α). For the steep-cusp host we read for the diffuse and compact satellites respectively $m/m_v \sim 0.01, 0.23$, $W_c/K_c(\text{single}) \sim 0.001, 0.56$ and $W_c/K_c(\text{cosmo}) \sim 0.003, 1.53$. The satellite stripping is stronger due to the steeper cusp, but the heating by compact satellites is still significant during half a virial time.

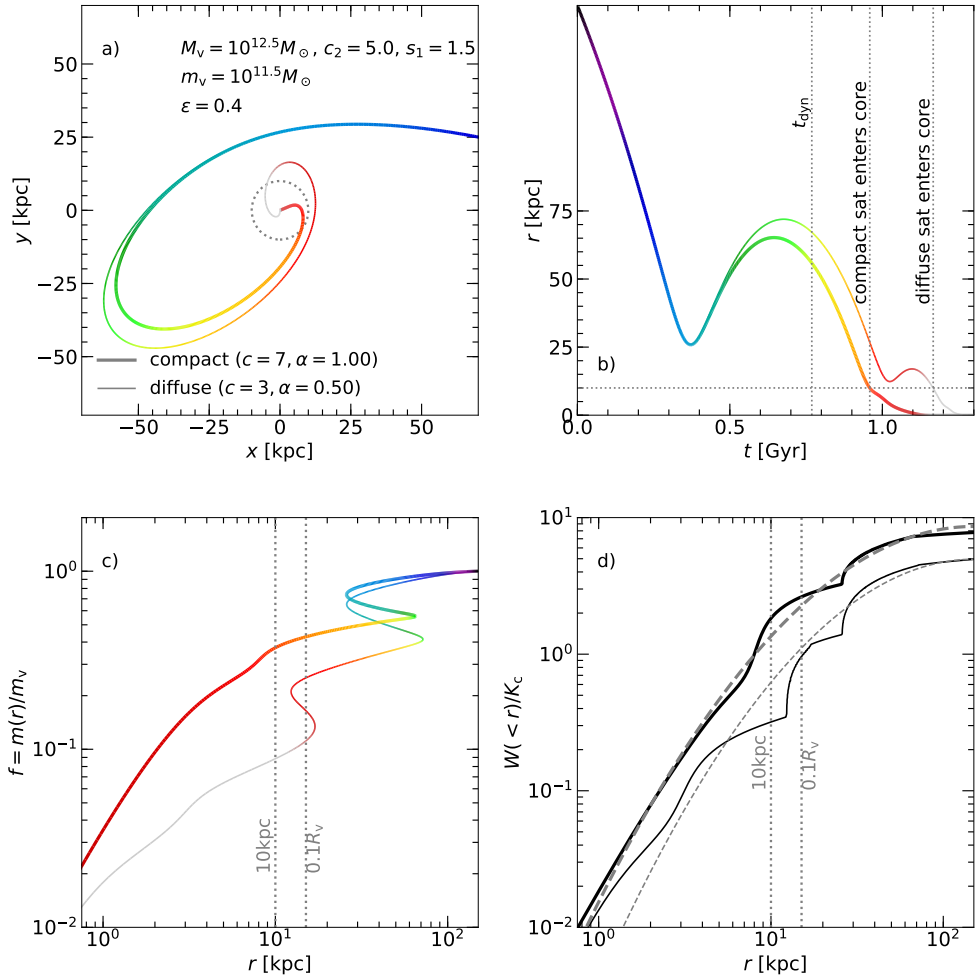

Figure G2. Same as Fig. 4, for an $m_v/M_v = 0.1$ single satellite, but for a steep-cusp host halo with DZ parameters $(c_2, s_1) = (5, 1.5)$. The penetrating satellite mass and the DF energy deposited in the steep cusp are similar to the case of an NFW halo, indicating significant heating by the compact satellite and only partial heating by the diffuse satellite. The functional fits (dashed) to be used by CuspCore are listed in Table 1.

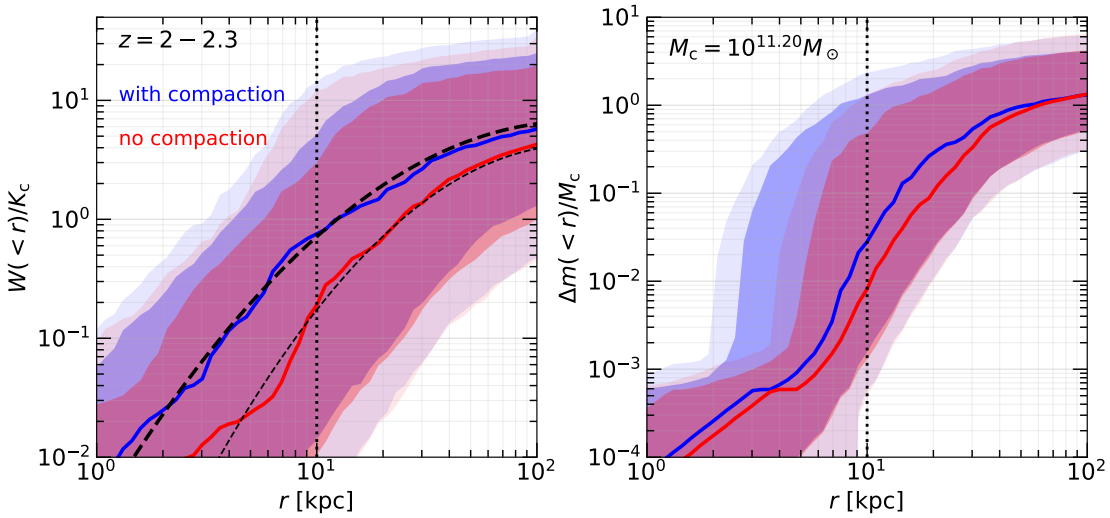

Figure G3. Same as Fig. 5 but for the steep-cusp host. Shown are the results of a SatGen simulation of a cosmological sequence of satellites during one halo virial time at $z \sim 2$, where $t_v \simeq 0.5$ Gyr. The host halo starts with a DZ steep-cusp profile of $(c_2, s_1) = (5, 1.5)$. The fits for the deposited energy by DF to be used in CuspCore is marked (dashed black), and listed in Table 1. Most of the satellite mass is deposited near the outer edge of the cusp, with only a small fraction of the mass penetrating to the inner cusp. The energy deposited in the steep cusp by DF on a sequence of compact satellites is lower than that deposited in the NFW cusp (Fig. 5) by a factor of ~ 2.5. However, it is 75% of the cusp kinetic energy, implying heating also in the steep-cusp host. The heating by diffuse satellites is weaker, only 20% of the cusp energy.

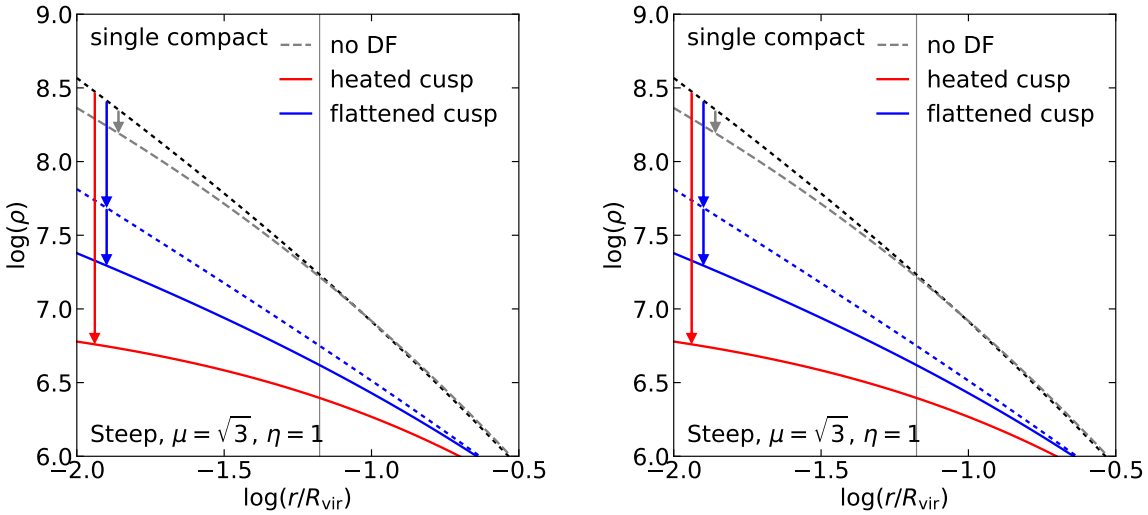

Figure G4. Density profiles as in Fig. 9, for DF heating by single satellite, but for a steep-cusp initial host halo with $(c_{2h}, s_{1h}) = (5, 1.5)$ instead of NFW. Here, in order to obtain a significant effect, the initial satellite mass is $m_v = 0.1 \mu M_v$ with $\mu = \sqrt{3}$ and the outflow is with $\eta = 1$, namely involving all the available gas of $10^{10.6} M_\odot$ (compared to $\mu = \sqrt{2}$ and $\eta = 0.5$ in Fig. 9. The DF heating is based on the energy deposit profile by SatGen, bottom-left panel of Fig. G2, which turns out to be comparable to the energy deposit of the NFW host. We learn that the steep cusp is more resilient than the NFW cusp both to DF heating and to outflows, requiring more massive satellite and outflow for generating an extended core.

ity magnitude V_{in} and the circularity $\epsilon = V_{tan}/V_v$. For reference, the orbit eccentricity is $e^2 = 1 - \epsilon^2$, and the corresponding spin parameter is

$$\lambda = \frac{V_{tan} R_v}{\sqrt{2} V_v R_v} = \frac{\epsilon}{\sqrt{2}}, \qquad \text{(I1)}$$

independent of V_{in}. The orbit, and the effects of dynamical friction and tidal stripping, depend in addition on V_{in}, which for $M_v \sim 10^{12} M_\odot$ at $z \sim 2$ is roughly $V_{in}/V_v \simeq 1.15 \pm 0.15$ (Wetzel 2011, Figs. 2,5,9).

According to the cosmological N-body simulations of Wetzel (2011), Figs. 5 and 8, for minor mergers of $m_v/M_v \sim 0.02$, ϵ tends to decrease with increasing redshift, where the average is $\langle \epsilon \rangle \simeq 0.55$ and 0.45 at

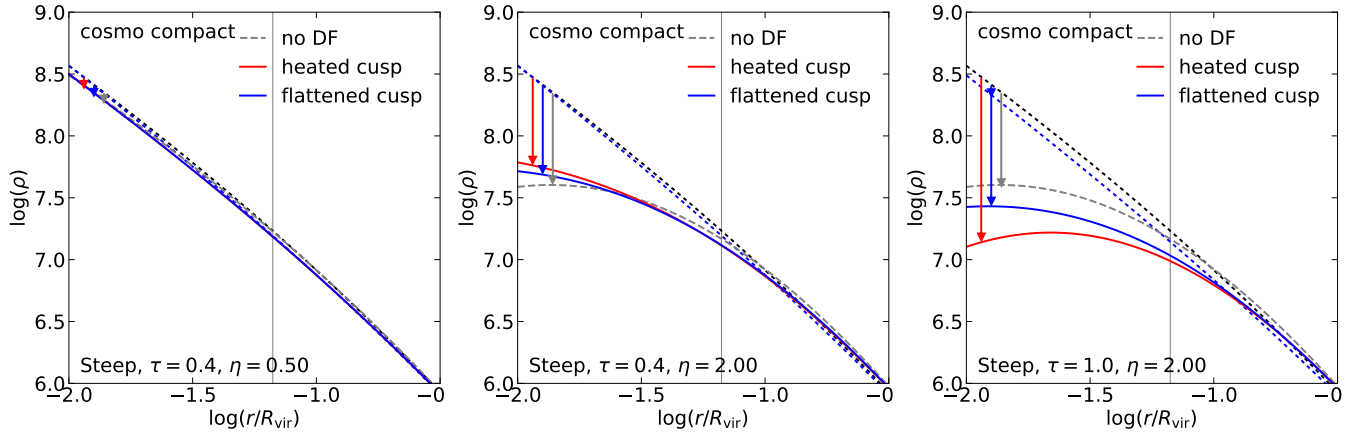

Figure G5. Same as Fig. 10, for DF heating by a cosmological sequence of satellites, but for a steep-cusp host halo with $(c_{2h}, s_{1h}) = (5, 1.5)$ instead of NFW. The energy deposited in a virial time is based on the fit to the SatGen run shown in Fig. G3. **Left:** Compact satellites and outflows with the parameters that produced a core in the NFW cusp cause a negligible effect on the steep cusp. **Middle:** An outflow of $\eta = 2$ is sufficient for forming a moderate core without DF heating. In this case, the DF actually steepens the profile, making it a little harder to produce a core by inflow. **Right:** Similar but using a slightly flatter slope of $K(<r)$ than produced by SatGen within the cusp, which enables a better fit to the DZ profile and thus a convergence of CuspCore, leading to a flat core after DF heating with $\tau = 1$. The higher value of τ can be interpreted as roughly representing the top 1/3 of the random realizations drawn from the mass function and the circularity distribution of satellites during $0.4\,t_v$. Alternatively it can be interpreted as the median energy during t_v or or a longer duration.

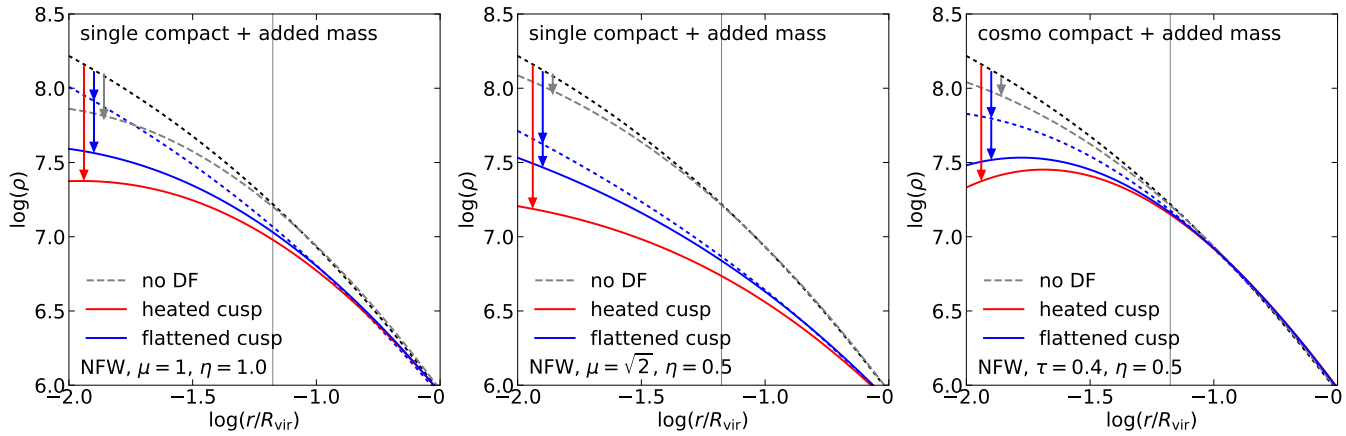

Figure H1. The effect of added satellite mass to the cusp. Same as Fig. 9 (left and middle) and Fig. 10, for DF heating by a single compact satellite or a sequence of compact satellites, in an NFW host, but with the mass of the satellite added to the host cusp where it is stripped or at the center, based on the bottom-left panel of Fig. 4 and the right panel of Fig. 5. The difference from the results obtained without this additional mass in Fig. 9 and Fig. 10 is small. For the single satellite, the satellite mass deposited in the cusp is comparable to the cusp mass, slightly steepening the final core. For the sequence of satellites, the satellite mass is only $0.2M_c$, causing almost no change to the final core. This justifies ignoring the added mass in our main analysis.

$z = 0$ and 2.5, respectively. The corresponding median of the pericenter of the orbit is roughly $\langle r_{peri} \rangle / R_v \simeq 0.24$ and 0.17, respectively, namely a deeper penetration at higher redshifts. The distribution of ϵ is found to be approximately universal for a given host halo mass when measured with respect to the non-linear Press-Schechter mass $M_{ps}(z)$. For a given satellite mass, ϵ and r_{peri} tend to decrease with increasing host halo mass. These redshift and mass dependencies are in the desired sense, but they are rather mild, possibly not sufficient by themselves for explaining the redshift and mass dependencies of the DM core phenomenon.

A qualitatively similar redshift dependence is obtained from hydro cosmological simulations, via the analysis of the angular momentum carried by the cosmic-web cold streams that build the galaxies at high redshift, and contain the incoming satellites (Danovich et al. 2015). According to their figure 15, the dominant stream carries on average 84% of the angular-momentum inflow rate, and 64% of the mass influx. In order to relate the measured spin parameter to measured eccentricity, we consider for an upper limit only one dominant stream, and obtain $\lambda \sim \epsilon/\sqrt{2} \sim 0.35$ for $\epsilon = 0.5$. For a lower limit, we consider three comparable streams with random orientation and impact parameter, and obtain $\lambda \sim (\epsilon/\sqrt{2})/\sqrt{3} \sim 0.2$ for $\epsilon = 0.5$. We can

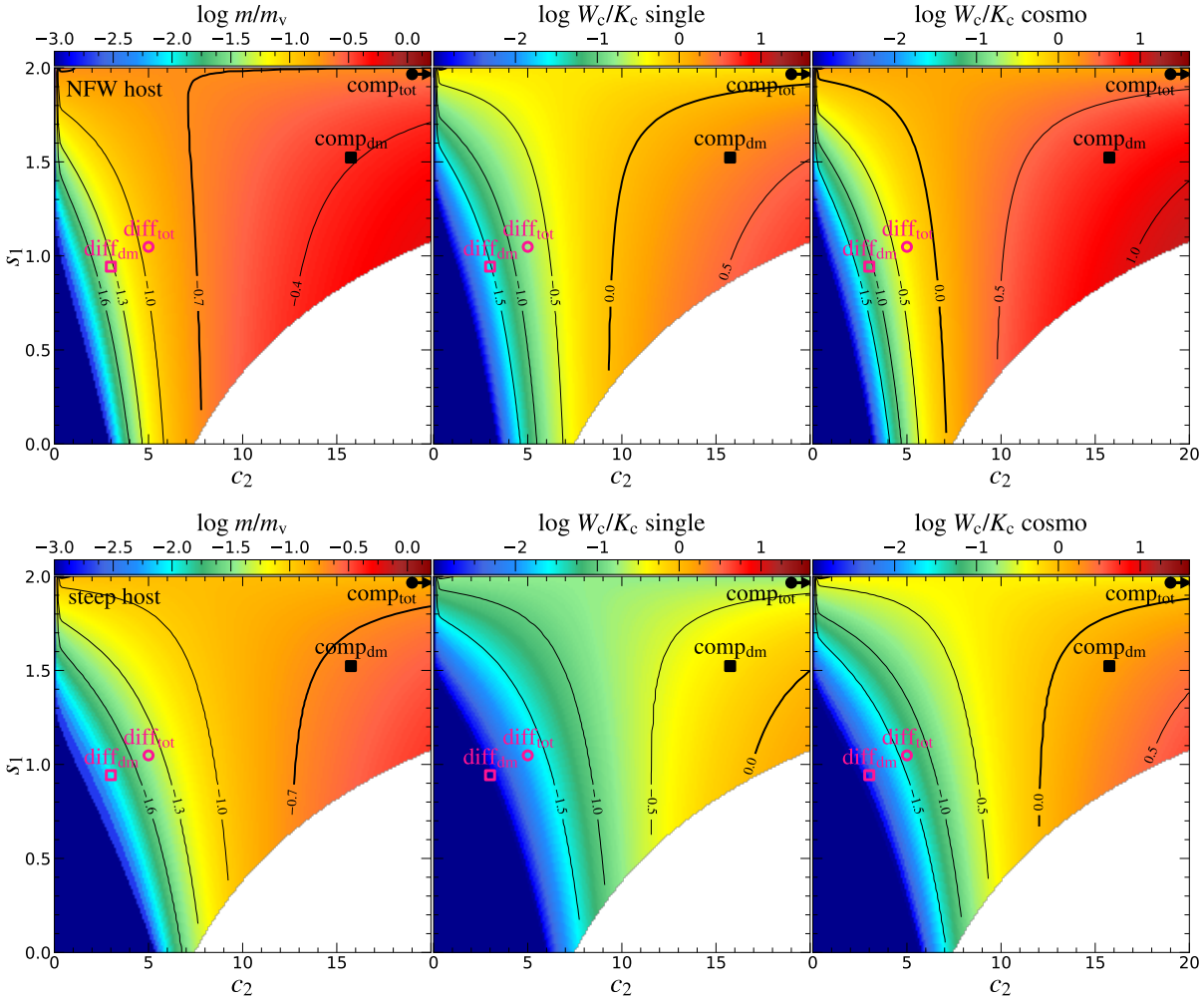

Figure I1. Toy-model estimates for satellite penetration and energy deposited in the host cusp by dynamical friction, similar to Fig. 3, but for the more accessible parameters (c_2, s_1) instead of the natural Dekel-Zhao parameters (c, α). The concentration parameter c_2 refers to the radius where the local log slope of the density profile is -2 (as in the concentration of the NFW profile), and s_1 is the minus the inner local log slope at $r = 0.01R_v$. The transformation between the two alternative pairs of parameters is given in eq. (A12) to eq. (A15). While there is a valid DZ profile for any values of c (>0) and α (<3), a valid profile is not guaranteed for arbitrary values of c_2 and s_1. For example, $c_2 \to \infty$ for $\alpha = 2$. The case of compact total $(c, \alpha) = (3, 1.8)$ has $(c_2, s_1) = (170, 2.05)$, which is outside the box of this figure. We truncate the plot where $c > 100$ (or where $\alpha \geqslant 2$).

therefore assume that the mean eccentricity measured by Wetzel (2011), $\epsilon \sim 0.5$, would typically correspond to $\lambda \sim 0.3$. However, from Danovich et al. (2015, Fig. 1), at R_v, we read for the dark matter that $\lambda \simeq 0.13$ and it is not varying with redshift, while for the cold gas $\lambda \simeq 0.3, 0.2, 0.15$ at $z = 1.5, 2.5.3.5$, respectively. Smaller λ values are measured at higher redshifts also in their Fig. 7. Similar results are obtained for cold gas at $z = 1.6-3$ in their Fig. 14, where $\lambda \sim 0.2$. This indicates that at $z \sim 2.5$ one should assume $\epsilon \sim 0.25$ for the dark matter and $\epsilon \sim 0.33$ for the cold gas. These values are lower than the average value obtained for satellites by Wetzel (2011), indicating more radial orbits and thus stronger dynamical friction. The redshift dependence in Danovich et al. (2015) for the cold gas is stronger than in Wetzel (2011), but this may be balanced by the weaker redshift dependence for the dark matter.

Qualitatively similar redshift and mass dependencies can be deduced from the analysis of random Gaussian fluctuation fields by Bardeen et al. (1986), who predict that higher-sigma density peaks have lower λ values and more radial orbits. This is consistent with the trends found in Wetzel (2011) and Danovich et al. (2015), and with the core phenomenon being more pronounced at higher redshifts and masses, being higher-sigma peaks.

REFERENCES

Agertz O., Kravtsov A. V., Leitner S. N., Gnedin N. Y., 2013, ApJ, 770, 25
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Barro G. et al., 2017, ApJ, 840, 47
Benson A. J., 2017, MNRAS, 467, 3454

Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Ceverino D., Dekel A., Bournaud F., 2010, MNRAS, 404, 2151
Ceverino D., Dekel A., Mandelker N., Bournaud F., Burkert A., Genzel R., Primack J., 2012, MNRAS,
Ceverino D., Dekel A., Tweed D., Primack J., 2015, MNRAS, 447, 3291
Ceverino D., Klypin A., 2009, ApJ, 695, 292
Ceverino D., Klypin A., Klimek E. S., Trujillo-Gomez S., Churchill C. W., Primack J., Dekel A., 2014, MNRAS, 442, 1545
Ceverino D., Primack J., Dekel A., 2015, MNRAS, 453, 408
Chandrasekhar S., 1943, ApJ, 97, 255
Danovich M., Dekel A., Hahn O., Ceverino D., Primack J., 2015, MNRAS, 449, 2087
Dekel A., Birnboim Y., 2006, MNRAS, 368, 2
Dekel A., Ginzburg O., Jiang F., Freundlich J., Lapiner S., Ceverino D., Primack J., 2020a, arXiv e-prints
Dekel A., Ishai G., Dutton A. A., Maccio A. V., 2017, MNRAS, 468, 1005
Dekel A., Krumholz M. R., 2013, MNRAS, 432, 455
Dekel A., Lapiner S., Ginzburg O., Jiang F., Ceverino D., Primack J., 2020b, arXiv e-prints
Dekel A., Sarkar K. C., Jiang F., Bournaud F., Krumholz M. R., Ceverino D., Primack J. R., 2019, arXiv e-prints
Dekel A., Zolotov A., Tweed D., Cacciato M., Ceverino D., Primack J. R., 2013, MNRAS, 435, 999
Dutton A. A. et al., 2016, MNRAS, 461, 2658
Eddington A. S., 1916, Monthly Notices of the Royal Astronomical Society, 76, 572
Ferland G. J., Korista K. T., Verner D. A., Ferguson J. W., Kingdon J. B., Verner E. M., 1998, PASP, 110, 761
Freundlich J., Dekel A., Jiang F., 2019, in SF2A-2019: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, Di Matteo P., Creevey O., Crida A., Kordopatis G., Malzac J., Marquette J. B., N'Diaye M., Venot O., eds., p. Di
Freundlich J., Dekel A., Jiang F., Ishai G., Cornuault N., Lapiner S., Dutton A. A., Macciò A. V., 2020a, MNRAS, 491, 4523
Freundlich J. et al., 2020b, MNRAS, 499, 2912
Green S. B., van den Bosch F. C., Jiang F., 2021, Monthly Notices of the Royal Astronomical Society, 503, 4075
Haardt F., Madau P., 1996, ApJ, 461, 20
Hopkins P. F., Quataert E., Murray N., 2012, MNRAS, 421, 3522
Inoue S., Dekel A., Mandelker N., Ceverino D., Bournaud F., Primack J., 2016, MNRAS, 456, 2052
Jiang F., Dekel A., Freundlich J., van den Bosch F. C., Green S. B., Hopkins P. F., Benson A., Du X., 2020, arXiv e-prints, arXiv:2005.05974
Jiang F. et al., 2019, MNRAS, 488, 4801
Kazantzidis S., Zentner A. R., Kravtsov A. V., 2006, Astrophys. J., 641, 647
Komatsu E., Dunkley J., Nolta M. R., Bennett C. L., Gold B., Hinshaw G., Jarosik N., et al., 2009, ApJS, 180, 330
Kravtsov A. V., 2003, ApJ, 590, L1
Kravtsov A. V., Klypin A. A., Khokhlov A. M., 1997, ApJS, 111, 73
Krumholz M. R., Thompson T. A., 2013, MNRAS, 434, 2329
Mandelker N., Dekel A., Ceverino D., DeGraf C., Guo Y., Primack J., 2017, MNRAS, 464, 635
Mandelker N., Dekel A., Ceverino D., Tweed D., Moody C. E., Primack J., 2014, MNRAS, 443, 3675
Moody C. E., Guo Y., Mandelker N., Ceverino D., Mozena M., Koo D. C., Dekel A., Primack J., 2014, MNRAS, 444, 1389
Murray N., Quataert E., Thompson T. A., 2010, ApJ, 709, 191
Parkinson H., Cole S., Helly J., 2008, MNRAS, 383, 557
Penarrubia J., Benson A. J., Walker M. G., Gilmore G., McConnachie A. W., Mayer L., 2010, MNRAS, 406, 1290
Roca-Fàbrega S. et al., 2018, arXiv e-prints
Rodríguez-Puebla A., Primack J. R., Avila-Reese V., Faber S. M., 2017, MNRAS, 470, 651
Snyder G. F., Lotz J., Moody C., Peth M., Freeman P., Ceverino D., Primack J., Dekel A., 2015, MNRAS, 451, 4290
Springel V., Hernquist L., 2005, ApJ, 622, L9
Strawn C. et al., 2020, MNRAS
Tacchella S., Dekel A., Carollo C. M., Ceverino D., DeGraf C., Lapiner S., Mandelker N., Primack J. R., 2016a, MNRAS, 458, 242
Tacchella S., Dekel A., Carollo C. M., Ceverino D., DeGraf C., Lapiner S., Mandelker N., Primack Joel R., 2016b, MNRAS, 457, 2790
Tomassetti M. et al., 2016, MNRAS, 458, 4477
van den Bosch F. C., 2017, MNRAS, 468, 885
van den Bosch F. C., Ogiya G., 2018, MNRAS, 475, 4066
Wetzel A. R., 2011, MNRAS, 412, 49
Zhao D. H., Jing Y. P., Mo H. J., Börner G., 2009, ApJ, 707, 354
Zhao H., 1996, MNRAS, 278, 488
Zolotov A. et al., 2015, MNRAS, 450, 2327