

# Gravitational recoil: effects on massive black hole occupation fraction over cosmic time

Marta Volonteri,<sup>1\*</sup> Kayhan Gültekin<sup>1</sup> and Massimo Dotti<sup>1,2</sup>

<sup>1</sup>*Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA*

<sup>2</sup>*Max-Planck-Institut für Astrophysik, Garching bei München, 85740, Germany*

Accepted 2010 January 26. Received 2010 January 22; in original form 2009 December 21

## ABSTRACT

We assess the influence of massive black hole (MBH) ejections from galaxy centres due to gravitational radiation recoil, along the cosmic merger history of the MBH population. We discuss the ‘danger’ of recoil for MBHs as a function of different MBH spin-orbit configurations and of the host halo cosmic bias, and on how that reflects on the occupation fraction of MBHs. We assess ejection probabilities for mergers occurring in a gas-poor environment, in which the MBH binary coalescence is driven by stellar dynamical processes and the spin-orbit configuration is expected to be isotropically distributed. We contrast this case with the ‘aligned’ case. The latter is the more realistic situation for gas-rich, i.e. ‘wet’, mergers, which are expected for high-redshift galaxies. We find that if *all* haloes at  $z > 5$ –7 host an MBH, the probability of the Milky Way (or similar size galaxy) to host an MBH today is less than 50 per cent, unless MBHs form continuously in galaxies. The occupation fraction of MBHs, intimately related to halo bias and MBH formation efficiency, plays a crucial role in increasing the retention fraction. Small haloes, with shallow potential wells and low escape velocities, have a high ejection probability, *but* the MBH merger rate is very low along their galaxy formation merger hierarchy: MBH formation processes are likely inefficient in such shallow potential wells. Recoils can decrease the overall frequency of MBHs in small galaxies to  $\sim 60$  per cent, while they have little effect on the frequency of MBHs in large galaxies (at most a 20 per cent effect).

**Key words:** black hole physics – galaxies: formation – cosmology: theory.

## 1 INTRODUCTION

It is well established that galaxies contain massive black holes (MBHs) in their nuclei (e.g. Kormendy & Richstone 1995; Richstone et al. 1998). Observations of the host elliptical galaxy and the galactic spheroid luminosity and the stellar velocity dispersion (Magorrian et al. 1998; Ferrarese & Merritt 2000; Gebhardt et al. 2000). Such relationships, as well as their small intrinsic scatter (Tremaine et al. 2002; Gültekin et al. 2009), hint at a connection between the formation of the MBH and the formation of the spheroid.

In the context of the favoured cold dark matter cosmology, hierarchical assembly of galaxies and protogalaxies with MBHs at their centres naturally leads to the formation of MBH binaries (Begelman, Blandford & Rees 1980). If the MBH binary can achieve a small enough separation through stellar dynamical hardening or through viscous gas dynamical drag, gravitational radiation from the binary can become significant enough to cause the system to

merge. Gravitational waves carry away energy from the binary system, driving the system to merger. If the black holes have unequal masses or misaligned spins, anisotropic gravitational radiation will impart a net momentum flux on the binary, causing the centre of mass to recoil.

Until recently, the magnitude of the recoil velocity was uncertain. For non-spinning black holes the maximum recoil has now been calculated to be  $v_{\text{recoil,max}} \approx 200 \text{ km s}^{-1}$  (Baker et al. 2006), and a similar range is expected for black holes with low spins, or with spins (anti-)aligned with the binary orbital angular momentum. Several studies have also found consistent results for the maximum recoil of spinning black holes  $v_{\text{recoil,max}} = 1000$ – $4000 \text{ km s}^{-1}$  (Baker et al. 2007; Campanelli et al. 2007; Gonzalez et al. 2007; Herrmann et al. 2007; Koppitz et al. 2007; Schnittman & Buonanno 2007). Such velocities are significant because they will eject the binary from the galaxy since most galaxies have escape speeds  $< 1000 \text{ km s}^{-1}$ .

Volonteri, Haardt & Gültekin (2008b) studied the effect of recoil on the MBH occupation fraction in nearby galaxies. They assumed that the relative orientation between the orbital angular momentum of the binary and the spins of the two MBHs were isotropically distributed. This configuration can result in high recoil velocities, up to

\*E-mail: martav@umich.edu

thousands  $\text{km s}^{-1}$ . However, Bogdanović, Reynolds & Miller (2007) proposed that MBHs orbiting in gaseous circumnuclear discs, such as those expected in advanced stages of gas-rich galaxy mergers (Mayer et al. 2007), align their spins with the orbital angular momentum of the binary. This configuration leads to small recoils for the MBH remnant. The accreting gas exerts gravito-magnetic torques that suffice to align the spins of both the MBHs with the angular momentum of the large-scale gas flow. Dotti et al. (2010) have quantified the efficiency of this alignment process through the analysis of high-resolution  $N$ -body Smoothed Particle Hydrodynamics (SPH) simulations. They apply the algorithm presented in Perego et al. (2009) to evolve masses, magnitudes and orientation of the spins of two MBHs.

We extend the investigation of Volonteri et al. (2008b) considering different degrees of alignment between the spins and the angular momentum of the MBHs. We will use either isotropic spins and angular momentum configurations or the ‘quasi-aligned’ distributions of spin orientations from Dotti et al. (2010). Those two cases bracket all the possible values of recoil velocities. We apply these distributions to Monte Carlo realizations of galaxy merger trees, and we study the effect of gravitational recoil on the occupation fraction of MBHs in galaxies at different redshifts.

## 2 RECOIL VELOCITIES

### 2.1 Fitting formulae

The recoil velocity may be broken down into components arising solely from mass asymmetry,  $v_m$ , which is perpendicular to the orbital angular momentum vector  $\mathbf{L}_{\text{pair}}$ , and components arising from spin asymmetry,  $v_{\perp}$  and  $v_{\parallel}$ , which are perpendicular and parallel to  $\mathbf{L}_{\text{pair}}$ , respectively:

$$v_{\text{recoil}} = \sqrt{v_m^2 + v_{\perp}^2 + 2v_m v_{\perp} \cos(\xi) + v_{\parallel}^2}, \quad (1)$$

where  $\xi$  is the angle between  $v_m$  and  $v_{\perp}$  in the orbital plane. We assumed  $\xi = 145^\circ$ , as suggested by Lousto & Zlochower (2009). Campanelli et al. (2007) and Lousto & Zlochower (2009, fit CL) propose the following fitting formulae for the recoil components:

$$v_m = A\eta^2 \sqrt{1 - 4\eta} (1 + B\eta), \quad (2)$$

$$v_{\perp} = H\eta^2 (1 + q)^{-1} (a_1^{\parallel} - qa_2^{\parallel}), \quad (3)$$

$$v_{\parallel} = K\eta^2 (1 + q)^{-1} \cos(\Theta - \Theta_0) |a_1^{\perp} - qa_2^{\perp}|, \quad (4)$$

where  $q = M_2/M_1 \leq 1$  is the mass ratio of the black holes and  $\eta \equiv q/(1+q)^2$  is the symmetric mass ratio. The components of the spins of the two MBHs are broken into projections parallel and perpendicular to  $\mathbf{L}_{\text{pair}}$ ,  $\mathbf{a}^{\perp} = \mathbf{a} \sin(\theta)$  and  $\mathbf{a}^{\parallel} = \mathbf{a} \cos(\theta)$ . In the isotropic case,  $\cos(\theta_1)$  and  $\cos(\theta_2)$  are distributed uniformly between  $-1$  and  $1$ . We compare the results from the isotropic case with those obtained assuming quasi-aligned spin-orbit configurations. In this case we assumed that  $\theta_1$  and  $\theta_2$  were distributed according to Dotti et al. (2010, see Section 2.2).  $\Theta$  is the angle between  $(\mathbf{a}_2^{\perp} - q\mathbf{a}_1^{\perp})$  and the separation vector at coalescence, and  $\Theta_0$  depends on the initial separation between the holes. We assume a uniform distribution of  $\Theta - \Theta_0$  between  $0$  and  $2\pi$ . Note that for  $qa_{1\perp} = a_{2\perp}$  (including  $a_{1\perp} = a_{2\perp} = 0$ ),  $v_{\parallel}$  vanishes and there is no recoil out of the orbital plane. In the absence of spins, the recoil is  $v_m$ , which is maximized for  $q = (3 - \sqrt{5})/2 \approx 0.38$ . The best-fitting parameters found by Lousto & Zlochower (2009, which we refer to as fit CL) are

$$A = 1.2 \times 10^4 \text{ km s}^{-1}, B = -0.93, H = 6900 \text{ km s}^{-1} \text{ and } K = 6.0 \times 10^4 \text{ km s}^{-1}.$$

Baker et al. (2008, referred to as fit B) found an alternative form for  $v_{\parallel}$  that scales as  $\eta^3$ :

$$v_{\parallel} = K\eta^3 (1 + q)^{-1} (a_1^{\perp} \cos(\Phi_1) - qa_2^{\perp} \cos(\Phi_2)), \quad (5)$$

where  $\Phi_1$  and  $\Phi_2$  are the differences between two angles that depend on the system of reference. We assume  $\Phi_1 = \Phi_2$  uniformly distributed between  $0$  and  $2\pi$  (see Dotti et al. 2009). The best-fitting parameters from Baker et al. (2008) are  $A = 1.35 \times 10^4 \text{ km s}^{-1}$ ,  $B = -1.48$ ,  $H = 7540 \text{ km s}^{-1}$  and  $K = 2.4 \times 10^5 \text{ km s}^{-1}$ .

Herrmann et al. (2007, fit H in the following) parametrize the recoil velocity in terms of  $\theta_H$ , the angle between  $\mathbf{L}_{\text{pair}}$  and

$$\boldsymbol{\Sigma} = (M_1 + M_2) \left( \frac{\mathbf{J}_1}{M_1} - \frac{\mathbf{J}_2}{M_2} \right), \quad (6)$$

where the  $\mathbf{J}$  are the spin angular momenta of the holes. Taking  $\mathbf{L}_{\text{pair}}$  to be in the  $z$  direction, they find the Cartesian components of the recoil velocity to be

$$\begin{aligned} V_x &= C_0 H_x \cos(\theta_H), \\ V_y &= C_0 H_y \cos(\theta_H), \\ V_z &= C_0 K_z \sin(\theta_H), \end{aligned} \quad (7)$$

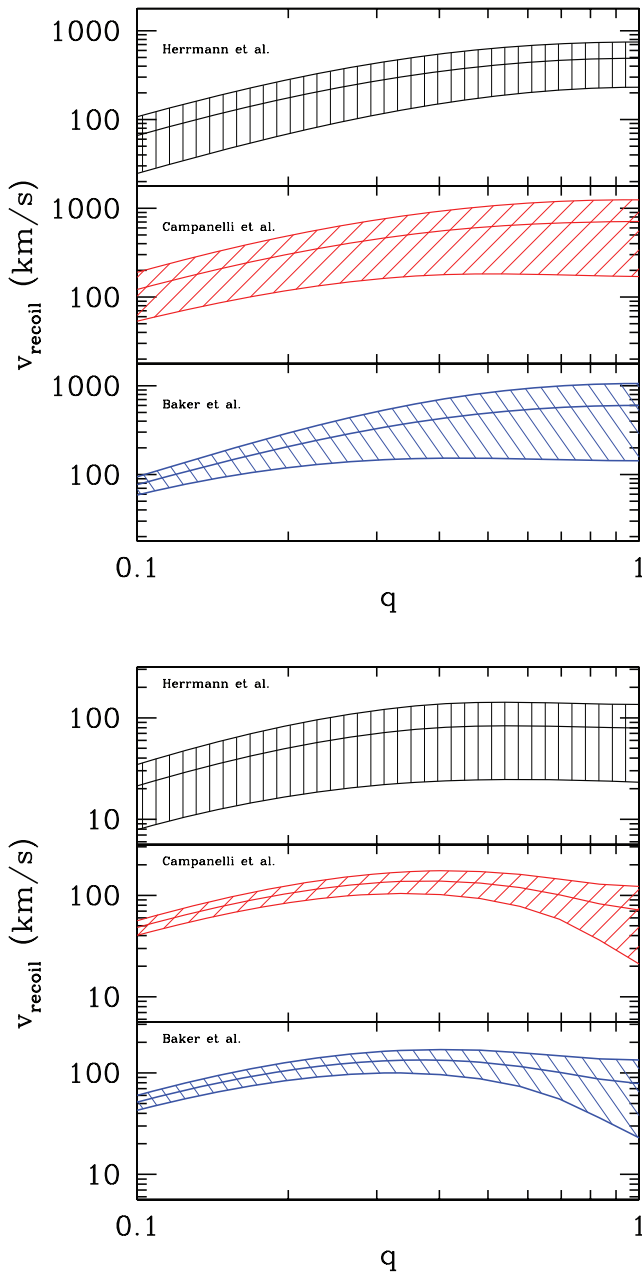
where  $C_0 = \Sigma^2 (M_1 + M_2)^{-2} (1 + q)^{-4}$  with the best-fitting parameters  $H_x = 2.1 \times 10^3$ ,  $H_y = 7.3 \times 10^3$  and  $K_z = 2.1 \times 10^4$ .

In Fig. 1 we plot recoil velocities from fits H, CL and B as a function of  $q$  for a sample of isotropically distributed MBH spins and for aligned configurations. This figure demonstrates the agreement of the three formulae for different spin-orbit configurations. For each value of  $q$  between  $0.1$  and  $1$ , we produce 500 000 realizations of the two MBH spins with magnitudes uniformly distributed between  $0$  and  $1$ . For the isotropic case we also assign isotropically distributed spin directions. We do not show results for  $q < 0.1$  because the recoil velocity drops below the typical escape velocity from galaxies.

Fit B decreases faster as  $q$  decreases, with respect to fits CL and H. This is because the component of the recoil velocity parallel to  $\mathbf{L}_{\text{pair}}$  scales as  $\eta^3$  in fit B, and as  $\eta^2$  in fits CL and H. The average values of  $v_{\text{recoil}}$  obtained with the different prescriptions for isotropically distributed spins are consistent within  $1\sigma$  (i.e. within a factor of  $\sim 2$ ). Assuming perfect alignment between the two spins and  $\mathbf{L}_{\text{pair}}$ , the results from fits CL, H and B are consistent within a factor of  $2.5$  for any  $q$ . For the aligned case with  $q \approx 0.1$ , the difference between fits CL and B disappears because, for perfect alignment,  $a = a^{\parallel}$  and  $v_{\parallel} = 0$ , resulting in a perfect agreement between the two fits.

### 2.2 Quasi-aligned configurations

As mentioned in Section 1, gravo-magnetic torques exerted by accreting flows on to the MBHs tend to align  $\mathbf{a}_1$  and  $\mathbf{a}_2$  to  $\mathbf{L}_{\text{pair}}$ . The distributions of relative angles between the two MBH spins, and between each spin and  $\mathbf{L}_{\text{pair}}$  after the formation of an MBH binary, have been computed in Dotti et al. (2010, see fig. 2 therein). Their results stem from the analysis of high-resolution  $N$ -body/SPH simulations. In the suite of runs discussed in that paper, the authors varied the thermodynamical properties of the gas. In particular they assumed two different polytropic equations of state, with polytropic index  $\gamma = 5/3$  (‘hot’ runs) and  $7/5$  (‘cold’ runs), respectively. The hot case corresponds to an adiabatic monoatomic gas, as if radiative cooling were completely suppressed during the merger. This case mimics gas radiatively heated by an AGN (Mayer et al. 2007). The cold case, instead, has been shown to provide a good approximation



**Figure 1.** Recoil velocity of the MBH remnant as a function of  $q$ , for MBH spins isotropically distributed (top) or for aligned configurations (bottom). The upper, middle and lower panel refer to fit H, fit CL and fit B, respectively. The thick lines show the average values of  $v_{\text{recoil}}$  ( $\bar{v}_{\text{recoil}}$ ) and  $\bar{v}_{\text{recoil}} \pm \sigma$ .

to a gas of solar metallicity heated by a starburst (e.g. Spaans & Silk 2000).

In each run, Dotti et al. (2010) found a significant degree of alignment already present before the formation of a binary. However, their runs always start with an equal-mass MBH pair. Because the dynamical evolution of the two MBHs is faster than the Salpeter time, the masses of the two MBHs do not change significantly, and at the end of the runs a nearly equal-mass binary forms. Here we discuss the dependence of spins/orbit configurations discussed by Dotti et al. (2010) on the masses of the MBHs, by comparing the dynamical time-scale ( $\tau_{\text{dyn}}$ ) of the binary formation with the alignment time-scale ( $\tau_{\text{align}}$ ). Before we consider the ratio of these two

time-scales, we look at each in slightly more detail in the context of our simulations; but for a full discussion, see Dotti et al. (2009).

Dynamical friction is the process that drives the dynamical evolution of the pair. Since the longer time-scale is what determines the time until coalescence,  $\tau_{\text{dyn}}$  is the dynamical friction time-scale of the less massive MBH, which depends on a number of factors:

$$\tau_{\text{dyn}} \propto K M_{\text{BH}}^{-1}, \quad (8)$$

where  $K$  is a function of the properties of the circumnuclear disc. In Dotti et al. (2010) runs,  $M_{\text{BH}} = 4 \times 10^6 M_{\odot}$  and  $\tau_{\text{dyn}} = 4.5\text{--}7.5$  Myr, depending on the initial orbital parameters of the BHs and on the effective equation of state used to evolve the gas thermodynamics. Here we are considering the circumnuclear disc parameters in Dotti et al. (2009) as fiducial. Considering different values for these is out of the scope of this paper, but it is likely to change only the details and not our qualitative results.

The alignment time-scale also depends on the BH spin,  $a$ , and may be expressed in terms of the Eddington fraction [see equation (43) in Perego et al. 2009]:

$$\tau_{\text{align}} \propto a^{5/7} M_{\text{BH}}^{-2/35} f_{\text{Edd}}^{-32/35}. \quad (9)$$

For  $M_{\text{BH}} = 4 \times 10^6 M_{\odot}$  in the simple case of a fixed orbital plane of the outer disc, i.e. coherent accretion, the value of  $\tau_{\text{align}} \approx 10^5$  yr. For the simulations in Dotti et al. (2009), however, the orbital plane of the accreted particles is never constant, and thus the alignment time-scale will be longer:  $\tau_{\text{align}} \approx 1\text{--}4$  Myr, depending on the initial orbital parameters of the BHs and the effective equation of state used.

Since both black hole spins must achieve alignment with the disc angular momentum for the spins to be aligned with each other, the alignment time-scale we are interested in is the longer of the two. For most cases, this will be the time-scale for the less massive binary. Before the formation of a binary, the more massive of the two MBHs always accretes more mass, and has the shortest alignment time-scale (Dotti et al. 2010). Thus, we may use  $M_{\text{BH}} = M_2$  in the expression for  $\tau_{\text{align}}$ .

The situation after the formation of a hard binary, however, is slightly different. In this case, the torques exerted by the binary on to the gas carve out a low-density region in the middle of the disc, and the secondary MBH, being closer to the outer and denser gas region, can grow faster than the primary (Hayasaki, Mineshige & Sudou 2007; Cuadra et al. 2009), further aligning its spin with  $\mathbf{L}_{\text{pair}}$ . For our purposes, we conservatively assume that after forming a binary, the spin-orbit configuration at the time of binary formation does not significantly evolve.

The ratio of these two quantities is obtained by dividing equation (9) by equation (8) and shows the dependence on the accretion behaviour of the two MBHs:

$$\tau_{\text{align}}/\tau_{\text{dyn}} \sim a^{5/7} M_2^{-2/35} f_{\text{Edd}}^{-32/35}. \quad (10)$$

For a fixed value of  $a$  and rewriting in terms of  $f_{\text{Edd}} \sim \dot{M} M_2^{-1}$ , this expression becomes

$$\tau_{\text{align}}/\tau_{\text{dyn}} \sim \dot{M}_2^{-32/35} M_2^{13/7}. \quad (11)$$

If they are accreting at the Bondi rate,  $\dot{M}_2 \propto M_2^2$ , and consequently  $\tau_{\text{align}}/\tau_{\text{dyn}} \propto M_2^{1/35}$ . In this case the ratio is almost independent of the mass of the secondary, and the assumption that the two MBHs are almost aligned is always correct.

If, instead, the secondary is accreting at the Eddington limit,  $\dot{M}_2 \propto M_2$ . From Dotti et al. (2009), we know that for

$M_2 \approx 4 \times 10^6 M_\odot$ ,  $\tau_{\text{align}}/\tau_{\text{dyn}} \approx 0.3$ .<sup>1</sup> Using this as our normalization, we get  $\tau_{\text{align}}/\tau_{\text{dyn}} \approx 0.1 M_{2,6}^{33/35}$ , where  $M_{2,6}$  is the mass of the secondary in units of  $10^6 M_\odot$ . This scaling implies that a binary can form before the two MBHs align their spins to  $\mathbf{L}_{\text{pair}}$  only for very massive secondaries ( $M_2 \gtrsim 2 \times 10^7 M_\odot$ ) and rapidly accreting pairs.

To estimate the relevance of such mergers, we consider the prevalence of the *secondary* MBH having mass  $M_2 \gtrsim 2 \times 10^7 M_\odot$ . Assuming the scaling of MBH masses with the velocity dispersion of the host determined for nearby galaxies (Gültekin et al. 2009), the host of an MBH with mass  $M_2 \gtrsim 2 \times 10^7 M_\odot$  has  $\sigma \gtrsim 130 \text{ km s}^{-1}$ , and an escape velocity  $\gtrsim 1000 \text{ km s}^{-1}$ . The recoil velocity has a significant probability of being  $> 1000 \text{ km s}^{-1}$  only for nearly equal mass mergers ( $q > 0.3$ ). Such mergers are extremely rare for the heaviest MBHs (see Section 4; Volonteri 2007; Sesana, Volonteri & Haardt 2007). Furthermore, as mentioned above, further accretion can increase the degree of alignment of the binary system. As a consequence, in this study we may neglect the dependence of the spin-orbit configurations on the MBH masses and adopt the same distributions of relative angles between the two MBH spins for all MBH mergers in our sample. Note, also, that the isotropic distribution case we test below provides an upper limit to the ejection probability at all masses.

### 3 EJECTIONS AND HALO BIAS

In models of structure formation based on gravitational instabilities in Gaussian primordial fluctuations, the number density and bias properties of a halo can be expressed as a function of its rms fluctuation of the linear density field at redshift  $z$ ,  $\sigma(M_h, z)$ , and on the threshold density for collapse of a homogeneous spherical perturbation at redshift  $z$ ,  $\delta_c(z)$ .

In the Press & Schechter (1974) formalism the comoving number density of haloes of mass between  $M$  and  $M + dM$  can be expressed as

$$\frac{dn}{dM} = \sqrt{\frac{2}{\pi}} \frac{\rho_m}{M} \frac{-d(\ln \sigma)}{dM} \nu_c e^{-\nu_c^2/2}, \quad (12)$$

where  $\nu_c = \delta_c(z)/\sigma(M, z)$  is the number of standard deviations that the critical collapse over-density represents on mass scale  $M$ . At any redshift we can identify the characteristic mass (i.e.  $\nu_c = 1$ ), and its multiples. The higher the  $\nu_c$ , the more massive and rarer the halo, and the higher its bias and clustering strength. Although this formalism is derived from a Press–Schechter analysis, it agrees fairly well with the results of  $N$ -body simulations.

Martini & Weinberg (2001) suggested that clustering analysis of quasar samples can be deconvolved to yield the typical mass of haloes hosting quasars, and their typical  $\nu_c$ . Martini & Weinberg (2001) and subsequent investigations (Porciani, Magliocchetti & Norberg 2004; Shen et al. 2007; Myers et al. 2007) have found that high-redshift quasars are highly biased objects with respect to the underlying matter, and that their  $\nu_c$  increases with  $z$ . Shen et al. (2007) and Myers et al. (2007) find that the bias increases from  $\nu_c \simeq 3$  at  $z = 2$  to  $\nu_c \simeq 3.5$  at  $z = 3$  and  $\nu_c \simeq 5.5$  at  $z = 5$ .

We evaluate the ejection probabilities, using the techniques described in Section 2 with fit CL, for haloes representing peaks of the density fluctuations  $\nu_c = 1, 2, 3, 4, 5, 6$  as a function of redshift in a concordance  $\Lambda$ CDM cosmology (Spergel et al. 2007). For every

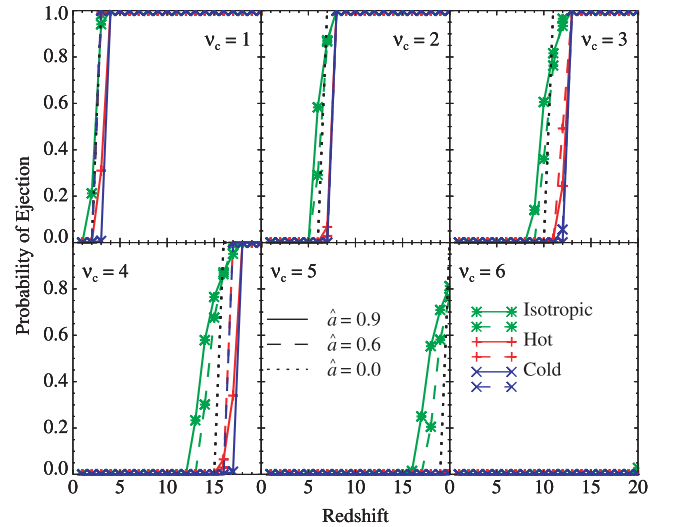
halo mass we estimate the ejection probability by comparing the recoil velocity ( $v_{\text{recoil}}$ ) to the escape velocity from the dark matter halo potential well (truncated at the virial radius). We model the halo potential with a Navarro, Frenk & White (1997) density profile, where the halo properties evolve as suggested by Bullock et al. (2001).

The ejection probabilities thus defined are a lower limit to the probability that the recoil voids a galaxy of its central MBH, as if the recoil velocity is lower than the escape velocity but higher than the velocity dispersion of a halo, the time-scale for the recoiled MBH to return to the centre under the effect of dynamical friction is likely longer than the Hubble time (Madau & Quataert 2004; Gualandris & Merritt 2008; Volonteri & Madau 2008; Devecchi et al. 2009; Guedes et al. 2009, and references therein).

With this caveat in mind we can estimate lower limits to the probability that MBH binaries are ejected or displaced due to the gravitational recoil in  $\nu_c$ -peaks haloes (Fig. 2).

The redshift at which the probability drops to 50 per cent increases with increasing  $\nu_c$  for all binary mass ratios and spins. The probability of ejection for MBH binaries in haloes that host very high redshift quasars ( $\nu_c \simeq 5.5$  at  $z = 5$ ) drops to 0 at  $z \gtrsim 19$  if MBHs are non-spinning, while for spinning MBHs the ejection probability is significant down to  $z \gtrsim 13$  in the absence of any alignment process. Lower  $\nu_c$  peaks remain dangerous for merging binaries for longer. A  $\nu_c \simeq 3$  peak has an ejection probability of 100 per cent until  $z = 13$ . If *all*  $\nu_c \simeq 3$  haloes host an MBH, and *all* experience a major merger before  $z = 13$ , then no MBHs are left for further evolution. In this scenario, MBHs must form at  $z < 11$  in order to evolve all the way to  $z = 3$  and host the observable quasars.

However, the  $\nu_c$ -peak formalism does not allow a clear evaluation of the merger history of haloes. A halo that represents a  $\nu_c$ -peak at a given redshift  $z$ , would be incorporated into a halo representing a lower peak at  $z - \Delta z$ . An MBH in a given halo would then belong to different  $\nu_c$ -peaks in its lifetime. In order to evaluate the number



**Figure 2.** Ejection probability as a function of redshift for different  $\nu_c = \delta_c(z)/\sigma(M, z)$  peaks of the density fluctuations field. Line styles indicate the assumed spin of the merging black holes: Solid lines:  $\hat{a} = 0.9$ ; dashed lines:  $\hat{a} = 0.6$ ; dotted lines:  $\hat{a} = 0.0$ . Line colours indicate assumed distribution of spin orientations: green: isotropically distributed; red: orientation distribution from hot simulations by Dotti et al. (2010); blue: orientation distribution from cold simulations by Dotti et al. (2010). All probabilities assume  $q = 0.1$ .

<sup>1</sup> We took the cold, prograde case as our fiducial case, but the results are not very different for the other cases.

of MBH mergers that a halo experienced, it is necessary to use techniques that can trace the whole merger history of a given halo as a function of mass and time.

#### 4 INDIVIDUAL HALO HISTORIES

We now turn to evaluate the merger histories of different galaxy haloes. We adopt a statistical approach, based on merger histories extracted from full  $\Lambda$ CDM merger trees. This work has a more general approach than Volonteri et al. (2008b), and addresses in an astrophysical context the analysis by Schnittman (2007). While Volonteri et al. (2008b) looked at the role of ejections in the merger histories for galaxies that were in clusters but not in the ‘main trunk’, here we examine merger histories with a focus on the main halo of the merger tree. We evaluate the average ejection probability as a function of cosmic epoch, and consider how it depends on the environmental bias. We consider here the halo merger histories leading to the formation of haloes with masses  $M_0 = 4 \times 10^{13} M_\odot$ ,  $M_0 = 2 \times 10^{12} M_\odot$ ,  $M_0 = 2 \times 10^{11} M_\odot$  at  $z = 0$ . We average our results over 20 realizations of the same mass, to account for cosmic variance. Our technique and cosmological framework is similar to the one described in Volonteri, Haardt & Madau (2003a). We track the dynamical evolution of MBHs ab initio and follow their assembly down to  $z = 0$ .

Several theoretical arguments indicate that MBH formation occurs at very high redshift, and probably in biased haloes (e.g. Madau & Rees 2001; Volonteri & Rees 2006). Additional parameters, such as the angular momentum of the gas, its ability to cool and its metal enrichment, are likely to set the exact efficiency of MBH formation and the redshift range when the mechanism operates (for a thorough discussion, see Volonteri, Lodato & Natarajan 2008a). We consider the effect of varying the initial conditions, by assuming frequent or rare MBH formation process. As a reference model, we assume that MBH formation is effective in all haloes with  $v_c > 3$  at  $z > 20$  (corresponding to masses  $> 10^6 M_\odot$ ). This model is based on a scenario where MBHs form as remnants of the first generation of metal-free stars (Population III). The main features of the hierarchical assembly of MBHs left over by the first stars in a  $\Lambda$ CDM cosmology have been discussed by Volonteri et al. (2003a), Volonteri & Rees (2006) and Volonteri & Natarajan (2009). A model similar to the one presented in this paper has been shown to reproduce observational constraints on MBH evolution (luminosity function of quasars and Soltan’s argument,  $M_{\text{BH}}-\sigma$  relationship at  $z = 0$ , mass density in MBHs at  $z = 0$ ; Volonteri et al. 2008a). We also analysed a very different case, where whenever a halo grows above a given mass threshold ( $> 10^{10} M_\odot$ ) and it does not already contain an MBH in its centre, then it forms one (in case the previous MBH had been displaced or ejected, a new MBH materializes), regardless of redshift. This model is not based on a specific physical model, but it is meant to be used for comparison with existing numerical simulations. MBH formation mechanisms define when and how often haloes are populated with MBHs. They provide the initial occupation fraction.

Additionally, we have to follow the dynamical evolution of haloes and embedded MBHs all the way to  $z = 0$  in order to determine the effect of recoils on the occupation fraction of MBHs. Since the magnitude of the recoil depends on the mass ratio of the merging MBHs, we have to model the mass-growth of MBHs and the merger efficiency. We base our model of MBH growth on the empirical correlation found between MBH masses and the properties of their hosts, and on the suggestion that these correlations are established during galaxy mergers that fuel MBH accretion and form bulges.

We therefore assume that after every merger between two galaxies with a mass ratio larger than 1:10, their MBHs climb to the same relation with the velocity dispersion of the halo, as it is seen today (Gültekin et al. 2009):

$$M = 1.3 \times 10^8 M_\odot \left( \frac{\sigma}{200 \text{ km s}^{-1}} \right)^{4.24}. \quad (13)$$

We link the correlation between the black hole mass and the central stellar velocity dispersion of the host with the empirical correlation between the central stellar velocity dispersion and the asymptotic circular velocity ( $V_c$ ) of galaxies (Ferrarese 2002; see also Pizzella et al. 2005; Baes et al. 2003):

$$\sigma = 200 \text{ km s}^{-1} \left( \frac{V_c}{304 \text{ km s}^{-1}} \right)^{1.19}. \quad (14)$$

The latter is a measure of the total mass of the dark matter halo of the host galaxies. A halo of mass  $M_h$  collapsing at redshift  $z$  has a circular velocity

$$V_c = 142 \text{ km s}^{-1} \left[ \frac{M_h}{10^{12} M_\odot} \right]^{1/3} \left[ \frac{\Omega_m}{\Omega_m^z} \frac{\Delta_c}{18\pi^2} \right]^{1/6} (1+z)^{1/2}, \quad (15)$$

where  $\Delta_c$  is the over-density at virialization relative to the critical density. For a *WMAP5* cosmology, we adopt here the fitting formula (Bryan & Norman 1998)  $\Delta_c = 18\pi^2 + 82d - 39d^2$ , where  $d \equiv \Omega_m^z - 1$  is evaluated at the collapse redshift so that  $\Omega_m^z = \Omega_m(1+z)^3 / (\Omega_m(1+z)^3 + \Omega_\Lambda + \Omega_k(1+z)^2)$ .

Therefore, if a dark matter halo of mass  $M_h$  at redshift  $z$  hosts an MBH, we can derive the MBH mass via  $V_c$  and  $\sigma$  (see also Volonteri et al. 2003a; Rhook & Wyithe 2005; Croton 2009). The relationship between black hole and dark matter halo mass is

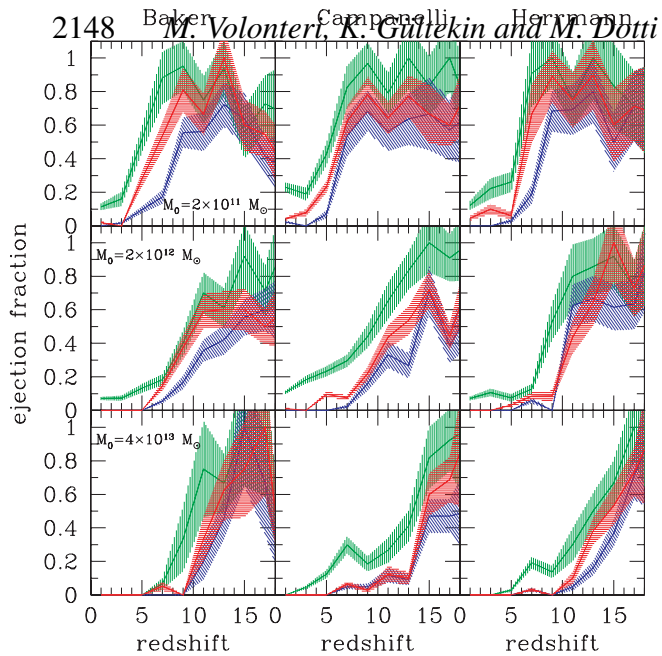
$$M = 1.3 \times 10^8 M_\odot \left[ \frac{M_h}{9.410^{13} M_\odot} \right]^{5/3} \left[ \frac{\Omega_m}{\Omega_m^z} \frac{\Delta_c}{18\pi^2} \right]^{5/6} (1+z)^{5/2}, \quad (16)$$

in good agreement with the recent result by Bandara, Crampton & Simard (2009).

We further assume that MBHs merge within the merger time-scale of their host haloes ( $t_{\text{merge}}$ ), which is a likely assumption for MBH binaries formed after gas-rich galaxy mergers (Dotti et al. 2007, and references therein). We adopt the relations suggested by Taffoni et al. (2003) for the orbital decay of merging haloes. We treat as MBH mergers at a given  $z$  those MBH binaries that are expected to merge during that specific time-step. If two haloes start interacting at  $z_{\text{in}}$ , corresponding to a Hubble time  $t_H(z_{\text{in}})$  with a merger time-scale  $t_{\text{merge}}$ , then we consider these MBHs merged at  $z_{\text{fin}}$  corresponding to  $t_H(z_{\text{fin}}) = t_H(z_{\text{in}}) + t_{\text{merge}}$ . Although we follow the dynamical evolution of each MBH along cosmic time, we note that dynamical friction appears to be efficient for mergers with mass ratio of the progenitors larger than 1:10. Small satellites suffer severe mass losses by the tidal perturbations induced by the gravitational field of the primary halo. This progressive mass loss increases the decay time that can be of order the Hubble time when the mass ratio is smaller than 1:10.

Finally, the spin of MBHs is fixed to be  $\hat{a} = 0.9$ , in order to obtain upper limits to the recoil consequences.

This set of models, though simple, provides a wide range of histories that can be considered to bracket some extreme behaviours. We find that the environment of the MBH population plays an important role. Along the merger history of massive galaxies, the fraction of ejected MBHs decreases more rapidly with decreasing



**Figure 3.** Probability of MBH ejections as a function of redshift along the merger history of the main halo in each merger tree. Each row corresponds to different halo masses  $M_0$  at  $z = 0$  as indicated in the left panels. Each column as indicated corresponds to the different fitting function used to calculate recoil speed. The *green, vertically hatched* regions are for an isotropic distribution of spin orientations; the *red, horizontally hatched* regions are for the ‘hot’ simulations by Dotti et al. (2010); and the *blue, diagonally hatched* regions are for the ‘cold’ simulations by Dotti et al. (2010). In every panel the MBH spins are fixed at  $\hat{a} = 0.9$ . The trends and results are in strong qualitative agreement with the results from the simpler simulations presented in Fig. 2. There is also general agreement among the different fitting functions.

redshift, dropping below 50 per cent by  $z \sim 7$  for  $M_0 = 4 \times 10^{13} M_\odot$ . The 50 per cent threshold is reached at later times in the ‘trees’ of less massive haloes:  $z \sim 5$  for  $M_0 = 2 \times 10^{12} M_\odot$  and  $z \sim 2$  for  $M_0 = 2 \times 10^{11} M_\odot$ .

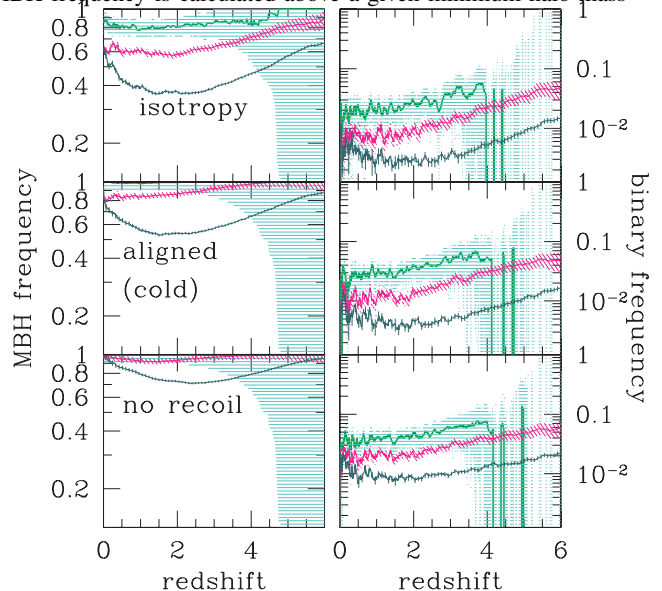
It is particularly instructive to examine the ejection history of the tree’s main halo, which is the galaxy we would see today (Fig. 3). The main halo is usually among the most massive haloes in the tree at any time, implying the largest escape velocities. The fraction of ejected MBHs is therefore lower. The central MBH in a large galaxy has a very small probability of being lost after  $z = 5$ . Since most of the growth of MBHs happens between  $z = 3$  and  $z = 1$ , corresponding to the peak in quasar activity, we argue that, if accretion is responsible for creating the correlations between MBHs and their hosts (Silk & Rees 1998; Fabian 1999; King 2003), then MBHs hosted in large galaxies will sit close to the expected correlation. The central MBH in a small galaxy (e.g.  $M_0 = 2 \times 10^{11} M_\odot$ ) has instead a large probability ( $\sim 20$  per cent) of being ejected all the way to today. If such a low-redshift ejection happens, it is not immediately clear if the galaxy can re-acquire an MBH and if the latter can grow to the hypothetical mass that the correlation with the host would suggest.

The alternative case of MBH formation (haloes with  $> 10^{10} M_\odot$ ) gives quantitatively similar results for large galaxies, e.g.  $M_0 = 4 \times 10^{13} M_\odot$ , and qualitatively similar results for small galaxies. Black holes form much later in this model (as the emergence of more massive galaxies must wait until  $z \simeq 5-7$ ), and this pushes the ejection fraction, causing it to first rise steeply, as more and more

MBHs are available to merge, and then decrease as galaxies grow in mass deepening the potential wells.

It is most interesting that the cosmic ejection rate is very similar for fit B compared to fits CL and H, notwithstanding the different  $\eta$  dependence ( $\eta^3$  versus  $\eta^2$ ). The reason is simple. From Fig. 1 it is evident that the mass-ratio dependence picks up at  $q < 0.1$  (where, incidentally, recoil velocities become much smaller than escape velocities from galaxies). However, mergers with  $q \ll 1$  are rather uncommon (Volonteri et al. 2003a; Sesana et al. 2005), due to dynamical effects. During a galactic merger, it is dynamical friction that drags in the satellite, along with its central MBH, towards the centre of the more massive progenitor. When the orbital decay is efficient, the satellite hole moves towards the centre of the more massive progenitor, leading to the formation of a bound MBH binary. The efficiency of dynamical friction decreases with the mass ratio of the merging galaxies: only nearly equal mass galaxy mergers (‘major mergers’, mass ratio larger than  $\simeq 1:10$ ) lead to efficient MBH binary formation within time-scales shorter than the Hubble time (Taffoni et al. 2003). These effects must be convolved with the mass-ratio probability distribution. As the mass function of haloes (and galaxies) is steep, the probability of halo mergers decreases with increasing mass ratio. That is, dynamically efficient major mergers are rare, and minor mergers are common but inefficient at forming MBH binaries.

We can now derive how strong or mild recoils influence the frequency of MBHs in galaxies. We analyse here only the  $v_c > 3$  case, as by construction the alternative case (mass threshold  $> 10^{10} M_\odot$ ) has an occupation fraction of unity for all haloes above the threshold, since if an MBH is lost, a new immediately materializes. We first derive a control case, in which we ignore recoils altogether. This control case is shown in the bottom panel of Fig. 4. The MBH frequency is calculated above a given minimum halo mass



**Figure 4.** Left: frequency of MBHs in galaxies as a function of redshift for haloes above three different mass thresholds:  $M > 10^{10} M_\odot$  (grey, vertically hatched)  $M > 10^{11} M_\odot$  (magenta, diagonally hatched)  $M > 10^{12} M_\odot$  (cyan, horizontally hatched) for isotropic spin orientations (top) and spin orientations from the ‘cold’ simulations of Dotti et al. (2010) (middle). The bottom panel shows the black hole occupation fraction when the recoil is arbitrarily set to zero at each MBH merger. Right: binary MBH frequency for the same models and mass thresholds.

( $>10^{10} M_{\odot}$ ;  $>10^{11} M_{\odot}$ ;  $>10^{12} M_{\odot}$ ). As discussed by Menou, Haiman & Narayanan (2001) and Volonteri et al. (2003a), the frequency of MBHs in haloes above a fixed mass threshold initially decreases with cosmic time as lower mass haloes lacking MBHs become more massive than the assumed threshold. Eventually, the occupation fraction starts to increase as the total number of individual haloes drops.

Given that different fitting formulae give very similar results, we show here the results that we obtain using fit CL. The worst-case scenario, the case with the largest number of ejections, is the isotropic case, which leads to the largest recoils (see Fig. 1). The top panel of Fig. 4 indicates that recoils can decrease the overall frequency of MBHs in small galaxies to  $\sim 60$  per cent, while they have little effect on the frequency of MBHs in large galaxies (at most a 20 per cent effect). The middle panel depicts the cold aligned case. We consider the latter the more realistic situation for gas-rich mergers, which are expected for high-redshift galaxies, at least. It is reassuring to notice that recoils are not dangerous for most galaxies.

This is the result of a combination of effects: on the one hand, MBHs hosted in small haloes, with shallow potential wells and low escape velocities, have a high ejection probability (Fig. 3); on the other hand, the MBH merger rate is very low along their galaxy formation merger hierarchy: MBH formation processes are inefficient in such shallow potential wells, and the anti-hierarchical nature of the galaxy assembly implies that not much action happens in low-bias systems at high redshift. This is exemplified in the right panels of Fig. 4, in which we show the frequency of MBH close binaries for the same mass thresholds we used to calculate the MBH frequency (note the different y-axis scales). Naïvely, one would expect the frequency of double MBHs to scale as the square of the MBH frequency, but the frequency of close binaries is suppressed with respect to the frequency of pairs because of the requirement of efficient orbital decay. Our simulations show that gravitational recoil is not expected to be efficient at ejecting MBHs with mass  $\simeq 10^6 M_{\odot}$ . This mass range is where the *Laser Interferometric Space Antenna* (LISA) gravitational wave observatory will be most sensitive to extreme mass ratio inspiral (EMRI) events (Gair 2009). Thus the event rate for EMRIs is not likely to be affected by recoils.

Summarizing, the larger ejection probability that MBHs hosted in low-bias haloes have is counteracted by their lower merger probability, thus leading to an overall small change in the MBH frequency as a function of the recoil strength.

## 5 DISCUSSION

Ejection of MBHs may explain the unusual case of NGC 3115. NGC 3115 is an S0 galaxy with MBH measured to have mass  $M = 1 \times 10^9 M_{\odot}$  (Kormendy & Richstone 1992), yet the surface-brightness profile of the central bulge has a small break radius  $r_b \sim 2$  pc with a steep inner slope consistent with a power law  $\gamma = 0.52$  (Lauer et al. 2007b). This presents a contradiction of theoretical expectations for a galaxy of this size ( $M_h \approx 10^{13} M_{\odot}$ ).

A galaxy as massive as NGC 3115 is expected to have merged frequently enough to have had a binary black hole at some point in its history (Volonteri, Madau & Haardt 2003b). According to the standard picture of core scouring (Begelman et al. 1980; Faber et al. 1997; Milosavljević & Merritt 2003; Lauer et al. 2007b), the binary black hole would eject stars on elongated orbits as orbital energy is transferred from the binary to kinetic energy in the stars. This process shrinks the binary separation and flattens the stellar number density profile so that the surface-brightness profile appears as a core at the centre. The galaxy's core can be described by its

deficit in stellar mass, i.e. the mass in stars ejected from what was previously a power-law profile. The most recent numerical simulations on black hole mergers find that for nearly equal masses, the mass deficit should scale with the total mass of the binary (Merritt 2006); but for mass ratios far from unity, the mass deficit should scale with the mass of the secondary (Sesana, Haardt & Madau 2008). Observationally, the mass deficit is found to scale linearly with existing  $M_{\text{BH}}$  (Lauer et al. 2007a; Kormendy & Bender 2009). Regardless of its magnitude, the mass deficit is expected to persist after the binary coalesces because the remnant black hole acts as a ‘guardian’ that prevents dynamical refilling of the core (Lauer et al. 2007a).

The observations of NGC 3115, however, reveal that this is not what happened. There is an existing MBH at its centre that would have acted as the core's guardian, but the surface-brightness profiles show no core. There are several possible explanations. First, core scouring may be greatly diminished if the final major merger had a significant amount of gas (Kormendy et al. 2009). The gas drag speeds up the process of MBH merger before the MBH binary can eject many stars. Even if core scouring occurred, a subsequent gas-rich merger could refill the core by depositing gas that would form stars. Here we suggest another possibility. The final MBH binary is ejected by gravitational recoil (or three-body encounters, Gültekin, Miller & Hamilton 2004, 2006; Hoffman & Loeb 2007), but the MBH is replaced in a subsequent merger. Taking into consideration the typical mass ratios of merging MBHs as a function of galaxy bias and cosmic time, the probability that an MBH in NGC 3115 has been ejected at redshift  $z < 5$  is 10 per cent in the case of a ‘dry merger’, but less than 1 per cent in the case of a ‘wet merger’. This is because at late cosmic times the binary mass ratio distribution becomes shallow, with  $q \ll 1$  becoming less probable (Volonteri 2007). In general, in order to have a probability larger than 10 per cent that an MBH recoils with a velocity larger than the escape velocity of the progenitor of NGC 3115 at  $z \simeq 5$ , the mass ratio of the merging MBHs must be  $q > 0.3$  for the isotropic and hot cases (assuming spins  $\hat{a} = 0.9$ ). For the cold case, even for  $q = 1$  and  $\hat{a} = 0.9$ , the probability is 0.8 per cent. In the ejection scenario, NGC 3115 should have had at least one major merger that involved two MBH-hosting galaxies where substantial star formation or AGN feedback kept the gas pressurized (corresponding to our hot simulations).

## 6 SUMMARY

We assessed the effects of spin alignment on the strength of the gravitational recoil along the cosmic build-up of galaxies and MBHs. We determined ejection probabilities for mergers in gas-poor galaxies, where the MBH binary coalescence is driven by stellar dynamical processes, and the spin-orbit configuration is expected to be isotropically distributed. We contrast this case of gas-rich mergers, where we expect MBH spins to align with the orbital angular momentum. This is because in gas-rich environments MBHs accrete gas, which exerts gravito-magnetic torques that align the spins of both the MBHs with the angular momentum of the large-scale gas flow. We find that for aligned configurations the ejection probability is strongly suppressed (by at least a factor of 2). Along the merger history of a large elliptical, the ejection probability becomes negligible at  $z < 5$ , while small galaxies have ejection probabilities of order 20 per cent even today. However, the MBH merger rate is very low along their merger hierarchy of small galaxies: MBH formation processes are likely inefficient in such shallow potential wells. The occupation fraction of MBHs, intimately related to halo bias and MBH formation efficiency, therefore plays a crucial role in

increasing the retention fraction. Recoils can effectively decrease the overall frequency of MBHs in small galaxies to  $\sim 60$  per cent, while they have little effect on the frequency of MBHs in large galaxies (at most a 20 per cent effect).

## ACKNOWLEDGMENTS

MV acknowledges support from a Rackham faculty grant.

## REFERENCES

- Baes M., Buyle P., Hau G. K. T., Dejonghe H., 2003, MNRAS, 341, L44  
 Baker J. G., Centrella J., Choi D.-I., Koppitz M., van Meter J. R., Miller M. C., 2006, ApJ, 653, L93  
 Baker J. G., Boggs W. D., Centrella J., Kelly B. J., McWilliams S. T., Miller M. C., van Meter J. R., 2007, ApJ, 668, 1140  
 Baker J. G., Boggs W. D., Centrella J., Kelly B. J., McWilliams S. T., Miller M. C., van Meter J. R., 2008, ApJ, 682, L29  
 Bandara K., Crampton D., Simard L., 2009, ApJ, 704, 1135  
 Begelman M. C., Blandford R. D., Rees M. J., 1980, Nat, 287, 307  
 Bogdanović T., Reynolds C. S., Miller M. C., 2007, ApJ, 661, L147  
 Bryan G. L., Norman M. L., 1998, ApJ, 495, 80  
 Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001, MNRAS, 321, 559  
 Campanelli M., Lousto C. O., Zlochower Y., Merritt D., 2007, Phys. Rev. Lett., 98, 231102  
 Croton D. J., 2009, MNRAS, 394, 1109  
 Cuadra J., Armitage P. J., Alexander R. D., Begelman M. C., 2009, MNRAS, 393, 1423  
 Devecchi B., Rasia E., Dotti M., Volonteri M., Colpi M., 2009, MNRAS, 394, 633  
 Dotti M., Colpi M., Haardt F., Mayer L., 2007, MNRAS, 379, 956  
 Dotti M., Ruzsowski M., Paredi L., Colpi M., Volonteri M., Haardt F., 2009, MNRAS, 396, 1640  
 Dotti M., Volonteri M., Perego A., Colpi M., Ruzsowski M., Haardt F., 2010, MNRAS, 402, 682  
 Faber S. M. et al., 1997, AJ, 114, 1771  
 Fabian A. C., 1999, MNRAS, 308, L39  
 Ferrarese L., 2002, ApJ, 578, 90  
 Ferrarese L., Merritt D., 2000, ApJ, 539, L9  
 Gair J. R., 2009, Classical Quant. Gravity, 26, 094034  
 Gebhardt K. et al., 2000, ApJ, 539, L13  
 Gonzalez J. A., Hannam M. D., Sperhake U., Bruggmann B., Husa S., 2007, Phys. Rev. Lett., 98, 231101  
 Gualandris A., Merritt D., 2008, ApJ, 678, 780  
 Guedes J., Madau P., Kuhlen M., Diemand J., Zemp M., 2009, ApJ, 702, 890  
 Gültekin K., Miller M. C., Hamilton D. P., 2004, ApJ, 616, 221  
 Gültekin K., Miller M. C., Hamilton D. P., 2006, ApJ, 640, 156  
 Gültekin K. et al., 2009, ApJ, 698, 198  
 Hayasaki K., Mineshige S., Sudou H., 2007, PASJ, 59, 427  
 Herrmann F., Hinder I., Shoemaker D., Laguna P., Matzner R. A., 2007, Phys. Rev. D, 76, 084032  
 Hoffman L., Loeb A., 2007, MNRAS, 377, 957  
 King A., 2003, ApJ, 596, L27  
 Koppitz M., Pollney D., Reisswig C., Rezzolla L., Thornburg J., Diener P., Schnetter E., 2007, Phys. Rev. Lett., 99, 041102  
 Kormendy J., Bender R., 2009, ApJ, 691, L142  
 Kormendy J., Richstone D., 1992, ApJ, 393, 559  
 Kormendy J., Richstone D., 1995, ARA&A, 33, 581  
 Kormendy J., Fisher D. B., Cornell M. E., Bender R., 2009, ApJS, 182, 216  
 Lauer T. R. et al., 2007a, ApJ, 662, 808  
 Lauer T. R. et al., 2007b, ApJ, 664, 226  
 Lousto C. O., Zlochower Y., 2009, Phys. Rev. D, 79, 064018  
 Madau P., Quataert E., 2004, ApJ, 606, L17  
 Madau P., Rees M. J., 2001, ApJ, 551, L27  
 Magorrian J. et al., 1998, AJ, 115, 2285

- Martini P., Weinberg D. H., 2001, ApJ, 547, 12  
 Mayer L., Kazantzidis S., Madau P., Colpi M., Quinn T., Wadsley J., 2007, Sci, 316, 1874  
 Menou K., Haiman Z., Narayanan V. K., 2001, ApJ, 558, 535  
 Merritt D., 2006, ApJ, 648, 976  
 Milosavljević M., Merritt D., 2003, ApJ, 596, 860  
 Myers A. D., Brunner R. J., Nichol R. C., Richards G. T., Schneider D. P., Bahcall N. A., 2007, ApJ, 658, 85  
 Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493  
 Perego A., Dotti M., Colpi M., Volonteri M., 2009, MNRAS, 399, 2249  
 Pizzella A., Corsini E. M., Dalla Bontà E., Sarzi M., Coccatto L., Bertola F., 2005, ApJ, 631, 785  
 Porciani C., Magliocchetti M., Norberg P., 2004, MNRAS, 355, 1010  
 Press W. H., Schechter P., 1974, ApJ, 187, 425  
 Rhook K. J., Wyithe J. S. B., 2005, MNRAS, 361, 1145  
 Richstone D. et al., 1998, Nat, 395, A14  
 Schnittman J. D., 2007, ApJ, 667, L133  
 Schnittman J. D., Buonanno A., 2007, ApJ, 662, L63  
 Sesana A., Haardt F., Madau P., Volonteri M., 2005, ApJ, 623, 23  
 Sesana A., Volonteri M., Haardt F., 2007, MNRAS, 377, 1711  
 Sesana A., Haardt F., Madau P., 2008, ApJ, 686, 432  
 Shen Y. et al., 2007, AJ, 133, 2222  
 Silk J., Rees M. J., 1998, A&A, 331, L1  
 Spaans M., Silk J., 2000, ApJ, 538, 115  
 Spergel D. N. et al., 2007, ApJS, 170, 377  
 Taffoni G., Mayer L., Colpi M., Governato F., 2003, MNRAS, 341, 434  
 Tremaine S. et al., 2002, ApJ, 574, 740  
 Volonteri M., 2007, ApJ, 663, L5  
 Volonteri M., Madau P., 2008, ApJ, 687, L57  
 Volonteri M., Natarajan P., 2009, MNRAS, 400, 1911  
 Volonteri M., Rees M. J., 2006, ApJ, 650, 669  
 Volonteri M., Haardt F., Madau P., 2003a, ApJ, 582, 559  
 Volonteri M., Madau P., Haardt F., 2003b, ApJ, 593, 661  
 Volonteri M., Lodato G., Natarajan P., 2008a, MNRAS, 383, 1079  
 Volonteri M., Haardt F., Gültekin K., 2008b, MNRAS, 384, 1387

## APPENDIX A: DEPENDENCE OF THE RESULTS ON THE MBH-HOST RELATIONSHIPS

The estimates presented in the main text assume MBH masses correlate with the velocity dispersion of their host halo (Bandara et al. 2009), and also that this relationship is redshift-independent.

To test the robustness of our results against the assumed MBH-host relationship, we implement a very different relationship, that is, we assume that the MBH mass scales with the mass of the dark matter halo:  $M = 10^{-6} M_h$ . This scaling leads to different dependencies of the MBH mass ratio,  $q$ , with the properties of the merging hosts (redshift, halo mass). The redshift dependence, however, cancels out in the expression for the MBH mass ratio,  $q$ , as long as the MBHs grow almost coevally. More important is the different scaling of  $q$  with the properties of the merging galaxies. In the ‘sigma-based’ relationship, we have  $q \propto (M_{h,2}/M_{h,1})^{5/3}$ , while for the ‘mass-based’ relationship,  $q \propto (M_{h,2}/M_{h,1})$ . For  $M_{h,2}/M_{h,1} < 1$ , the sigma-based scaling leads to lower  $q$  values (this is true for any scaling relationship with an exponent larger than 1), and milder recoils.

We find that the occupation fraction of small galaxies is mostly unaffected by the choice of the MBH-host relationship, and the difference for larger galaxies is within 20 per cent, with the linear mass scaling leading to a lower occupation fraction. The redshift dependence of the MBH frequency is almost identical.

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.