Dark halo microphysics and massive black hole scaling relations in galaxies

Curtis J. Saxton,1,2* Roberto Soria3* and Kinwah Wu1*

1 Mullard Space Science Laboratory, University College London, Holmbury St Mary, Surrey RH5 6NT, UK
2 Physics Department, Technion - Israel Institute of Technology, Haifa 32000, Israel
3 International Centre for Radio Astronomy Research, Curtin University, GPO Box U1987, Perth, WA 6845, Australia

ABSTRACT
We investigate the black hole (BH) scaling relation in galaxies using a model in which the galaxy halo and central BH are a self-gravitating sphere of dark matter (DM) with an isotropic, adiabatic equation of state. The equipotential where the escape velocity approaches the speed of light defines the horizon of the BH. We find that the BH mass \(m_\bullet\) depends on the DM entropy, when the effective thermal degrees of freedom \(F\) are specified. Relations between BH mass and galaxy properties arise naturally, with the BH mass and DM velocity dispersion following \(m_\bullet \propto \sigma^{F/2}\) (for global mean density set by external cosmogony). Imposing observationally derived constraints on \(F\) provides insight into the microphysics of DM. Given that DM velocities and stellar velocities are comparable, the empirical correlation between \(m_\bullet\) and stellar velocity dispersions \(\sigma_\star\) implies that \(7 \lesssim F < 10\). A link between \(m_\bullet\) and globular cluster properties also arises because the halo potential binds the globular cluster swarm at large radii. Interestingly, for \(F > 6\) the dense dark envelope surrounding the BH approaches the mean density of the BH itself, while the outer halo can show a nearly uniform kpc-scale core resembling those observed in galaxies.

Key words: black hole physics – globular clusters: general – galaxies: haloes – galaxies: kinematics and dynamics – galaxies: structure – dark matter

1 INTRODUCTION
Over the last two decades, empirical correlations between different galaxy components – nuclear supermassive black hole (SMBH), stellar bulge and disc, dark matter (DM) halo – have shaped our understanding of galaxy structure evolution and of SMBH/galaxy co-evolution (Kormendy & Ho 2013, for a review). The most significant correlations are the ones observed between SMBH masses \((m_\bullet)\) and velocity dispersions \((\sigma)\) of their host stellar bulges or spheroids \((m_\bullet-\sigma)\) relation; Ferrarese & Merritt 2000; Gebhardt et al. 2000; Tremaine et al. 2002; Graham et al. 2011; Xiao et al. 2011); and the ones between SMBH masses and bulge masses (Magorrian et al. 1998; Laor 2001; Haring & Rix 2004; Graham & Scott 2013; Scott, Graham & Schombert 2013). The kinetic or potential energy of the bulge also correlates with \(m_\bullet\) (Feoli & Mele 2005, 2007; Aller & Richstone 2007; Hopkins et al. 2007; Feoli & Mancini 2009; Mancini & Feoli 2012; Benedetto, Fallarino & Feoli 2013), as does the momentum-like quantity \(M_\bullet \sigma / c\), where \(M_\bullet\) is the bulge stellar mass (Lahav, Meiron & Soker 2011; Soker & Meiron 2011).

This leads to the proposal of a ‘BH Fundamental Plane’ with \(m_\bullet\) depending on two input quantities (Marconi & Hunt 2003; Barway & Kembhavi 2007; Hopkins et al. 2007b). The correlation may also take other forms, such as a dependence on the Sérsic (1968) shape index of the stellar profile (Graham et al. 2001; Graham & Driver 2007; Savorgnan et al. 2013). More recently, it was also found (Burkert & Tremaine 2010; Harris & Harris 2011; Harris, Harris & Alesi 2013) that the total number of globular clusters (GCs) in a galaxy correlates with the SMBH mass and with the dynamical mass \(M_{\text{dyn}} = 4R_\epsilon \sigma_\star^2 / G\) of the spheroidal component, where \(R_\epsilon\) is the effective radius enclosing half of the galaxy light, and \(\sigma_\star\) is the stellar velocity dispersion. The specific number of GCs of galaxies is not a fundamental physical property, but it is a useful proxy for the total stellar mass contained in the GCs.

An interpretation of this finding (Snyder, Hopkins & Hernquist 2011) is that both the total mass of GCs and the SMBH mass correlate with the host spheroid’s binding energy \(E_b \sim M_{\text{dyn}} \sigma_\star^2\) (see also Aller & Richstone 2007; Hopkins et al. 2007a,b). Hence, the total number of GCs and the SMBH mass also show a correlation with each other.

Taking all these empirical correlations together points to the presence of a general scaling relation between SMBH mass, stellar mass
in the host spheroid and the total mass/number of GCs in the galaxy. This scaling appears straightforward to understand, at least qualitatively. Rapid growth of the nuclear BH of a galaxy (particularly at redshifts $2 \lesssim z \lesssim 6$) might be fuelled by a massive inflow of cold gas towards the centre of the galaxy. The gas inflow would trigger starbursts and the formation of new GCs. Numerical simulations often show massive gas inflows in mergers of gas-rich galaxies (e.g. Hernquist 1989; Barnes & Hernquist 1991, 1996; Hopkins et al. 2005). This scenario expects a coeval growth of SMBH and stellar components (which includes the spheroid and GCs), regulated by the gas supply that reaches the inner region of the galaxy, and ultimately primed by the merger rate (Volonteri & Natarajan 2009). The parallel growth of the SMBH and stellar component cannot continue indefinitely, and it terminates when the gas supply ceases. The accretion into a BH at super-Eddington rates will emit copious radiation, which exerts radiative pressure on the inflowing gas, leading to a massive galactic-scale outflows. When the central BH in a galaxy has grown to a sufficiently large mass (and can therefore attain a sufficiently high Eddington luminosity), the momentum-driven, expanding shell of the swept-up gas will achieve a velocity higher than the escape velocity from the galaxy (Silk & Rees 1998; King 2003; Murray, Quataert & Thompson 2005). When most of the gas is expelled, star formation and SMBH accretion are quenched.

However, the reality could be more complicated than described above, as there is evidence that SMBH accretion and star formation do not always trace each other (Zheng et al. 2009). Thus, there could be pathways (or even multiple pathways) of SMBH and spheroid growth without invoking self-regulation (e.g. Anglés-Alcázar, Özel & Davé 2013) that lead to the SMBH scaling relations that we observe today (Zheng 2013). It worth noting that the duration of SMBH growth in the co-evolution scenario depends on the initial mass of their seed BHs. Some authors (e.g. Shibata & Shapiro 2002; Volonteri & Madau 2008; Begelman 2010) argued that seed BHs may come from direct collapse of supermassive stars, which were formed directly from large-scale gas inflows in the DM halo. As such, the seed BH mass distribution would be a function of the DM halo virial temperature and the BH spin. Also, there would be an angular momentum ceiling for the DM halo, only below which inflows can occur and supermassive stars can form.

The existence and nature of a correlation between GCs, SMBH, and dark halo is not free from disputes. GCs have a bimodal colour distribution, probably the signature of two physically distinct populations: younger, metal-rich red and older, metal-poor blue clusters (Brodie & Strader 2006). Co-evolution of stellar populations and SMBH due to major mergers should produce a correlation only between red GCs (formed during the starburst phase and located closer to the nucleus) and SMBH (Kormendy & Ho 2013). The correlation is indeed tighter for red GCs (Sadoun & Colin 2012), but the fraction of red/blue GCs is similar for most galaxies (Burkert & Tremaine 2010), indicating some residual correlation also with the blue (old) population, or perhaps an initial correlation between blue GCs and seed BH. Intriguingly, it was recently noted (Harris et al. 2013) that the relation between GC mass fraction (i.e. fraction of a galaxy mass that is contained in GCs) and galaxy mass is not a constant but has a characteristic U-shape: both dwarf and giant ellipticals have a larger fraction of baryonic mass located in GCs, than intermediate-mass galaxies. This could be due to different rates of GC formation or subsequent GC destruction. Alternatively, perhaps dwarf and giant galaxies have formed field stars less efficiently, owing to gas losses from superwinds and SMBH activity, respectively. Only in intermediate-mass systems is the observed GC mass fraction a true indication of how much gas was initially present in the galaxy potential well. Proponents of collisionless cold DM theories also invoke a scenario of gas blowouts to explain the differences between the simulated halo mass spectrum and the visible baryonic mass function, especially at the low-mass and high-mass ends (e.g Persic & Salucci 1992; Bell et al. 2003; Read & Treurnth 2005; Papastergis et al. 2012). Either way, it follows that the mass in GCs is determined by the amount of gas initially present in the (DM-dominated) potential well of a galaxy, and therefore there must be some correlation between GC mass and DM halo mass (Georgiev et al. 2010; Harris et al. 2013). In particular, Harris et al. (2013) propose a linear correlation with $M_{\text{DM}} \approx 6 \times 10^{-5} M_{\text{halo}}$.

In summary, there are empirical hints of correlations between SMBHs, DM haloes and GCs in galaxies despite the widely different scales of the three types of objects, but it is still not clear to what extent the associations are truly intrinsic or they are mere by-products of other physical processes, such as galaxy mergers. In this work, we search for physical processes that could give rise to such correlations and demonstrate a physical mechanism that naturally links the properties of the SMBH, DM haloes and GCs. Motivated by the extent of GC swarms – rounded and far from the direct reach of active galactic nuclei (AGN) in normal galaxies – we seek explanations in which the DM halo is the component controlling the scaling relations.

For galaxies that are large or small; rich or poor in baryons; pristine, star-forming or aged, observations indicate that DM haloes feature a kpc-scale central core of nearly uniform density, surrounded by outskirts where the density declines radially till it becomes unmeasurable at $\sim 100$ kpc distances (e.g. Flores & Primack 1994; Moore 1994; Burkert 1995; Salucci & Burkert 2000; Kelson et al. 2002; Kleva et al. 2003; Simon et al. 2003; Gentile et al. 2004; de Blok 2005; Thomas et al. 2005; Goerdt et al. 2006; Kuzio de Naray et al. 2006; Gilmore et al. 2007; Oh et al. 2008; Weijmans et al. 2008; Donato et al. 2009; Inoue 2009; de Blok 2010; Pu et al. 2010; Memola, Salucci & Babić 2011; Murphy, Gebhardt & Adams 2011; Richtler et al. 2011; Walker & Peñarrubia 2011; Agnello & Evans 2012; Amorisco & Evans 2012; Jardel & Gebhardt 2012; Lora et al. 2012, 2013; Salucci et al. 2012; Schubert et al. 2012; Amorisco, Agnello & Evans 2013; Hague & Wilkinson 2014). However, early theories of collisionless and non-interacting DM predicted steep power-law central density cusps and not the observed cores (e.g. Gurevich & Zybin 1988; Dubinski & Carlberg 1991; Navarro, Frenk & White 1996). More microphysics may be needed. The core sizes can be set by the effective thermal properties of DM in equilibrium (Nunez et al. 2006; Saxton & Ferreras 2010; Saxton 2013), or temporally by heat conduction (Kochanek & White 2000; Davé et al. 2001; Ahn & Shapiro 2005; Rocha et al. 2013). Alternatively, one may invoke stellar and/or SMBH feedback (see e.g. Navarro, Eke & Frenk 1996; Mashchenko, Couchman & Wadsley 2006). While it is conceivable to mechanically shake the halo to create a uniform core, it requires certain parameter fine tuning in the feedback approach, which is not always feasible in certain classes of galaxies (Gnedin & Zhao 2002; Peñarrubia et al. 2012). Here, we advance a more generic theory that is independent of episodic astrophysical events, by attributing the halo structure to the innate microphysics of DM. Studies (see e.g. Ullio, Zhao & Kamionkowski 2001; MacMillan & Henriksen 2002; Merritt 2004; Zakharov et al. 2007; Ghez et al. 2008; Saxton & Wu 2008; Zakharov et al. 2010) have shown the central density profile can rise locally in a sharp spike in the sub-pc to pc-scale gravitational sphere of influence around the SMBH. Our paper builds upon this finding, allowing a direct material coupling between the halo and SMBH, with a smooth transition from a DM density spike around the horizon of the SMBH to a cored DM halo at
galaxy scales, that in turns binds the swarm of GCs located at larger distances. We consider constraints at both scales, and show how SMBH–GC relations emerge from the SMBH–halo connection.

We organize the paper as follows. Section 2 presents the model and the formulation. Section 3 shows the solutions, and in Section 4 we discuss the astrophysical implications.

2 MODEL AND FORMULATION

2.1 Halo model properties

We assume a self-gravitating, spherically symmetric and structurally stationary DM halo, in dynamical equilibrium, with the thermodynamics of the DM described in terms of a polytropic equation of state. The DM in the halo is well mixed, without sub-halo clump structures. The SMBH develops from within the DM halo as an integral part of a self-gravitating system, instead of being inserted artificially into the halo centre as a massive external point-like object. Moreover, the BH has a physically defined horizon directly interfacing with the surrounding DM in the halo core. The gravity in the system is dominated by the DM components, i.e. the halo and SMBH, with insignificant contribution by the baryonic components, i.e. gas, stars and globular GCs.

2.2 Equation of state of the DM

The equation of state of the DM takes the form

\[ P = \rho \sigma^2 = s \rho' , \]  

or equivalently

\[ \frac{\rho}{Q} = \frac{s}{\sigma^2} , \]  

where \( P \) is the pressure, \( \rho \) is the density and \( \sigma \) is the isotropic velocity of the particles. The quantity \( s \) is the (pseudo-)entropy, and \( Q \equiv s^{-1/2} \) is the phase-space density. The adiabatic index \( \gamma \) is determined by the DM microphysics. It is related to the effective thermal degrees of freedom of the dark particles \( F \) via

\[ \gamma = 1 + \frac{2}{F} . \]  

Many DM scenarios entail a functionally equivalent equation of state (Section 4.4). Generally, \( F \) describes the number of modes in which the microscopic energies of DM particles can be equipartitioned. For translational motions in three dimensions, \( F = 3 \). When self-interacting DM particles are composite or have internal structure and modes of rotation, vibration and excitation at a comparable energy scale (Cline et al. 2014b), then \( F > 3 \). The specific heat capacity at constant volume is \( c_v \equiv F k/2 \) (where \( k \) is Boltzmann’s constant) and the energy density is \( FP/2 \). If instead DM is a sterile neutrino, then \( F \geq 3 \) in the degenerate halo core (cf. neutrino-ball SMBH; Viollier, Trautmann & Tupper 1993). If DM is a boson scalar field, then \( F \) derives from the index of the self-coupling potential (Peebles 2000). If DM experiences phase changes, then the equation of state is more complicated, but a polytropic law would remain a fair working approximation in limited ranges of temperature and density.

In principle, \( \gamma \) and \( s \) vary radially, if the halo is stratified, e.g. due to a history of mergers and accretion, or if dynamically significant energy exchange processes are present (e.g. ‘dark radiation’; Ackerman et al. 2009; Fan et al. 2013). In that case, buoyant stability could appear, when \( \rho' / \rho > 0 \) and \( (F \rho') / (\rho') < 0 \). However, we have assumed that the adiabatic DM in the halo does not have sub-structures. Thus, \( Q \) and \( s \) are constant for each galaxy in our calculations.

For \(-2 < F < 10 \), the outer radius of the halo and total mass enclosed are finite, safeguarding the existence of realistic solutions for the DM halo–SMBH system. In this paper, we discard models with \( F < 0 \), since they have minimum density at the centre and greatest densities outside (which seems inappropriate for galaxies). Isolated polytropes with \( F > 6 \) are sometimes susceptible to interesting dynamical instabilities (Ritter 1878; Emden 1907; Chandrasekhar 1939). The instability can none the less be moderated by interactions with the baryonic matter components (Saxton 2013) or by a confining external pressure (e.g. McCrea 1957; Bonnor 1958; Horedt 1970; Umemura & Ikeuchi 1986).

2.3 Halo profile

A realistic halo requires that the density \( \rho \) falls to zero at a certain outer radius \( R \), which defines the size of the halo. The mass enclosed by \( R \) is the total mass \( M \) of the halo. Inside the halo, the mass distribution is the solution to

\[ \frac{dm(r)}{dr} = 4\pi r^2 \rho (r) , \]  

where \( m(r) \) is mass contained within radius \( r \). The gravitational field strength is given by

\[ g(r) = -\frac{Gm(r)}{r^2} \]  

and the gravitational potential \( \Phi(r) \) by

\[ \frac{d\Phi(r)}{dr} = -g(r) . \]  

The escape velocity \( v(r) \) satisfies the relation

\[ \frac{dv(r)^2}{dr} = 2g(r) . \]  

If the pressure in the halo were deficient near a central gravitating mass, adiabatic accretion would proceed (Bondi 1952), which, conceivably, feeds the growth of the SMBH (e.g. Peirani & de Freitas Pacheco 2008; Guzmán & Lora-Clavijo 2011a,b; Pepe, Pelizza & Romero 2012; Lora-Clavijo, Gracia-Linares & Guzman 2014). Without losing generality, we ignore the complications of accretion inflow and focus on the stationary halo, which is pressure-supported everywhere. Under these conditions, the velocity dispersion of DM is then given by

\[ \frac{dv(r)^2}{dr} = \frac{2}{F + 2} g(r) . \]  

The DM velocity dispersion \( v^2(r) \) can be considered as a measure of the local thermal ‘temperature’. Within the halo, this thermal temperature is related to the local escape velocity and gravitational potential by

\[ \sigma(r)^2 = \frac{1}{F + 2} \left[v(r)^2 - V^2\right] , \]  

where \( V \) is the escape velocity at the outer boundary of the halo \( r = R, \rho = 0, \sigma = 0 \). The above expression can be obtained by carrying out an integration after combining equations (7), (6) and (8). The escape velocity \( V \) depends on whether there is any non-DM material extending beyond the outer halo radius \( R \). Otherwise, it takes the value \( V = \sqrt{2GM/R} \).
Given either the inner or outer boundary conditions, locating the other boundary is performed by numerical integration (Section 2.5). We define a dimensionless gravitational compactness parameter:

\[
\chi \equiv \left(\frac{V}{c}\right)^{2} = \frac{2GM}{c^{2}R} \ll 1. \tag{10}
\]

Empirical values of \(\chi\) could be estimated from a characteristic velocity dispersion or mass–radius relation of self-bound objects. For example, massive galaxy clusters have \(\chi \lesssim 10^{-8} \quad (V \lesssim 3000 \text{ km s}^{-1})\); giant galaxies have \(\chi \lesssim 10^{-6} \quad (V \lesssim 300 \text{ km s}^{-1})\); and faint dwarf galaxies have \(\chi \lesssim 10^{-8} \quad (V \lesssim 30 \text{ km s}^{-1})\).

### 2.4 Central BH and horizon surface

Most galaxies are expected to possess a central BH, but observations indicate that some actually do not. In some cases, there might never occur a mass concentration dense enough to collapse gravitationally. Effects such as rotational support might help avert BH formation in certain late-type galaxies (Section 4.7). Also, merger events could eject a SMBH from the host galaxy. Here however, we investigate only galaxies that have formed a nuclear SMBH and retain it in equilibrium with its DM surroundings.

The escape velocity of a test mass is \(c\), the speed of light, at the event horizon of a (Schwarzschild) BH. If it is appropriate to consider the ‘formation’ of a BH in this Newtonian model, then the BH is defined by the sphere where the escape velocity is \(v = c\) at its surface (i.e. the horizon). This BH contains a mass \(m_{\rho}\), inside a horizon radius, which is given by \(r_{h} \approx 2GM_{h}/c^{2}\). In a dense DM envelope enclosing the central BH, the horizon radius is larger than the ideal Schwarzschild value in vacuum. We parametrize the ratio between the horizon radius and the Schwarzschild radius \(r_{h}\), by \(\eta \equiv r_{h}/r_{s}\). In a fully relativistic treatment, \(\eta = 1\) always. Here, the value of \(\eta\) is generally of the order unity. The mean density of the BH is then

\[
\bar{\rho}_{s} \equiv \frac{3m_{s}}{4\pi r_{s}^{3}} = \frac{3c^{6}}{32\pi G^{2}m_{h}^{2}\eta^{3}}.
\]

The velocity dispersion of the DM at the horizon surface is

\[
\sigma_{s} = \left(1 - \frac{\chi}{F + 2}\right) c^{2}. \tag{11}
\]

If the halo is adiabatic all the way down to the horizon surface of the central BH, then from the equation of state (2) we obtain

\[
\rho(r) = \frac{\rho_{s}}{\sigma_{s}^{2}} = \left[\frac{(\sigma(r))^{2}}{\sigma_{s}^{2}}\right]^{F/2}. \tag{12}
\]

Define a parameter \(\psi \equiv \rho_{s}/\rho_{*}\), which is the density ratio of the BH to DM near its horizon surface. Then, we have

\[
\begin{align*}
m_{*} &= \sqrt{\frac{3c^{6}}{32\pi G^{3}}} \left(1 - \frac{F}{1 - \chi}\right)^{F/4} , \\
&= \sqrt{\frac{3c^{6-F}}{32\pi G^{3}}} \left(1 - \frac{F}{1 - \chi}\right)^{F/4} \frac{1}{\sqrt{Q\psi}}. \tag{14}
\end{align*}
\]

Substituting any observed set of \((\rho, \sigma)\) values of the DM from elsewhere in the halo’s adiabatic region yields an estimate of the natural mass of the central compact object. The values of \(\rho\) and \(\sigma\) in the above expression are local. They can be constrained by the observations. The dimensionless correction factors \(\psi\) and \(\eta\) are, however, obtained by numerical solution of a particular halo model. In Section 3.3, we will show that, for the physically relevant models of polytropic DM haloes, the correction factors are moderate. Appendix A expresses the mass prediction (14) in absolute physical units.

### 2.5 Numerical integration scheme

The radial profiles of particular polytropic haloes are obtained from direct numerical integration of equations (1), (4) and (8). This is an initial value problem, with radius \(r\) as the independent variable, starting from either the inner boundary \((r = 0)\) or the outer boundary \((r = R)\). The phase-space density \(Q\), (pseudo-)entropy \(s\) and thermal degrees of freedom \(F\) are mutually consistent constants.

We adopt the embedded eighth-order Runge–Kutta Prince–Dormand method with ninth-order error estimate (Prince & Dormand 1981; Hairer, Nørsett & Wanner 2008)\(^1\) in our integration. When integrating outwards from known inner values, we can express the differential equations in their original form. When integrating inwards, each equation is multiplied by \(-1\), and \(-r\) is used as the independent variable. Near the inner and outer boundaries, the radius is not known a priori, but the velocity dispersion is known exactly \((\sigma = \sigma_{s}\) and \(\sigma = 0\), respectively). Near those limits, it is more desirable to adopt \(\sigma^{2}\) as the independent variable in the differential equations, i.e. re-expressing each quantity \(y\) the equations in the form of \(dy/\sigma^{2}\) or \(dy/d(\sigma^{-2})\).

In the numerical integration, we first consider small steps (but much larger than the round-off level) until it is appropriate to switch to another independent variable and then continue in the same integration mode. Doing so we can integrate accurately either from the outer boundary of the halo towards the SMBH horizon, or from the SMBH horizon to the outer boundary of the halo.

### 3 RESULTS

Before presenting the results of our DM halo–SMBH calculations, we briefly review the general properties of adiabatic self-gravitating polytropic spherically symmetric bodies. This class of spheroids has been investigated previously, but more often in the context of stars instead of larger spheroids such as galaxies or galaxy clusters. There are three sub-classes (Fig. 1) with these characteristics:

(i) **Non-singular**, with a zero density gradient at the centre. The density declines outward until reaching zero at a large radius \(R\). The Lane–Emden spheres are examples of these (Lane 1870; Emden 1907). Observations show that galaxy haloes often have a uniform density core with a profile resembling that of these polytropic spheres.

(ii) **Singular**, with a density spike around a massive nuclear object. Shallower density gradients further out resemble that of a galaxy core. The profile of the outer fringe is similar to that of the non-singular polytropic haloes.

(iii) **Terraced**, with the radial density profile alternating between power-law slopes and cores, nested inside each other. The centre is, however, singular. Medvedev & Rybicki (2001) studied terraced polytropes with \(F \approx 10\).

Non-singular polytropic spheroids can be obtained by setting variables according to inner boundary conditions and then integrating the system of profile differential equations outwards. In the context of DM halo–SMBH model considered in our paper,
this kind of spheroid does not provide a self-consistent description for the circumnuclear properties of the DM, as the central mass (SMBH) is absent. Singular and terraced polytropic spheroids can be obtained by integrating the profile ordinary differential equations inwards with the outer boundary conditions $M(R) = M$ and $\sigma(R) = 0$. The central singularity gives rise to the SMBH, with the horizon determined by setting the escape velocity equal to the speed of light.

3.1 Relevant solutions

The polytropes have a non-zero compact central mass surrounded by a density spike, where $\rho \sim r^{-2}$ (e.g. Huntley & Saslaw 1975; Quinlan, Hernquist & Sigurdsson 1995; Ullio et al. 2001). The gravitational potential is Keplerian near the origin, with $\Phi \sim r^{-1}$, and the velocity dispersion peaks in the same manner, i.e. $\sigma^2 \sim r^{-1}$. The escape velocity reaches $c$ at some sufficiently small radius.

In computing the radial profile, we set a fiducial outer radius, say $R = 1$, where $\sigma = 0$, and choose trial values of the total mass $M$. This implies a specific value for the compactness parameter $\chi$. Keeping these fixed, we test trial values of the phase-space density $Q$, and integrate the profile differential equations inwards. If the condition (12) is satisfied, then we record the conditions of that inner boundary, $(r_*, m_*, \eta, \psi)$. If the origin is reached, or if a condition of $m \leq 0$ is encountered at any $r > 0$, then no horizon for the central gravitating object is obtained, and the trial value of $Q$ is recorded as an unphysical case.

Fig. 2 depicts the radial profiles of solutions for haloes with $F = 9$ and compactness appropriate for a galaxy ($\chi = 10^{-6}$). The inner tip (left) of each curve locates a horizon $(r_*)$; the outer tip (corresponding to $R = 1$, which is by construction) is where the halo truncates itself. Taking the outer radius to be of the order of $R \sim 300 \text{kpc}$, the nearly uniform core has a radius $\sim 1 \text{kpc}$. The non-singular solution (black curve) is uniform at the origin. For a model near the non-singular limit (e.g. blue curve), the DM velocity dispersion (and density) rise at radii within a parsec. This is the sphere of gravitational influence of the central mass. In this particular solution, the dense DM envelope surrounding the horizon outweighs the influence of central object $(m_*)$ by almost an order of magnitude at radii $r \lesssim 10 r_\star$. The density contrast between the horizon and envelope is low ($\psi = 3.00$). The central mass fraction is $m_*/M \approx 3.24 \times 10^{-5}$, consistent with the observed ratios between the SMBH and their host galaxy haloes (e.g. under slightly different assumptions, the $m_*$ versus $M$ results of equations 4–7 and fig. 5 of Ferrarese 2002). For models with lower $Q$ (green and red curves), the inner dense-hot spike is radially larger, $m_*$ is heavier, and the halo core is more compact. At the opposite extreme (large $Q$), we have $m_* \to 0$ and obtain the biggest possible halo core. Its maximal mass and radius depend on $F$. When $F$ is smaller, the maximal core is wide and contains much of the halo mass. (For an incompressible fluid, $F = 0$, the entire halo is a core.) When $F$ is larger, the maximal core is radially smaller and is relatively lightweight.

3.2 Configuration space

Particular radial profiles can be obtained for choices of $(\chi, Q)$ across a two-dimensional configuration space at fixed $F$. This task can be wrapped within a root-finding routine or an amoeba-like minimizer, seeking a specific or optimal value of any desired property of the central object (e.g. $m_*/M$ or $\psi$). We explore the $(\chi, Q)$ plane numerically at high resolution. Fig. 3 maps the varying properties of the
Figure 3. Characteristics of the central massive object in a halo with $F = 9.5$, in terms of the compactness ($\chi$) and adjusted phase-space density ($q = QV^F / \bar{\rho}$). The upper panels show the density contrast at the horizon ($\psi = \bar{\rho} / \rho$); and the correction to Schwarzschild radius ($\eta = r_s / r_s$). Lower panels show the mass and radius fractions ($m_\star / M$ and $r_\star / R$).

central object, for $F = 9.5$ halo models. For clarity of presentation, the vertical axis value is a dimensionless adjusted version of the phase-space density

$$q \equiv QV^F / \bar{\rho}, \quad (15)$$

where $\bar{\rho} = 3M / 4\pi R^3$ is the mean density of the system, and $V$ is the surface escape velocity. The four panels show results for: the horizon density contrast ($\psi$); the horizon radius correction ($\eta$); the central object’s fractional mass ($m_\star / M$) and its radius ($r_\star / R$). Several distinct domains appear. The top-left panel labels these domains:

(i) **forbidden zone**: for sufficiently high $q$, there are no self-consistent solutions. The halo is too dense and cold to reach the assumed outer radius $R$;

(ii) **border zone**: for $q$ slightly below the forbidden zone, there is a thin domain of solutions with extremely high or low $\psi$ values (and steep gradients of $\partial \psi / \partial q$). The upper edge of the border is where the non-singular solutions occur ($m_\star / M \rightarrow 0$ and $r_\star / R \rightarrow 0$, which cannot describe a galaxy hosting an SMBH);

(iii) **moderate plateau**: if $6 < F < 10$, then there is a domain of $q$ values below the border, where $m_\star / M$ and $r_\star / R$ are small but finite (and astronomically significant). The envelope density is non-negligible compared to the mean density of the BH ($\psi \lesssim 100$). This plateau zone is more extensive in $q$ (or in $Q$) if $\chi$ is small (systems with low escape velocities). Viewed in $(\chi, q)$ or $(\chi, Q)$ planes, the plateau is roughly triangular. Terraced haloes occur here;

(iv) **valleys**: within the plateau, there are local minima in $\psi$, coinciding with spikes in $\eta$. This implies gradual density continuity between the BH and its immediate dark envelope. Valleys are more numerous for smaller $\chi$;

(v) **hole-dominated**: At $q$ values lower than the plateau zone, the BH mass becomes dominant, $m_\star / M \rightarrow 1$. The halo is relatively tenuous: $r_\star / R$ remains small. The BH is effectively decoupled from the density and pressure of its diffuse surroundings. The density contrast $\psi$ rises by orders of magnitude, and the gradient $\partial \psi / \partial q$ is steep.
The plateau and $\psi$-valleys only exist for haloes with $6 < F < 10$. For $0 \leq F \leq 6$, the transition from non-singular border to the hole-dominated domain is much narrower than 1 dex in $q$. For $6 < F < 10$, the plateau becomes wider with increasing $F$, until the plateau vanishes suddenly around $F = 10$ (infinite profiles including the classic Plummer 1911 model). Fig. 4 depicts the $\psi$ landscape of the plateau in the $(\chi, q)$ plane, for various equations of state ($F = 6, 7, 8$ and 9). The $\psi$-valleys are conspicuous diagonal stripes. The valleys are more numerous for greater $F$. For large-$F$ models, the $\psi$-valleys coincide with steps in the ratio $m_*/M$ (Fig. 3, compare left-hand panels). For lower $F$, the steps are less distinct (gradients $\partial(m_*/M)/\partial q$ and $\partial(m_*/M)/\partial \chi$ are steadier). Across most of the plateau, the $\psi$ contours (such as the valleys) are approximately parallel to contours of $m_*/M$. As noted below (Section 4.1.3), the $\psi$-valleys are locally energetically favoured states. Haloes in $\psi$-valleys have varied properties:

(i) in the valley at lowest $Q$, the BH is obese ($m_*/M \gtrsim 0.1$) with only a tenuous halo. This is unrealistic for a galaxy.

(ii) Valleys at intermediate $Q$: dark diagonals in Fig. 4. Running along these valleys keeps $m_*/M$ nearly constant, resembling a Magorrian et al. (1998) relation.

(iii) Valleys near the non-singular border have branches and irregular $\psi$-topography. Values of $m_*/M$ are lowest here.

For larger $F$, the valleys reach lower values of $m_*/M$ at any given $\chi$. For $\chi \approx 10^{-6}$, the lowest valley haloes with $F = 8$ and $F = 9$ give $m_*/M \sim 10^{-3.3}$ and $\sim 10^{-4.5}$, respectively. This range may be consistent with observed SMBH–bulge relations if the DM halo is $\sim 10^1$ times the baryonic mass. Thus, on the one hand, when $7 \lesssim F < 10$ some of the $\psi$-valleys are consistent with realistic SMBH masses relative to the host galaxy. Conversely, assuming these theories of $F$, we predict that some galaxies host SMBH in low-$\psi$ configurations: the dark envelope is dense near the horizon. At least in the present Newtonian model, the edge of such a SMBH is blurry. This deserves further investigation through general relativistic calculations.
3.3 Relation of central mass to halo core

Formally, the configuration space at fixed $F$ is two-dimensional, with gravitational compactness and phase-space density parameters $(\chi, Q)$ or $(\chi, q)$. In practical applications, the compactness parameter $(\chi = 2GM/Rc^2)$ may be difficult to estimate, since it depends on the total mass $M$ contained within the dark halo truncation radius $R$, which is not directly observable. It may be more useful to specify models in terms of quantities pertaining to the measurable DM core. If the local index of the density profile is

$$\alpha \equiv -\frac{d \ln \rho}{d \ln r}, \quad (16)$$

then we can annotate slope-radii ($R_\alpha < R$) at a standard chosen $\alpha$. The radii $R_\alpha$ can be multivalued (in terraced haloes) and this is more likely when $F$ is larger. We shall define the DM core to be the region enclosed by the outermost locations where $\alpha = 1, 2, 3$, i.e. $R_1 < R_2 < R_3 < R$. The mass contained within these radii satisfies $m_* < M_1 < M_2 < M_3 < M$. The gravitational compactness of the core can be written as $\chi_\alpha = -2\Phi_\alpha/c^2$. Similarly, we might define the core at the half-mass radius $R_m$ and potential $\Phi_m$ (where $m(R_m) = \frac{1}{2}M$), with core compactness $\chi_m = -2\Phi_m/c^2$. In our discussions below, we can abbreviate the core compactness $\chi_\alpha$ standing for $\chi_1, \chi_2, \chi_3$ or $\chi_m$. The particular choice does not change the qualitative conclusions. These notations disregard the tenuous outskirts of the halo, which in any case are difficult to measure astronomically.

When the $(\chi, q)$ plane transforms to $(\chi_c, q)$, the plateau region straightens from a wedge with diagonal stripes to a rectangular region with vertical stripes (see Fig. 5). Except near the upper $q$ border (the non-singular limit), the models of fixed $\chi_c$ are almost independent of $q$. The two-dimensional parameter-space is almost (but not quite) reduced to a one-dimensional space in terms of the core compactness.

Fig. 6 shows the variation of $m_* / M_c$ with respect to core compactness, for equations of state with $F = 6.5, 7.0, 7.5, 8.0, 8.5, 9.0$. Each ribbon depicts the entire plateau region, apart from the border strip. The physically uninteresting 'hole-dominated' region hides in the top-right point where $m_* \approx M$. The thinness of the ribbons in this projection shows how $q$ becomes inconsequential compared to $\chi_c$. The non-singular solutions (not shown) have smaller $m_* / M_c$. 

Figure 5. Mass fraction of the central object (shaded) in relation to the DM core, for various values of $F$ (as annotated). Use of core compactness (e.g. $\chi_c = \chi_2, \chi_3$ or $\chi_m$) instead of global compactness $\chi$ reveals a projection in which the model properties are insensitive to $q$ except near the upper border (non-singular solutions).
for given $\chi_c$: down to arbitrarily small values as $q$ approaches its maximum. In Fig. 6, they occupy the region of the $(\chi_c, m^*/M_c)$ plot below the ribbon of given $F$. Thus, each ribbon represents the maximum possible $m^*$ hosted within a DM halo core of given compactness.

Now, we can interpret a characteristic velocity dispersion of tracer objects in the core region, $\sigma_g \propto c \sqrt{\chi_c}$. Observable velocity dispersions of the GC swarm should match this value to within a factor of a few. For an assumed halo $F$ and known $\sigma_g$, one can estimate $\chi_c$ and infer a narrow range of possible $m^*/M$ (if the galaxy is in the ‘plateau’ regime) or an upper limit on $m^*/M$ (if it is a nearly non-singular case). An estimate $\sigma \propto \sigma_g$ can be substituted in equations (14) and (A3). Estimates should be most robust for galaxies where GCS projected $\sigma_g$ appears nearly constant within the DM core (e.g. Côté et al. 2003; Bridges et al. 2006; Norris et al. 2012; Napolitano et al. 2014).

The shading of the ribbons in Fig. 6 shows the horizon correction factor $\log_{10} \sqrt{\eta^3 \Psi}$ that applies in equation (14) when predicting SMBH masses in physical units. For the $F > 6$ plateau halo models, the variation of this term is small. For galaxy compactness ($\chi_c < 10^{-4}$), the correction factors vary by less than 0.6 dex. The variation is smaller for lower $F$. Thus, equation (14) is robust enough to apply approximately, even when $Q$ is unobservable.

4 DISCUSSION

4.1 Preferred solutions and $m^*$

Although the $\chi_c$ core representation simplifies the projected configuration space, the present spherical halo model still has two free parameters: the compactness $\chi$, and some measure of the orderliness (such as $Q$ or $q$). Given the apparent simplicity of the empirical relations between $m^*$ and host galaxy properties, it is worth seeking a simple causal explanation. Is there any physical principle that constrains $q$ as a function of $\chi_c$ or $\chi$? A satisfactory model would involve a simple intuitive rule involving instantaneous halo properties, without complexities involving fine-tuning processes, such as those invoked in feedback scenarios, local contingencies or accidents of evolutionary history. Here, we shall discuss some conceivable rule-of-thumb explanations.

4.1.1 Cosmic density

In the theory and simulations of cosmological collapse of collisionless haloes, galaxy-like objects lack a clearly defined outer boundary surface. Instead, they are described in terms of a virial radius containing some multiple of the cosmic critical
density (e.g. $\rho_v = 100 \rho_{crit}$ with $\rho_{crit} \approx 9.2 \times 10^{-30}$ g cm$^{-3}$, when $H_0 \approx 70$ km s$^{-1}$ Mpc$^{-1}$; e.g. Hinshaw et al. 2013).

It is not obvious whether haloes in polytropic theories would also separate from the cosmic background at a standard density. If they do, then their individual radii and compactness are linked,

$$R = \sqrt{\frac{3c^2}{\hbar \pi G \rho_v}}$$

(17)

and the corresponding mass is $M = x^2 R / 2G \propto \sqrt{\chi / \rho_v}$. Thus, the dimensionless parameter $\chi$ is linked to empirical properties of each galaxy. Halo star and GC velocity dispersions would scale as $\sigma \sim \sqrt{\chi}$. Local densities would scale in proportion to the standard $\rho_v$. The internal mass distribution still depends non-trivially upon $q$ or $Q$ however, which determines whether a particular galaxy halo is non-singular, highly singular, or any condition in between. If $\eta = \eta(\chi, q)$ were weakly dependent on $q$, then equation (14) would imply a power-law correlation, $m_* \sim \sigma^{1/2}$. If $\eta = \eta(\chi, q)$ also has non-trivial variations, the assumption of a universal virial density $\rho_v$ would not predict $m_*$ tighter than the ribbon relations in Fig. 6. At least one extra principle is needed.

4.1.2 Halo entropy

The total entropy of the DM in the halo is

$$S = -Nk \ln(Q / Q_0),$$

(18)

where the constant $Q_0$ depends on universal particle properties, $N = (M - m_\gamma) / \mu$ is the number of DM particles and $\mu$ is the particle mass. The event horizon also contributes entropy, $S_r \approx \tau(\tau_r / l_0)^2$, where $l_0$ is the Planck length (Bekenstein 1973). When models are normalized (Appendix B) to the same total mass $M$, the total entropy is

$$S = 4\pi \left(\frac{M_{\text{Pl}}}{m_p}\right)^2 \left(\frac{m_*}{M}\right)^2 \eta^2 \left(1 - \frac{m_*}{M}\right) k \ln \left(\frac{Q}{Q_0}\right),$$

(19)

where $m_p$ is the Planck mass. The left-hand (horizon) term of equation (19) dominates if $\mu \gg m_\gamma^2 / M$ and vanishes if $\mu \ll m_\gamma^2 / M$. At fixed $\chi$, the right-hand term is monotonic in $Q$, and so is the left-hand term, except subtle wrinkles within 1 dex of the non-singular border. Maximal entropy prefers a massively massive BH with only a tenuous dark envelope. Realistic SMBH scaling relations cannot derive from a simple entropic principle.

4.1.3 Energetic constraints

For the $F > 6$ scenarios, some mass profiles are energetically more or less favourable, depending non-trivially on the system parameters. For fixed $M$ and $\chi$, the gravitational potential energy $|W|$ and total energy are extremal at the $\psi$-valley where $q$ is lowest. This is the valley where $m_*/M \gtrsim 0.1$, which is excessive. The BH mass is significant compared to the envelope, but not dominant. This solution is energetically favoured because DM is mostly concentrated deep in the potential well. At lower $q$, the BH dominated profiles ($m_*/M \approx 1$) are less energetically favourable because there is not much matter in the tenuous halo. In the medium-$q$ plateau domain, the configurations are less energetically favourable because the mass is less concentrated.

However, the other $\psi$-valleys (where the $m_*/M$ ratios are more astronomically realistic) are suboptimal extrema of $|W|$. In energetic terms, these states may be locally preferred to adjacent configurations in $(\chi, q)$ space.

It is not obvious whether or not these energetically favourable states are effective attractors in galaxy halo evolution. An evaluation of realistic evolutionary tracks in $(\chi, q)$-space might require Monte Carlo simulations that apply hierarchical mergers to an initial population of primordial mini-haloes. As in toy-model studies of SMBH demographics (e.g. Yu & Tremaine 2002), it would be necessary to assume whether adiabatic agglomeration or mass–energy conservation takes priority during mergers, flybys and fission events. Dark shocks and mixing would introduce inelastic and dissipative factors. These issues are non-trivial and deserve a separate investigation.

4.1.4 Landscape of $\psi$

The density ratio of the central object to its envelope, $\psi$, is a diagnostic of the halo solutions. One might wonder whether a sensible physical condition involving $\psi$ might select the astronomically realistic models. Large values of $\psi$ imply a central object with a high density contrast to its surroundings. The rare cases with $\psi < 1$ are perhaps unnatural as they would imply an overdense inner envelope, and a density inversion in whatever primordial object formed the SMBH seed in the first place. Small values of $\psi \lesssim 1$ are of special interest, as they imply systems where the inner DM envelope is comparable to the mean density of the central object. In some sense, this implies a SMBH that is maximally blended and coupled with the galaxy halo.

An inner condition $\partial m_*/ \partial r = dm_*/dr$ would describe a seamless continuity between the SMBH and its envelope. If the horizon occurred at the Schwarzschild radius ($r_s \approx R$ and $\eta \approx 1$), the optimal continuity condition would imply $\psi \approx 3$. When the envelope is massive enough that $\eta \neq 1$ by a significant amount, the seamless condition will prefer another value of $\psi$ (depending on the actual non-Schwarzschild $\partial m_*/ \partial r$ at fixed $\chi$). Ideally that rule should be calculated from the numerical maps of $\partial m_*/ \partial q$ and $\partial r_s / \partial q$. We would expect the special value of $\psi$ to be a number of order three or unity: probably near the minima in the $\psi$-valleys, and not in the large-$\psi$ regions outside the plateau.

The seamless envelope condition is theoretically interesting, but what would it imply about SMBH formation and growth? It might be the natural outcome if DM accretion was the main mass supply to the BH, either through gradual, adiabatic contraction or violent, supernova-like implosions. If the subtle energetic preference for $\psi$-valleys were an evolutionary attractor, this qualitative picture might become something more quantitatively predictive.

4.1.5 Summary

If haloes share a cosmologically determined mean density, then their individual masses are functions of $\chi$, but the variation of internal structure means this ansatz does not provide unique predictions for $m_*$. Entropy maximization ideally favours high $m_*/M$, which could not describe a realistic galaxy.

Gravitational energy also favours models with $m_*$ too large, but there is a subtler preference for moderate-$m_*$ profiles in the $\psi$-valleys of the $(\chi, q)$ space.

If galaxies tend to evolve to minimize $\psi$, then this implies that a relativistic dark envelope surrounds the SMBH horizon, where local densities of DM could be significant compared to the SMBH mean density. In that case, the preferred configurations include: tracks where $m_*/M$ is almost independent of $\chi$ (resembling Magorrian relations); a track of minimum $q$ with unrealistic $m_*/M$; and low
$m_\star/M$ cases near the maximum $q$ (non-singular border, minimal entropy). Detailed scaling relations depend on $F$ sensitively.

When the halo is described in terms of properties of its DM core, the plateau region of the parameter space simplifies considerably. In these terms, the $m_\star/M$ ratio falls within a narrow ribbon that depends on $F$ and $\chi$, but only weakly on $q$. The valley and inter-valley solutions shrink together into the same projected region. The ribbons are thinner than the present scatter in observational $m_\star$ values, so it is practically almost a one-dimensional $m_\star-\sigma$ scaling relation. When nearly non-singular models are allowed, the ribbons in Fig. 6 become strict upper limits on $m_\star/M_\star$. If we can link the velocity dispersion of halo tracers – such as GCs – to the core radius $r_c$ (or $\chi$, see above), then realistic $m_\star$ versus $\sigma$ relations emerge. These relations may have some intrinsic scatter, due to the model dependences of $\sqrt{\eta^*\psi}$ envelope correction factor. This factor reintroduces some $Q$-dependence, though it is subtle: for $6.5 \leq F < 9.5$ and $\chi > 10^{-8}$, we find $0.5 < \frac{1}{2} \log_{10}(\eta^*\psi) < 1.81$ across the plateau region (terraced haloes, as coloured in Fig. 6). For galaxy-like compactness $\chi \ll 10^5$, $0.5 \lesssim \frac{1}{2} \log_{10}(\eta^*\psi) < 1.1$

4.2 Comparison to observed SMBH

Empirically, the heaviest known ultramassive BHs amount to a few times $10^{10} m_\odot$ (McConnell et al. 2011, 2012; van den Bosch et al. 2012). The smallest confirmed SMBH are a few times $10^7 m_\odot$, residing in bulgeless disks and dwarf galaxies (Filippenko & Ho 2003; Barth et al. 2004; Peterson et al. 2005; Shields et al. 2008; Seth et al. 2010; Reines, Greene & Geha 2013). For the central BH of a massive elliptical galaxy, $m_\star = 10^9 m_\odot$ and the Schwarzschild radius $r_s = 2.95 \times 10^{14} \text{cm} \approx 10^{-3} \text{kpc}$. For a realistic galaxy-sized halo, this implies $r_s/R \lesssim 3^{-3}$, and $m_\star/M \lesssim 10^{-3}$ or $\lesssim 10^{-4}$ assuming a large galaxy with $\chi \sim 10^{-6}$. We adopt these values as a benchmark.

For haloes with $0 \leq F \leq 6$, astronomically realistic ratios of $m_\star/M$ only exist very close to the border of non-singular models. A small relative amount of heating (e.g. dissipative effects of a tidal flyby) could induce a significant jump in $m_\star$, unless the compactness $\chi$ also changes. In the thin band of ($\chi$, $q$) states where $F \leq 6$ is compatible with SMBH scaling trends, the density contrast between the hole and its dark envelope tends to be immense ($\psi \gg 10^{20}$). This means that when $F \leq 6$, an astronomically realistic BH is much denser than its surroundings, and effectively decoupled from the ambient halo pressure. It would be necessary to invoke elaborate, mundane non-DM physics to explain the observed correlations. The $6 < F < 10$ regime however enables observationally plausible $m_\star/M$ values throughout the small-$\psi$ plateau and valleys, as well as near the non-singular border. Many orders of magnitude are available in $q$ and pseudo-entropy $s$. Incrementally heating a galaxy halo need not be catastrophic for SMBH growth.

Observational constraints are met in the $\psi$-plateau when $6 < F < 10$ (for $r_s/R$) and $7.5 \lesssim F < 10$ (for $m_\star/M$, according to our Fig. 6). This then is an observationally favoured range of DM microphysics. We typically find $\psi < 100$ in the best models. This entails a dark envelope with gravitationally significant density near the event horizon. This envelope declines radially as a nuclear ‘spike’, though more steeply than the low-$F$ spikes of previous modelling (Gondolo & Silk 1999; Mouwad et al. 2005; Hall & Gondolo 2006; Zakharov et al. 2010). For large $F$, the spike’s steepness makes the combined DM envelope plus BH appear (from afar) as if it were a more-massive BH.

To an accuracy comparable to the present observational scatter, it is useful to represent the mass trend as a power law, $m_\star/M_\star \sim M_\star^{\beta-1}$. If we assume that galaxy haloes share a nearly universal cosmic mean density (Section 4.1.1, $M_\star \propto \sqrt{\Omega_c}$), then the expectation is $m_\star/M_\star \sim \chi_{\star}^{0.6-0.8}$. Our numerical results, the domain $6 < F < 10$ ensures $1 < \beta < 2$; while $F < 6$ only gives solutions near the non-singular limit ($\beta = 1$). Observations of $m_\star$ in local galaxies and AGN (Laor 2001) show $\beta = 1.54 \pm 0.15$. In our Fig. 6, this would correspond to slopes of $\delta \ln(m_\star/M_\star)/\delta \ln \chi_{\star} = 0.81 \pm 0.23$, which graphically is consistent with the ribbons of higher $F$ cases. Bandara, Crampton & Simard (2009) modelled the strong gravitational lensing effects of a set of elliptical galaxies that also have $m_\star$ estimates. They found a correlation that implies $\beta = 1.55 \pm 0.31$ or $\beta = 1.57 \pm 0.39$ depending on their fitting methods. Our equivalent ribbon slopes would be $\delta \ln(m_\star/M_\star)/\delta \ln \chi_{\star} = 0.82 \pm 0.47$ or $0.86 \pm 0.59$. These constraints are lax, but would seem to prefer $F \geq 7$.

In our model, the predicted ratios $m_\star/M$ refer to $M$ as the halo mass, not the stellar bulge ($M_\star$). Peculiar galaxies observed with high $m_\star/M_\star$ (Bogdán et al. 2012; van den Bosch et al. 2012) could be normal products of the SMBH–halo relationship, but impoverished in stars and gas for some other reason. Alternatively, if they are genuinely overweight in $m_\star/M$ terms, they might be high-entropy outliers: low $q$ or high $\chi$, due to an unlucky history of tidal buffeting or other halo heating processes.

Our model also has implications for the presence of intermediate-mass black holes (IMBH; $10^3 \lesssim m_\star/m_\odot \lesssim 10^6$) in the least-massive systems. Based on velocity dispersions, escape velocities and tidal radii, ultracompact dwarf and faint dwarf galaxies could have compactness parameters $\chi \lesssim 10^{-2}$. If they bind substantial amounts of DM, then the $\psi$-plateau of $F > 6$ haloes set upper limits on $m_\star/M_\star$ that are rather low (left extreme of Fig. 6). In a system amassing $M = 10^7 m_\odot$, the ‘plateau’ configurations of $F = 7, 8, 9$ haloes with $\chi \approx 10^{-3}$ predict a maximum central object of $m_\star \approx 10^5, 10^6$ and $10^7 m_\odot$, respectively. For objects with $\chi \approx 10^{-7}$, these $F = 7, 8, 9$ models give $m_\star \approx 10^6, 10^7$ and $10^8 m_\odot$, respectively. This object could be a stellar BH, rather than an IMBH. If there is a non-trivial central stellar density, then the predicted central mass is lost amidst stellar granularity, and the model breaks down. Even if an IMBH were formed, there are plausible processes that might remove it: the ‘gravitational rocket’ effect during high-spin BH mergers; random walks due to scattering in dense stellar environments; random walks due to momentary imbalances between the thrusts of two jets during a gas accretion episode. The rarity or non-observation of IMBH in dwarf galaxies and GC is unsurprising.

4.3 Possible observational tests of the DM envelope

The presence of a DM envelope around a BH contributes to the gravitational potential, which will produce observational consequences. Here, we list a few examples.

(i) The gravitational potential of the DM envelope will cause stellar orbits to deviate from the Keplerian orbits that are expected for motion around a bare spherically symmetric gravitating object (Rubilar & Eckart 2001; Mouwad et al. 2005; Hall & Gondolo 2006; Zakharov et al. 2007; Ghez et al. 2008; Will 2008; Zakharov et al. 2010; Iorio 2011). A possible means to detect this deviation is timing observations of pulsars, if present, around the central BHs in nearby galaxies (Wex & Kopeikin 1999; Kramer et al. 2004; Pfahl & Loeb 2004; Liu et al. 2012; Singh, Wu & Sarty 2014).

(ii) Stars with non-circular orbits traversing the DM envelope around a BH would experience a gentler tidal-force gradient than
around a bare BH of equal total mass. In a stellar tidal disruption process (see Rees 1988; Komossa 2002; Bloom et al. 2011; Saxton et al. 2012), the stellar debris tracks would have morphologies different to those resulting from a rapid change in the tidal force field.

(iii) AGN are powered by accretion of gas into a massive black hole (MBH). The inner accretion disc region unleashes most of the accretion power, in the form of radiation and outflows. For objects orbiting around a BH, there is a well-defined innermost stable circular orbit (ISCO). This orbit is assumed to be the inner boundary of the accretion disc, because beyond that, the inflow matter plunges towards the horizon without having time to dissipate and radiate energy. X-ray emission line profiles are often used as a diagnostic of space–time properties and conditions near the inner-disc radius (see e.g. Fabian et al. 1989, 2000; Stella 1990; Laor 1991; Fuerst & Wu 2004; Younsi, Wu & Fuerst 2012). However, the ISCO location does not have a simple analytic solution when a massive DM envelope is present. The gravity of the DM envelope modifies the accretion flow dynamics, and hence the thermodynamics and radiative properties of the inner disc. Accretion discs around BHs in the presence and in the absence of a massive DM envelope would show different spectral profiles (cf. Joshi, Malafarina & Narayan 2011, 2014; Bambi & Malafarina 2013). Thus, BH parameter estimates derived without accounting for DM could give incorrect results.

(iv) Interferometric imaging of SMBH in nearby galaxies will be possible with the development of the Event Horizon Telescope (EHT)2 and the Greenland Telescope (GLT).3 An SMBH that is heavily enveloped by DM might show an ‘event horizon’ shadow smaller than that expected from stellar-kinematic mass deductions (cf. Falcke, Melia & Agol 2000; Nusser & Broadhurst 2004; Doeleman et al. 2008).

4.4 DM physics and microphysics

There are many theories of DM physics that can viably describe the gravitational fields of galaxy haloes. At galactic scales, the only essential requirements are that the unknown material is electromagnetically invisible and has no discernable effect on nucleosynthesis or the stability of normal stars. Since halo shapes are spheroidal, the DM seems unable to lose energy as readily as the radiatively cooling gas in classic astrophysical discs (although see Fan et al. 2013).

Often DM is simply assumed to be collisionless: practically an invisible self-gravitating dust. This provides an easy prescription for cosmological simulations employing N-body methods. However, observations do not confirm the predicted density cusps (see Section 1). Simulations also overpredict numbers of dwarf galaxies (Klypin et al. 1999, 2014; Moore et al. 1999; D’Onghia & Lake 2004; Tikhonov & Klypin 2009; Zwaan, Meyer & Staveley-Smith 2010), and dense large satellites that are seen in reality (Boylan-Kolchin, Bullock & Kaplinghat 2011, 2012; Garrison-Kimmel et al. 2014; Kirby et al. 2014; Miller et al. 2014; Tollerud, Boylan-Kolchin & Bullock 2014). The question then is: what variety of modification or alternative theory is necessary? Suppose that some process drives the DM phase-space distribution function to become locally isotropic and proportional to a power of the single-particle energy, $f \propto (-E)^{\alpha}$. Then, a polytropic relation (1) emerges (Camm 1952). In collisionless DM simulations, the cuspy haloes have $Q$ following a power of $r$ when assuming $F = 3$ (Taylor & Navarro 2001; Ludlow et al. 2011), which implies a constant-$Q$ singular polytrope for some non-integer $F$ value. For those models, the severest challenge is to explain why kpc cores occur in real haloes. Driving processes might involve shaking by an elaborate baryonic feedback (e.g. Peirani, Kay & Silk 2008), or collective phenomena similar to bar-mode instabilities. This needs fine-tuning to achieve a realistic core radius. When simulations invoke ad hoc feedback recipes, these can be decisive or ineffectual, depending on numerical implementation (e.g. Governato et al. 2010; Vogelsberger et al. 2014).

The polytropic condition is claimed to be a natural equilibrium for self-gravitating systems, according to the Tsallis (1988) conjecture of extended thermostatistics. Collisionless spheres may settle as ‘stellar polytropes’ (Plastino & Plastino 1993; Vignat, Plastino & Plastino 2011). Our parameter $F$ is linearly related to Tsallis’ extensivity parameter, which is a non-integer. F´eron & Hjorth (2008) find that stellar polytropes are a poor representation of cuspy haloes that emerge in numerical simulations. However, this is not a fatal criticism of the thermostatistical models, since real observed galaxies have cored (not cuspy) profiles.

Another possibility is that DM is adiabatic and self-interacting (SIDM). The polytropic equation of state (1) arises from basic thermodynamics, in the absence of complications such as phase changes. SIDM interactions may consist of direct interparticle scattering, short-ranged Yukawa interactions or long-ranged dark forces analogous to magnetism (e.g. Spergel & Steinhardt 2000; Ahn & Shapiro 2005; Ackerman et al. 2009; Buckley & Fox 2010; Loeb & Weiner 2011). If the fluid consists of point-like particles with only translational motions, then $F = 3$. This is a common, unquestioned assumption, algorithmically built into many simulations of weakly interacting SIDM (Moore et al. 2000; Yoshida et al. 2000; Davé et al. 2001; Vogelsberger, Zavala & Loeb 2012; Peter et al. 2013; Rocha et al. 2013; Vogelsberger et al. 2014). However, if DM has additional internal energy then $F > 3$, e.g. $\frac{1}{3}kT$ for each degree of freedom of rotational kinetic energy of ‘dark molecules’. For diatomic dark molecules, $F = 5$. The $F$ value increases if DM particles have more composite complexity. Independently, some efforts to reconcile direct detection experiments invoke composite or inelastic DM (e.g. Smith et al. 2001; Chang et al. 2009; Alves et al. 2010; Kaplan et al. 2010, 2011). The astroparticle physics implications are increasingly recognized (e.g. Cline et al. 2014a,b; Boddy et al. 2014). This possibility is inherently beyond the scope of N-body codes. The quantity $F$ might effectively vary in some DM theories that yield pressure anisotropies and/or a more complicated equation of state (e.g. Sobouti, Hasani Zonoozi & Haghi 2009; Harko & Lobo 2011, 2012). For now, we assume constant $F$.

The nature of SIDM is still under debate. It is possible that SIDM is not a gas but a scalar field or boson condensate (e.g. Ji & Sin 1994; Sin 1994; Lee & Koh 1996; Hu, Barkana & Gruzinov 2000). A polytropic equation of state can be obtained from some boson models, with the value of $F$ depending on the self-coupling potential in the lagrangian. Many works assume $F = 2$ with $s$ and $Q$ fixed universally by particle properties (e.g. Goodman 2000; Arbey, Lesgourgues & Salati 2003; Böhmer & Harko 2007; Chavanis & Delfini 2011; Harko 2011), but other $F$ values are possible (Peebles 2000). It was also suggested that phase changes can occur in bosonic DM and this would alter the spatial variations in the properties of large astrophysical objects (see e.g. Arbey 2006; Slepian & Goodman 2012).

Alternatively, DM may consist of neutral fermions (Dodelson & Widrow 1994). Warm DM made of sterile neutrinos ($\sim 1–7 \text{ keV}$

---

2 http://www.eventhorizontelescope.org/
3 http://cfa.harvard.edu/greenland12m/
mass range, depending on the primordial particle distribution) might decay, producing X-ray emission lines (Boyer et al. 2014; Bulbul et al. 2014). In this case, a degenerate phase can act as a cored polytrope with \( F = 3 \) (e.g. Manuanean & Biermann 2005, 2006; Richter, Tupper & Voilier 2006; Chan & Chu 2008; Destri, de Vega & Sanchez 2013). For fermionic DM, the Pauli exclusion principle implies a universal maximum phase-space density, \( Q \leq Q_{\max} \). In a DM halo, \( q \leq Q_{\max} c^{3} (8\pi G R^{2}/3c^{2})^{2}(F-2)/2 \). Assuming that the haloes have roughly the same mean density gives \( q \leq Q_{\max} c^{3} (\rho_{c}/\rho_{s})^{2}(F-2)/2 \). Either way, a region of \( (\rho, q) \) map is excluded above a diagonal line. For sufficiently low \( \rho \), non-singular solutions are excluded. This implies that isolated haloes cannot form below some critical mass (if non-singular) or otherwise they must have a singular, terraced or BH-dominated centre. Assuming that dwarf galaxies obey this limit, observationally inferred \( Q \) values reveal or exclude the candidate particle properties (e.g. Tremaine & Gunn 1979; Boyarsky, Ruchayskiy & Iakubovskyi 2009; Destri et al. 2013; de Vega et al. 2014; Domcke & Urbano 2014; Horiuchi et al. 2014). Those studies implicitly assume point-like particles (\( F = 3 \)). The wider possibilities of \( F \neq 3 \) fermions remain unchecked.

Whatever the fundamental nature of DM, its distribution must deform within the gravitational sphere of influence of an SMBH. A dark density spike emerges either as a static equilibrium, or as the result of gradual capture of DM at the horizon (Ipsen & Sikivie 1987; Quinlan et al. 1995; Gondolo & Silk 1999; Ullio et al. 2001; MacMillan & Henriksen 2002; Peirani & de Freitas Pacheco 2008). A stellar density cusp may also develop in this region (Bahcall & Wolf 1976; Young 1980). In dense galaxy nuclei, the gravitational scattering of DM by the stars renders the halo indirectly collisional (Gnedin & Primack 2004; Ilyin, Zybin & Gurevich 2004; Merritt 2004, 2010; Zelnikov & Vasiliev 2005; Vasiliev & Zelnikov 2008). Unless DM annihilation or other effects overrule the dynamics, a polytropic description applies in the stellar cusp. For standard DM with point-like particles (\( F = 3 \)), the dark spike profile is \( \rho \sim r^{-3/2} \). If there is internal energy, then \( F > 3 \), and the spike is steeper, \( \rho \sim r^{-3/2} \). If \( F > 6 \), then the mean-free-path of self-scattering \( (\chi \propto \sigma_{\gamma} / \rho \sim r^{1-(F-6)/2}) \) shortens and vanishes at small radii in the spike, which justifies an adiabatic SIDM treatment regardless of DM collisionality properties in the outer halo. Our formulae (14) and (A3) relating the BH mass (\( m_{*} \)) to the DM properties therefore should hold locally in galactic nuclei.

The observational evidence that galaxies possess kpc-sized dark cores is well modelled by polytropic density profiles. Relative to the global \( M \) and \( R \), core sizes tend to shrink as \( F \) increases. Scaling relations among disc galaxies imply high \( F \) (Nunez et al. 2006; Zavala et al. 2006). From the kinematics of elliptical galaxies, the inference is \( 7 \lesssim F \lesssim 9 \) (Saxton & Ferreras 2010). Models of galaxy clusters comprising DM and cooling gas inflows predict realistic core sizes and enable realistic \( m_{*} \) if \( 7 \lesssim F < 10 \) (Saxton & Wu 2008, 2014). Now our simple analysis of BH plus adiabatic haloes also supports this range of \( F \). The case of \( F = 9 \) naturally leads to a rule \( m_{*} \sim \sigma_{*}^{4.5} \). If velocity dispersions of DM and stars both follow the shared gravitational potential, then this matches the observed correlation, \( m_{*} \sim \sigma_{*}^{4.5} \).

### 4.5 GCs as a tracer

GCs inhabit the host galaxy’s halo and provide a useful physical probes where other visible tracers are rare. The GC swarm diminishes with distance from the core, but can also develop central deficits (Capuzzo-Dolcetta & Mastrobuono-Battisti 2009). GC consist of uniformly old and metal-poor stellar populations. They appear to lack DM of their own: stellar mass suffices to explain the internal kinematics (e.g. Heggie & Hut 1996; Baumgardt et al. 2009; Sollima et al. 2009; Lane et al. 2010; Bradford et al. 2011; Conroy, Loeb & Spergel 2011; Hankey & Cole 2011; Sollima, Bellazzini & Lee 2012; Ibata et al. 2013).

GC formation was either a purely baryonic process, or else their miniature DM haloes were ablated later. The oldest GC apparently formed in brief single starbursts comparable to a dynamical time of the proto-galaxy, perhaps caused by thermal instabilities or shock compressions of clouds in the halo (e.g. Searle & Zinn 1978; Fall & Rees 1985). Newer (metal-rich) GC may form from shocked gas in wet mergers (Aumann & Zepf 1992; Zepf & Ashman 1993; Whitmore & Schweizer 1995; Hancock et al. 2009; Whitmore et al. 2010; Smith et al. 2014). Dry mergers of galaxies combine pre-existing GC swarms and preserve the ratios of SMBH, stellar and GC masses. GC on radial orbits traversing the inner galaxy can be destroyed by tidal shocking (e.g. Ostriker, Spitzer & Chevalier 1972; Fall & Rees 1977; Gnedin & Ostriker 1997; Gnedin, Lee & Ostriker 1999; Fall & Zhang 2001). Compared to ellipticals, disc galaxies seem more efficient as GC destroyers or less-efficient GC formers. (e.g. Harris 1988; Georgiev et al. 2010). The surviving GC population depends on: the primordial baryonic mass endowment; the subsequent formation and destruction processes; and the breadth and depth of the halo potential binding GC to the galaxy. By the virial theorem or Jeans modelling, the radial velocity dispersion of the GC system is proportional to the depth of the halo potential.

Since the GC swarm traces aspects and properties of the whole galaxy halo, it is significant that GC observables correlate with the SMBH (\( m_{*} \)) (Spitler & Forbes 2009; Burkert & Tremaine 2010; Harris & Harris 2011; Harris et al. 2013). Snyder et al. (2011) interpret the SMBH–GC correlations as consequences of the depth of the galaxy bulge’s gravitational potential. Sadoun & Colin (2012) relate the velocity dispersion of the GC system, \( \sigma_{*} \), to the depth of the halo potential. There are robust relations of the form \( m_{*} \sim \sigma_{*}^{3} \) with \( \beta = 3.78 \pm 0.33 \). Pota et al. (2013) also linked \( m_{*} \) with \( \sigma_{gc} \) (\( 3 \lesssim \beta \lesssim 6 \) or \( \beta \approx 4.45 \) on average) and Rh ode (2012) found \( \beta \approx 5.3 \) or 5.9. These \( \sigma_{gc} \) relations have great implications. This correlation could be evidence of a link between SMBH formation and the halo properties, not merely the properties of the stellar bulge. The stronger the \( m_{*} \sim \sigma_{gc} \) relation is, the less likely that these components are controlled by BH feedback, and the more likely that it depends somehow on the underlying DM potential.

Burkert & Tremaine (2010) and Rh ode (2012) have a different interpretation: attributing the correlation to the effect of mergers later on (more mergers produce more GCs and a bigger SMBH). We suggest that the correlation would not be so tight if the individual merging blocks did not already have a correlation on their own. Furthermore, mergers cannot have been the controlling process in bulgeless thin-disc galaxies that host an SMBH but have never experienced a major merger (Section 4.7). Mergers cannot be the universal explanation. Instead, we propose that the halo controls the SMBH origin and the GC properties separately. In each large galaxy, there will be a fraction of large GCs produced in situ during the initial collapse, and a fraction coming later from the disruption of nucleated satellite galaxies. Stellar populations and orbital kinematics are usually clues to which is which (e.g. M54 and \( \omega \) Cen may be satellite accretions). It would be interesting to predict the implications if local GCs are those formed without a DM potential well, and those coming from accreted galaxies are formed at the bottom of that galaxy’s DM potential, perhaps even with their own nuclear BHs.
4.6 SMBH formation and accretion

Our equilibrium configurations do not distinguish how the central object originated. We simply have a non-evolutionary description of the endpoint after the inner halo attains approximate pressure balance. Our model $m_\star$ limits do not apply while a system is dynamically disturbed, asymmetric and evolving into another state. However, the most realistic equilibrium solutions tend to have small $\psi$ values, meaning that a dark envelope is a significant presence around the horizon. This suggests that DM accretion may be relevant to SMBH seeding and growth. We are aware of at least three scenarios. Steady growth is perhaps via Bondi (1952) accretion of fluid (e.g. Manyanze & Biermann 2005, 2006; Richter et al. 2006; Peirani & de Freitas Pacheco 2008; Guzmán & Lora-Clavijo 2011a, b; Pepe et al. 2012; Lora-Clavijo et al. 2014) or gradual capture of collisionless orbiting particles accompanied by loss-cone refilling. (e.g. Peebles 1972; Ullio et al. 2001; Vasiliev & Zel’nikov 2008). If the DM self-interactions are weak (with a kpc-sized mean-free-path) but heat conduction is significant, then gravitational instability could form an SMBH (Ostriker 2000; Balberg & Shapiro 2002; Balberg, Shapiro & Inagaki 2002; Hennawi & Ostriker 2002). If SIDM is a fluid with $F > 6$, then collapse may proceed via a localized gravitational instability in a discrete ‘dark gulp’ lasting a dynamical time-scale of the nucleus (Saxton & Wu 2008, 2014). The gulfed dark mass could be an appreciable fraction of the SMBH total.

Initiating this process may require a steep central density gradient. BH seeding is probably helped if there is already a steep spike of stars or accumulation of inflowing gas. It may be necessary for baryons to become denser than some threshold, in order to pinch the DM (via adiabatic contraction, Blumenthal et al. 1986) and enable collapse of the innermost DM. Perhaps, this pinching can partly explain the observed correlations between SMBH and the Sérsic index of the stellar surface brightness profile (Graham et al. 2001; Graham & Driver 2007; Savorgnan et al. 2013). Evaluating the collapse thresholds needs multicomponent stability analyses, like Saxton (2013) but with a density spike.

Some comparisons of the mass function of the local SMBH population with the AGN and quasar luminosity distribution were consistent with most of the current SMBH mass coming from radiatively efficient gas accretion (Soltan 1982; Salucci et al. 1999; Yu & Tremaine 2001; Shankar et al. 2004; Shankar, Weinberg & Miralda-Escudé 2009). This does not invalidate our proposed scenario. If these audits of light and mass are complete, they are still consistent with an initial relation between bulge mass and seed BH mass, in which the latter could have been $\lesssim 10^{-4}$ bulge mass and much less than the final SMBH mass. That situation corresponds to $X \lesssim 10^{-6}$ in the $F \gtrsim 8.5$ halo models. Spatially, $R \sim 10^{12} r_s$ would be a plausible size for a seed BH, as $r_s \lesssim 10^{11}$ cm for large galaxies with $R \sim 10 s$ of kpc ($\sim 10^{15}$ cm). These seeds could have condensed according to our predicted scaling index, $m_\star \sim \sigma^{-5/2}$, and then grown through Eddington (1918) limited luminous accretion of gas. The final observed BH mass would be $10^6$ times the seed mass, after $\approx n$ Salpeter time-scales. The scaling relations would rise in normalization but retain the original slope: $m_\star \approx 10^6 m_\odot \sim \sigma^{-5/2}$, if $F \approx 9)$. The index of SMBH scaling is preserved from our simplistic gasless halo model.

Note that there are always uncertainties and complications in the accounting of total SMBH mass and radiative efficiency of their growth. For example, recoiling SMBHs can escape their galaxies after a merger (e.g. Redman & Rees 1989; Menou, Haiman & Narayanan 2001; Haiman 2004; Madau & Quataert 2004; Baker et al. 2006; Campanelli et al. 2007a, b; Gonzalez, Zaritsky & Zabludoff 2007; Schnittman & Buonanno 2007; Lousto & Zlochower 2011, 2013), and end up dormant in intergalactic space: in that case, simple counts of nuclear SMBHs would underestimate the total cosmic BH mass. The local SMBH density may also have been underestimated if there is a previously unrecognized population of SMBHs in ultracompact dwarf galaxies (Seth et al. 2014). The presence of ultramassive BHs may require more accretion (via radiatively inefficient modes) than reckoned before (McConnell et al. 2011, 2012; van den Bosch et al. 2012; Fabian et al. 2013). The discovery of modern-sized quasars at high redshift is likely to require a faster early growth than allowed by Eddington-limited luminous accretion (Fan et al. 2004; Shapiro 2005; Mortlock et al. 2011; Venemans et al. 2013); on the other hand, the X-ray background from high-redshift AGN is dimmer than expected, contradicting rapid radiatively efficient gas accretion in the $z > 5$ era (Willott 2011; Salvaterra et al. 2012; Treister et al. 2013). The radiative efficiency of quasar accretion may be lower than the standard $\varepsilon \sim 0.1$ disc efficiency during supercritical gas accretion phases (Novak 2013) or due to DM accretion; however, accretion can instead appear more radiatively efficient than $\varepsilon \sim 0.1$, for example if it taps into the BH spin (Narayan, Igumenshchev & Abramowicz 2003; Igumenshchev 2008; Lasota et al. 2014; Tchekhovskoy et al. 2014) or if DM envelope dominates the inner potential. Finally, SMBHs might grow via BH–BH coalescence without any radiative emission; however, constraints set by the cosmic gravitational-wave background imply that steady accretion (of gas or DM) dominates (Shannon et al. 2013). Thus, the issues of how early SMBHs were seeded, the role of DM in setting the seed mass (e.g. Mack, Ostriker & Ricotti 2007; Dotan, Rossi & Shaviv 2011; Lora-Clavijo et al. 2014) and the DM mass contribution are far from settled.

Our halo model has some similarities to the supermassive star scenario that aims to explain the early SMBH seeding. The proposal is that a $\gtrsim 10^6 m_\odot$ polytropic sphere of gas (e.g. Hoyle & Fowler 1963; Iben 1963; Fowler 1964; Shibata & Shapiro 2002) burns and collapses to produce a seed BH that is born supermassive, thereby reducing the feeding time needed to reach observed SMBH scales (e.g. Begelman, Volonteri & Rees 2006; Begelman 2010; Johnson et al. 2013). The main doubt about this scenario is that the gas may not collapse into a single supermassive object, and may instead fragment into clumps and star clusters because of its angular momentum. Even if a single supermassive star were formed, it may not survive long enough to develop a core and collapse into a single BH, due to mass losses in intense winds. Our model would create MBH seeds from polytropic DM instead. Eddington limits and winds do not apply to SIDM seeding. Whichever way real SMBH originated, we expect a scaling like $m_\star \sim \sigma^{5/2}$ to emerge from the direct or indirect coupling of the SMBH and halo in equilibrium, since the equilibrium state is independent of what fed the SMBH previously.

4.7 Late-type galaxies

It has long been a puzzle to explain why ellipticals, lenticulars and early-type spirals have a nuclear SMBH, while many late-type spirals have a nuclear star cluster but no SMBH. M33 and NGC205 are local examples of the latter (Gerhard et al. 2001; Merritt, Ferrarese & Joseph 2001; Valluri et al. 2005). Even more puzzling is the fact that the nuclear star cluster mass versus $\sigma$ relation runs parallel to the $m_\star$ scaling relations (Ferrarese et al. 2006; Graham & Spitler 2009; Graham 2012a). On the other hand, some bulgeless galaxies
do possess a nuclear SMBH (e.g. Filippenko & Ho 2003; Peterson et al. 2005; Shields et al. 2008; Araya Salvo et al. 2012; Reines et al. 2013; Simmons et al. 2013).

Salucci et al. (2000) observed that late-type galaxies have SMBH that are undersized compared to the usual trend with bulge mass ($M_\bullet$). It is argued those galaxies only have pseudo-bulges (evolved quiescently from the disc via secular processes), whereas SMBH correlate with classical bulges (Kormendy & Bender 2011). Alternatively, perhaps the $m_\bullet$ relation bends downwards in the low-$M_\star$ domain (Graham 2012b; Scott et al. 2013) and the SMBH relation to $\sigma$ is straighter. This hints that $\Phi$ plays the fundamental role, consistent with our thesis linking the SMBH to the halo. Either way, the hints of some dependence on luminous morphology (besides the DM halo) deserve an explanation within our theory.

It is worth noting some exemptions from the SMBH mass prediction of equations (14) and Appendix A. If the velocity dispersion $\sigma$ is non-relativistic everywhere in the profile, then there need not be an event horizon at the centre. A non-singular halo does not grow any central compact mass. This is the lowestentropy condition available. We propose that protogalaxies condensed in this initial state, and some would grow quiescently (without major mergers or gas expulsions) till the present epoch. Those are tranquill disc galaxies, near the non-singular border, lacking classical bulges, and having undersized SMBH or none at all. For other galaxies, tidal harassment or minor mergers would raise the entropy (lowering $q$), inducing a more centrally peaked density profile. Perhaps, if the central DM becomes concentrated enough, a seed BH forms. Subsequent large-scale gas inflows accrete on to the SMBH in a quasar phase. These galaxy haloes enter the ‘plateau region’; they follow the maximum $m_\bullet/M$ scaling relation. For those that suffer more major mergers, the luminous disc converts partially into a classical bulge, or totally into an elliptical. In contrast, for the undisturbed, high-$q$ non-singular galaxies, if the inner halo never became dense enough, it does not form the initial BH, and the same large-scale gas inflows produce a nuclear star cluster. The mass in this nuclear star cluster is comparable to the baryonic mass that would have fed the SMBH. We speculate that the knee in the $M_\bullet$ correlation (Graham 2012b) or the underweight SMBH of late-type galaxies (Salucci et al. 2000) may occur:

(i) because the latest-type galaxies are near the high-$q$ non-singular border and their $m_\bullet/M$ is below the relations in Fig. 6; or
(ii) because these galaxies are near one of the knees in a ribbon relation such as those in Fig. 6; or
(iii) the bulge is incidental and the halo determines $m_\bullet$.

Though it is beyond the scope of our spherical modelling, we speculate that the angular momentum of the halo and gas may also affect the outcome. If the inner halo possesses too much angular momentum (and cannot shed it via large-scale dark turbulence), then rotational support inhibits collapse. If the baryons have effective rotational support, then they may not achieve the central densities needed to trigger the inner halo to condense a seed. The result is a pure disc galaxy without a central BH.

4.8 Stellar components

Our gasless and starless model is a simplification. In principle, a galaxy’s stellar mass distribution affects the SMBH/halo equilibrium to some extent. In galaxy clusters, Saxton & Wu (2008, 2014) found that the continuity requirements of gas inflows impose lower limits on $m_\bullet$, however inserting a central galaxy’s stellar profile did not alter these constraints greatly. An isolated elliptical galaxy’s stellar spheroid compresses the dark core slightly (Saxton & Ferreras 2010; Saxton 2013). None the less, DM always dominates in the outer halo. DM should also dominate baryons at the centre: within the innermost stellar orbit, and perhaps throughout the SMBH sphere of influence. Visible matter is most influential at medium radii (kpc for an elliptical galaxy).

Our present models omit stellar profiles, as we are most interested in the link between the DM halo and the SMBH. Because observations already show that these properties correlate, we suspect that the stellar mass does not dominate SMBH scaling relations outright. This motivates our comprehensive exploration of baryon-free configurations. Our model has two components and three key parameters: thermal degrees of freedom ($F$), compactness ($Q$) and entropy ($s$, via $Q$ and $F$). Adding one more density component will increase the complexity of the formulation, if we want a self-consistent treatment. This topic is worth a separate study, and we intend to resume it elsewhere. However, we would also like to comment qualitatively here. The addition of a stellar spheroid entails three more free variables: total stellar mass ($M_\star$), a half-light radius ($R_e$) and Sérsic shape index ($n$), vastly increasing the system’s dimensionality. We ran restricted tests of $F = 9$ models where the stars comprise 10 per cent of the mass. In a preliminary way, we note:

(i) If the stellar component is compact ($R_e \ll R$), it exerts little influence on the scaling relations. This is understandable since this bulge behaves somewhat like a central concentrated point, which is effectively the same as the SMBH.
(ii) For any terraced or singular model, the DM dominates at a sufficiently small radius. The stellar component is also sub-dominant at the radius of halo core and outskirts beyond $r \gg kpc$. The stellar potential only perturbs the DM density profile locally at intermediate radii.
(iii) Theoretically, the worst scenario is when the DM and the stellar component have similar compactness. Even then, we find that the basic conclusion holds, except that $\psi$ and $m_\bullet/M$ values shift across the parameter plane. This shift is only significant near the non-singular border. We will leave the detailed discussion for our next paper.

The robustness of SMBH versus halo scaling relations, in spite of a stellar contribution and medium radii, might be foreseeable on qualitative grounds. The halo core depends on heat capacity and entropy. The location of the event horizon (which sets $m_\bullet$) coincides with an effectively universal maximum of $\sigma^2$. Both defining structures depend straightforwardly on the gravitational potential and velocity dispersion, which are linearly related. Their correlation arises naturally. Essentially and generally, the empirical SMBH scaling relations reveal how density $\rho$ is stratified with respect to potential $\Phi$ in galaxies.

5 CONCLUSIONS

We investigate the properties of spherical, adiabatic self-gravitating systems with the DM microphysics prescribed by an equation of state. These states form a halo of DM and a central compact object. We have found that the halo profile is determined by two necessary parameters. One possible combination is the gravitational compactness $\chi$ (equation 10) and a measure of (pseudo-)entropy $s$ (or equivalently the phase-space density $Q$). Characterization of such halo profile in terms of a single parameter (e.g. asymptotic or peak circular velocity; see Ferrarese 2002; Baes et al. 2003;
The recently observed correlations between SMBH and velocities feedback (which operates at the centre). The finding that controlled by the underlying DM potential rather than by AGN the far outskirts supports an idea that SMBH scaling relations are

The consistency of these correlations (especially GC properties at clusters (Saxton & Wu 2008, 2014; Saxton & Ferreras 2010).

The SMHB mass scales with the characteristic velocity dispersion, \( m_2 \sim \sigma_c^{1/2} \), with effective thermal degrees of freedom \( F \) as the scaling index. Given that bulge stars and DM particles bound in the same potential well have similar velocity dispersions, the observed \( m_2 \) versus \( \sigma_c \), scaling relation indicates that \( F \gtrsim 7 \) for the dark halo. The recently observed correlations between SMBH and velocities of halo GC swarm \( \sigma_\star \) are also consistent with this conclusion. The consistency of these correlations (especially GC properties at the far outskirts) supports an idea that SMBH scaling relations are controlled by the underlying DM potential rather than by AGN feedback (which operates at the centre). The finding that \( F \gtrsim 7 \) implies that DM has large effective degrees of freedom, which we might interpret as a large heat capacity, or perhaps a steep index of a self-interaction potential. These values agree with the range indicated in some previous modelling of elliptical galaxies and galaxy clusters (Saxton & Wu 2008, 2014; Saxton & Ferreras 2010).

These models also tend to predict that a dense dark envelope surrounds the SMBH. In at least some systems, the envelope may have non-negligible density compared to the SMBH itself. In extreme cases, the dark envelope outweighs the SMBH. This might have observable consequences in the relativistic vicinity of the event horizon. Useful tests might involve apparent sizes of SMBH horizons, the tidal disruption of stars, and the inner structure of AGN accretion discs.

ACKNOWLEDGEMENTS

We thank the referee, P. Salucci, for helpful criticisms and suggestions that improved the scope and focus of our results and commentary. We thank A.W. Graham for discussions of SMBH scaling, and M. Cropper for discussions of GC populations. This work has made use of NASA’s Astrophysics Data System. Our calculations employed mathematical routines from the GNU Scientific Library. This publication has made use of code written by James R. A. Davenport. Specifically, the figures’ colour scheme was developed by Green (2011).

REFERENCES


\[ \text{http://www.astro.washington.edu/users/jrad/idl.html} \]
\[ \text{http://www.mrao.cam.ac.uk/~dag/CUBEHELIX/} \]

Bloom J. S. et al., 2011, Science, 333, 203

Downloaded from http://mnras.oxfordjournals.org/ at University College London on February 3, 2015
APPENDIX A: SMBH PREDICTION IN ABSOLUTE TERMS

The equation (14) for the SMBH mass can be written in various absolute units for practical applications. The choice of units depends on context. For example, in the vicinity of the dark envelope and the circumnuclear region, velocity dispersions are almost relativistic. DM densities could become comparable to that of baryonic matter on Earth. In units suitting that environment, the SMBH mass in solar units is

\[
m_* \approx \frac{4.2919 \times 10^9}{\sqrt{\eta \psi}} \left( F + 2 \right)^{4/3} \left( \frac{1 \text{ kg m}^{-3}}{\rho} \right)^{1/2} \left( \frac{\sigma}{c} \right)^{2/3}.
\]

(A1)

Farther out, in the kpc-scale core of the galaxy’s halo, typical velocities drop to the order of 100 km s\(^{-1}\). DM core densities are multiples or fractions of 1 m\(_\odot\) pc\(^{-3}\). In these terms, the predicted central mass (solar units) is

\[
m_* \approx \frac{1.6495 \times 10^{19}}{\sqrt{\eta \psi}} \times \left( \frac{1 \text{ m} \odot \text{ pc}^{-3}}{\rho} \right)^{1/2} \left( \frac{\sigma}{100 \text{ km s}^{-1}} \right)^{2/3}.
\]

(A2)

An equivalent logarithmic form says

\[
\log_{10} \left( \frac{m_*}{\text{m}_\odot} \right) \approx 19.217 - 1.7384 F + \frac{F}{4} \log_{10} \left( \frac{F + 2}{1 - \chi} \right) - \frac{1}{2} \log_{10} \left( \frac{\psi}{\eta} \right) - \frac{1}{2} \log_{10} \left( \frac{\rho}{1 \text{ m}_\odot \text{ pc}^{-3}} \right) + \frac{F}{2} \log_{10} \left( \frac{\sigma}{100 \text{ km s}^{-1}} \right).
\]

(A3)

The third term on the right-hand side is \(\lesssim 2.7\) when \(\chi \ll 1\). The astronomical mass range \(m_* \lesssim 10^{10} \text{ m}_\odot\) implies that either \(F > 6\) (in the second term of the right-hand side) or there is a large correction factor \(\eta \psi\) (in the fourth term on the right).

APPENDIX B: MODEL HOMOLOGIES AND SCALE-INVARINAT PARAMETRIZATION

Given a particular polytropic halo model, a family of homologous models can be formed by multiplying each quantity \(y\) by a scale factor \(X_s\). Since we take the speed of light as an absolute reference scale for velocity dispersions, escape velocities and gravitational potentials, we necessarily have \(X_s = 1\). For a scale factor \(X_s = 1\) and...
X_0 = 1. It follows that model masses and distances must rescale by the same factor, X_0 = X_0 \equiv X, and densities rescale as \( X_0 = X_0/X_0 = X^{-2} \). The phase-space density rescales as \( X_0 = X_0/F = X^{-2} \). For example, if we choose to standardize a set of models so that they have the same total mass \( M \), we could transform the phase-space densities as \( Q \rightarrow Q/R^2 \).

We prefer to classify and compare models in terms of their dimensionless properties that remain constant under the homology transformations. Dimensionless quantities such as \( \chi, \eta \) and \( \psi \) remain constant under the homology transformations. If the outer boundary conditions are known, then it is possible to define a dimensionless variable related to \( Q \), for instance \( q \equiv Q/V F / \rho \) for which \( Xq = 1 \). Similarly, \( l \equiv M^2 Q \) for which \( Xl = 1 \). The properties of the central object are best described in terms of invariant fractional quantities such as the \( m_\bullet/M \) and \( r_\bullet/R \).

**APPENDIX C: ENTROPY CALCULATION**

The Bekenstein (1973) entropy of an event horizon is \( S_\bullet = kA/4l_P^2 \), where \( A \) is the surface area, \( k \) is Boltzmann’s constant, \( l_P = Gm_P/c^2 \) is the Planck length and \( m_P \) is the Planck mass. Substituting the area of the inner boundary of our model, \( A \approx 4\pi r_\bullet \), we have \( S_\bullet = \pi k(r_\bullet/l_P)^2 \), which simplifies:

\[
S_\bullet = \pi k \left( \frac{c^2 r_\bullet}{Gm_P} \right)^2 = \pi k \left( \frac{c^2 2Gm_\bullet \eta}{Gm_Pc^2} \right)^2 = 4\pi k \left( \frac{m_\bullet \eta}{m_P} \right)^2. \tag{C1}
\]

Since the total mass of the system is \( M \), the mass of DM outside the SMBH is \( M - m_\bullet \), and the number of dark particles is \( N = (M - m_\bullet)/\mu \). The DM halo entropy is \( S_\delta = -Nk \ln (Q/Q_0) \). For the total entropy,

\[
S = 4\pi k \left( \frac{M}{m_P} \right)^2 \left( \frac{m_\bullet}{M} \right)^2 \eta^2 - M/\mu \left( 1 - m_\bullet/M \right) \ln \left( \frac{Q}{Q_0} \right)
= \frac{Mk}{\mu} \left[ 4\pi \frac{M \mu}{m_P^2} \left( \frac{m_\bullet}{M} \right)^2 \eta^2 - \left( 1 - m_\bullet/M \right) \ln \left( \frac{Q}{Q_0} \right) \right]. \tag{C2}
\]

The first term (entropy of the horizon) dominates if \( \mu \gg m_P^2/M \), and the second term (entropy of the DM halo) dominates if \( \mu \ll m_P^2/M \).

Also note the trivial algebraic identity,

\[
\frac{m_\bullet}{M} \eta = \frac{m_\bullet r_\bullet}{Ml_P} = \frac{c^2 m_\bullet r_\bullet}{2GMm_\bullet} = \frac{c^2 r_\bullet}{2GM} = \frac{r_\bullet}{R \chi}. \tag{C3}
\]

For fixed \( \chi \), the ratio \( m_\bullet/M \) is a monotonic function of \( Q \), and \( \eta \) remains on the order of 1. Also for fixed \( \chi \), the ratio \( r_\bullet/R \) is monotonic in \( Q \) except for wrinkles within one dex of the non-singular border. Therefore, if the right-hand term of (C2) dominates, then \( S \) is monotonic in \( Q \); and if the left-hand term dominates, then \( S \) is also monotonic in \( Q \) (except for subtle features near the non-singular boundary).