From galactic nuclei to the halo outskirts: tracing supermassive black holes across cosmic history and environments

David Izquierdo-Villalba,1⋆ Silvia Bonoli,2,3 Massimo Dotti,4,5 Alberto Sesana,4
Yetli Rosas-Guevara,2 Daniele Spinoso1
1 Centro de Estudios de Física del Cosmos de Aragón (CEFCA), Plaza San Juan 1, Planta-2, Teruel, 44001, Spain.
2 Donostia International Physics Centre (DIPC), Paseo Manuel de Lardizábal 4, 20018 Donostia-San Sebastián, Spain
3 IKERBASQUE, Basque Foundation for Science, E-48013, Bilbao, Spain
4 Dipartimento di Fisica “G. Occhialini”, Università degli Studi di Milano-Bicocca, Piazza della Scienze 3, 20126 Milano, Italy
5 INFN, Sezione Milano-Bicocca, Piazza della Scienze 3, 1-20126 Milano, Italy

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

We study the mass assembly and spin evolution of supermassive black holes (BHs) across cosmic time as well as the impact of gravitational recoil on the population of nuclear and wandering black holes (wBHs) by using the semi-analytical model L-Galaxies run on top of Millennium merger trees. We track spin changes that BHs experience during both coalescence events and gas accretion phases. For the latter, we assume that spin changes are coupled with the bulge assembly. This assumption leads to predictions for the median spin values of \(z=0\) BHs that depend on whether they are hosted by pseudobulges, classical bulges or ellipticals, being \(a \sim 0.9, 0.7\) and \(0.4\), respectively. The outcomes of the model display a good consistency with \(z \leq 4\) quasar luminosity functions and the \(z=0\) BH mass function, spin values and black hole-bulge mass correlation. Regarding the wBHs, we assume that they can originate from both the disruption of satellite galaxies (orphan wBH) and ejections due to gravitational recoils (ejected wBH). The model points to a number density of wBHs that increases with decreasing redshift, although this population is always \(\sim 2\) dex smaller than the one of nuclear BHs. At all redshifts, wBHs are typically hosted in \(M_{\text{halo}} \gtrsim 10^{13} M_{\odot}\) and \(M_{\text{stellar}} \gtrsim 10^{10} M_{\odot}\), being orphan wBHs the dominant type. Besides, independently of redshift and halo mass, ejected wBHs inhabit the central regions \((\lesssim 0.1 R_{200})\) of the host DM halo, while orphan wBHs linger at larger scales \((\gtrsim 0.3 R_{200})\). Finally, we find that gravitational recoils cause a progressive depletion of nuclear BHs with decreasing redshift and stellar mass. Moreover, ejection events lead to changes in the predicted local black hole-bulge relation, in particular for BHs in pseudobulges, for which the relation is flattened at \(M_{\text{bulge}} > 10^{10.2} M_{\odot}\) and the scatter increase up to \(\sim 3\) dex.

Key words: galaxies:theoretical – galaxies:black holes – galaxies:quasar – methods:numerical

1 INTRODUCTION

Since the discovery of the first quasar (Schmidt 1963) our understanding of supermassive black holes (BHs), their growth and relation with the host galaxies has substantially increased, although a complete and exhaustive picture on their formation, evolution and impact on the environment is still missing. A couple of decades ago, observational results confirmed both the existence of BHs at the center of most massive galaxies and the presence of a correlation between their masses and the properties of their hosts (Soltan 1982; Haehnelt & Rees 1993; Faber 1999; O’Dowd et al. 2002; Haring & Rix 2004; Kormendy & Ho 2013; Savorgnan et al. 2016). Their effect on shaping the star formation of their hosts has also been extensively studied (Silk & Rees 1998; Birzan et al. 2004; Diamond-Stanic & Rieke 2012; Mullaney et al. 2012a,b; Eisenreich et al. 2017). Observational studies are finally finding direct evidence of AGN-driven winds

⋆ E-mail: dizquierd@cefca.es
(Tombesi et al. 2014, 2015; Cresci et al. 2015; Bischetti et al. 2017).

From a theoretical point of view, it is well established that black holes grow via gas accretion and coalescence with other black holes (Marulli et al. 2008; Bonoli et al. 2016; Hirschmann et al. 2012, 2014; Fanidakis et al. 2011; Dubois et al. 2012). While the former seems to be the main contributor to the total mass-assembly budget (Soltan 1982), the latter appears to have just a sub-dominant importance (Small & Blandford 1992; Fanidakis et al. 2011; Dubois et al. 2014). Concerning gas accretion, several physical processes are believed to be responsible for driving gas towards the black hole surroundings. One of these processes is galaxy mergers, whose gravitational torques can modify the galaxy cold gas structure and trigger gas infows towards the nuclear region (Hernquist 1989; Barnes & Hernquist 1991; Mihos & Hernquist 1996; Di Matteo et al. 2005; Springel 2005; Li et al. 2007). Secular processes such as the formation of bar structures or the redistribution of gas after supernovae explosions may have similar effects (Shlosman et al. 1989; Hopkins et al. 2009; Hopkins & Quataert 2010; Fanali et al. 2015; Dubois et al. 2015; Du et al. 2017). However, no clear observational conclusions about these two latter feeding mechanisms have been reached up to date. After being transported towards the nuclear parts, the gas has to lose a significant amount of its angular momentum to reach closer orbits around the BH to be ultimately accreted by it. Even though this process is not trivial and probably does not have an unique answer, many works have pointed out that both local viscous stress and gravitational torques (such as bars-within-bars, spiral waves or large-scale magnetic stress) can be very effective in transporting gas close to the BH (≈0.01 – 0.1pc) before being transformed into stars (Shlosman & Begelman 1989; Balbus & Hawley 1998; Thompson et al. 2005).

As a result of gas accretion onto the BHs, a large amount of energy can be released prompting the birth of a quasar or an Active Galactic Nucleus (Soltan 1982; Hopkins et al. 2007; Ueda et al. 2014). The amount of energy released depends on the mass of the black hole, the geometry of the accretion disk and on the spin of the black hole (Shakura & Sunyaev 1973; Merloni & Heinz 2008). The dimensionless spin parameter, \( a = c J_{\text{BH}}/GM_{\text{BH}}^2 \lesssim 1 \) where \( c \) is the speed of light, \( G \) is the gravitation constant and \( J_{\text{BH}} \) is the black hole angular momentum. Its specific value plays an important role in the BH assembly and feedback. While highly-spinning BHs can convert up to \( \sim 40\% \) of the accreted matter into radiation, slowly spinning ones can only reach up to \( \sim 5\% \). This reflects into potentially very bright radiative emission and relatively slow mass-growth of the first ones, as opposed to the generally high growth-rates and low luminosity of the latter ones (Bardeen 1970; Novikov & Thorne 1973; Page & Thorne 1974; Blandford & Znajek 1977). How BHs acquire their spin is not fully understood yet. Up to date, all the proposed theories have assumed that a BH reaches its final spin after repeated gas accretion episodes and BHs coalescence that modify the initial rotation (Fanidakis et al. 2011; Volonteri et al. 2013; Berti & Volonteri 2008; Barausse 2012). On the gas accretion side, several theoretical scenarios have been proposed. For instance, King et al. (2005) suggested the chaotic scenario where the gas accretion always proceeds via consumption of self-gravitating accretion disks in uncorrelated directions. In this scenario, BHs tend to spin-down towards a typical final value of \( a \lesssim 0.2 \) (King et al. 2008). On the other side, Volonteri et al. (2007) showed that a prolonged gas accretion in the form of a constant and coherent gas consumption, efficiently spins-up the BHs up to \( a \sim 1 \). These two can be seen as extreme cases, and intermediate scenarios have been explored (see e.g. Dotti et al. 2013; Fiacconi et al. 2018; Bustamante & Springel 2019). Over the last years, some works have attempted to measure BHs spin by using X-ray spectroscopy (Brenneman & Reynolds 2006; Reynolds 2013). However, the challenges and biases in the observations make it difficult to draw solid conclusions. In the near future, with the upcoming experiments such as Laser Interferometer Space Antenna (LISA, Amaro-Seoane et al. 2017) or Advanced Telescope for High-ENergy Astrophysics (ATHENA, Nandra et al. 2013) it will be possible to reach robust observational conclusions ruling out many theoretical models of BH spin evolution.

Even though BH-BH coalescences seem to play a sub-dominant role in the evolution of both BH mass and spin, they can have an important effect on the co-evolution between black hole and galaxy (Redmount & Rees 1989). Because of conservation of linear momentum, the propagation of gravitational waves imparts a recoil (or kick) in the remnant BH (Bekenstein 1973; Baker et al. 2008; Lousto et al. 2012). Small kicks could simply lead to a small displacement of the remnant BH from the galaxy center whose observational counterpart would be an offset AGN (Madau & Quataert 2004; Loeb 2007; Blecha & Loeb 2008b; Guedes et al. 2011). At the other extreme, if the recoil velocity is larger than the escape velocity of its host galactic center, the black hole can be kicked out of the galaxy becoming a wandering black hole. This last case could be more probable at high-\( z \), where the galaxy potential wells are smaller and relatively small recoil kicks can effectively expel the BH from the galaxy and possibly even from the halo potentials. Such BH ejections might have some important consequences to be taken into account when drawing a complete black hole-galaxy co-evolution picture. For instance Blecha et al. (2011), by using hydro-dynamical simulations, found that the bulge velocity dispersion and the BH mass relation changes (a shift in the velocity dispersion axis) when the simulation is run with and without the inclusion of recoil velocities. This deviation is due to the fact that the ejection interrupts the BH growth while the galaxy continue its evolution (see similar results of Blecha & Loeb 2008a; Gerosa & Sesana 2015). In the same work, Blecha et al. (2011) discussed the halting of AGN feedback as a consequence of the BH ejection. The outcome is an extended period of active star formation in the galaxy which yields a dense and massive nuclear stellar cusps. This type of nuclear regions could be a target property which would point out galaxies which underwent a BH ejection. Indeed, from an observational point of view, some galaxies have been tagged as candidates of having undergone a BH recoil (see e.g. BCG in A2261 cluster Postman et al. (2012) or SDSS J1056+5516 Kalfountzou et al. (2017) and Komossa (2012) for a review). On top of ejection via gravitational recoil, the complex dynamics of
the two BHs before the final coalescence might cause the formation of a wandering black hole. If the two black holes spend enough time orbiting around each other (Begelman et al. 1980; Quinlan & Hernquist 1997) a third BH could reach their sphere of influence. As a consequence, the three BHs are likely to undergo a complex 3-body scattering interaction with the final outcome of the ejection of the less massive BH and the shrinking of the separation between the two remnants BHs (Volonteri et al. 2003; Volonteri & Perna 2005; Bonetti et al. 2018a). Even though this type of interactions do not look like to have a main role in the population of wandering black holes (see e.g. Volonteri et al. 2003; Volonteri & Perna 2005) they seem to play importance in speeding up the final BH-BH coalescence (Bonetti et al. 2018b).

Wandering black holes can also be a consequence of the disruption of a satellite galaxy during the infall onto a larger system. Environmental processes such as ram pressure or tidal interactions, in fact, can gradually strip away satellite galaxies before they merge with the central object (Moore et al. 1999; Springel et al. 2001). The BHs hosted by the disrupted satellite thus are incorporated in the halo of the central galaxy and contribute to the population of wandering BHs. Even though these type of wandering BHs are interesting because they can give an idea of the galaxy disruption rate, they have been mainly studied by recent hydro-dynamical simulations (Bellovary et al. 2010; Tremmel et al. 2018; Pfister et al. 2019). Indeed, the challenges associated with their detection make comparisons between models and observations difficult (Miki et al. 2014).

Given the observational limitations in detecting wandering black holes, more theoretical work is needed to constrain where and how these black holes can be found with current and upcoming instruments. Several works have attempted to determine the orbits and time spent by a BH in a wandering orbit. For instance, by using 3D hydrodynamical simulation Blecha et al. (2011) found that at a fixed ratio between recoil and escape velocity, the recoil trajectories have a substantial variation with the galaxy gas fraction. On the other hand, under an analytic perspective, Choksi et al. (2017) explored different parameters and contributions in the orbital damping of wandering BHs finding a longer wandering phase in lighter halo and galaxy systems. These authors found that in many circumstances the wandering phase may be larger than the Hubble time.

In this work we use the L-Galaxies semi-analytical model in the version presented in Izquierdo-Villalba et al. (2019) to study the statistics and environment of black hole growth and spin evolution and to explore the population of wandering black holes across cosmic time. We include both galaxy mergers and galaxy disk instabilities as physical processes that can lead to BH growth. For tracking spin evolution we follow the approach introduced by Sesana et al. (2014), where spin changes are coupled to the bulge assembly. We also introduce a time-delay between galaxy and BH mergers that depends on the BHs mass ratio and the galaxy gas fraction. Recoil velocities after BH coalescences are calculated on the basis of the progenitors properties while the trajectories of kicked BHs depend on the physical properties of the host halos and galaxies. We also follow the trajectory of orphan black holes from stripped satellites to complete the picture of wandering black holes.

The paper is organised as follows: in section 2 we describe the main characteristics of the semi-analytical model and the merger trees used in this work. In section 3 we describe all the physics included in the semi-analytical model to properly follow the BH mass growth, spin evolution and wandering phases. In section 4 we presents our main findings regarding nuclear black holes (i.e., black holes in the center of galaxies). In section 5 we explore the properties of wandering black holes across cosmic time. In section 6 we present the effects of gravitational recoils in both BH occupation fraction and bulge-black hole mass relation. Finally, in section 7 we summarize our main findings. A lambda cold dark matter (ΛCDM) cosmology with parameters Ω_m = 0.315, Ω_Λ = 0.685, Ω_b = 0.045, σ_8 = 0.9 and H_0 = 67.3 km s^{-1} Mpc^{-1} is adopted throughout the paper (Planck Collaboration et al. 2014).

2 GALAXY FORMATION MODEL

In this section we briefly describe the L-Galaxies semi-analytical model (SAM) and the dark matter simulations used in this work. The bulk of the model is the one presented in Henriques et al. (2015), with the modifications detailed in Izquierdo-Villalba et al. (2019) which contribute to a better description of galaxy morphologies and radii for both the Millennium and MillenniumII merger trees.

2.1 Merger trees

The version of L-Galaxies used here is built on top of the merger-trees of the Millennium (hereafter MS, Springel 2005) and Millennium II (hereafter MSII, Boylan-Kolchin et al. 2009) dark matter (DM) N-body simulations. The former follows the cosmological evolution of 2160^3 DM particles with a mass of m_p = 8.6 \times 10^8 M_{\odot}/h inside a periodic box of 500 Mpc/h on a side, from z = 127 to the present. The latter follows the same number of particles with a 125 times higher mass resolution (m_p = 6.885 \times 10^6 M_{\odot}/h) in a box of 100 Mpc/h on a side. Both simulations were originally run with WMAP1 & 2dFGRS concordance cosmology: Ω_m = 0.25, Ω_b = 0.045, Ω_Λ = 0.75, h = 0.73 km s^{-1} Mpc^{-1}, σ_8 = 0.9 (Colless et al. 2001). However, the latest L-Galaxies version was tuned on a re-scaled versions of MS and MSII simulations (Angulo & White 2010) to match the cosmological parameters provided by Planck first-year data (Planck Collaboration et al. 2014): Ω_m = 0.315, Ω_Λ = 0.685, Ω_b = 0.045, σ_8 = 0.9 and h = 0.673 km s^{-1} Mpc^{-1}.

Data from the MS and MSII simulations were stored respectively at 63 and 68 epochs or snapshots (where Δt~300 Myr). DM halos and subhalos were identified within the snapshots by using SUBFIND algorithm (Springel et al. 2001) and arranged in merger trees structures with L-HALOTREE (Springel 2005). This latter allowed to follow the evolutionary path of any DM halo in the simulations. Even though these merger trees are the skeleton of the SAM, its
finite number of outputs is not enough to properly trace the baryonic physics. Therefore, to properly follow the galaxy evolution between two consecutive DM snapshots, the SAM does an internal time discretization between them with approximately ~ 5–20 Myr of time resolution. These extra temporal subdivisions used by the SAM are called sub-steps. These extra temporal subdivisions used by the SAM are called sub-steps. Even though this work is based on the merger trees of the Millennium simulation, we combine them with the ones of Millennium II (as detailed in Section 2.2.3) to improve the galaxy and black hole initialization.

2.2 Baryonic physics

We briefly describe here the baryonic processes included in L-Galaxies that lead to the build-up of the disk and bulge components.

2.2.1 Galaxy disk

The starting point of the galaxy evolution model is the infall of baryonic matter onto every newly-resolved DM halo (see e.g. White & Frenk 1991). This process is modeled by associating an amount of matter to each halo proportionally to its DM mass, \( M_{\text{halo}} \), based on the baryon fraction of the assumed cosmology. The baryonic component initially forms a diffuse, spherical, quasi-static, hot atmosphere whose cooling leads to the formation of a cold, rotationally supported gas component according to the bi-modal cooling prescriptions described in White & Rees (1978), De Lucia et al. (2004) and Guo et al. (2011). If the mass of the cold gas disk is large enough, star formation (SF) episodes are prompted, leading to the build-up of the stellar disk component. Shortly after the SF event, a fraction of new stars explode as supernovae, enriching the environment with newly formed heavy elements. As it is widely accepted, feedback from supernova can have a severe impact on quenching the galaxy star formation. To model this, supernovae in L-Galaxies inject energy in the cold-gas disk, re-heating a fraction of it to the hot atmosphere and eventually ejecting part of the hot gas beyond \( R_{200} \). The reincorporation of these ejecta at later times helps regulating the low-\( z \) star formation, especially in low-mass galaxies (Henriques et al. 2015). Besides, in order to prevent the stellar component of massive galaxies to over-grow, the model introduces feedback from central supermassive black holes (BHs) as an additional mechanism to regulate star formation at low redshifts (see Section 3).

2.2.2 Galaxy bulge

Galaxies are allowed to develop/grow a bulge component via galaxy mergers and disk instabilities. While the former is a natural consequence of the hierarchical growth of the DM halos, the latter play a crucial role in isolated galaxies and are linked to the internal star formation.

Galaxy encounters follow the merger of the two parent halos on a time-scale given by the dynamical friction presented in Binney & Tremaine (1987). According to the baryonic (gas and stars) mass ratio of the two involved galaxies, \( m_R \), the model differentiates between major (\( m_g > m_R^{th} \)) and minor (\( m_g < m_R^{th} \)) interactions (\( m_R^{th} \) is set to 0.2 as in Izquierdo-Villalba et al. 2019). Major mergers are assumed to be able to completely destroy the disks of the two interacting galaxies, leading to a pure spheroidal remnant which undergoes a collisional starburst. In minor mergers, instead, the disk of the larger galaxy survives and experiences a burst of star formation, while its bulge incorporates the entire stellar mass of the satellite that survived stripping (as modelled by Guo et al. 2011). Besides, the model used in this work includes the prescription of smooth accretion in order to deal with the physics of extreme minor mergers (Izquierdo-Villalba et al. 2019). During those events it is expected that the stellar remnant of the satellite (bulge and disk) gets diluted inside the disk of the central galaxy before being able to reach the nucleus, thus losing the possibility of make the bulge of the primary to grow.

On the other hand, disk instability (DI) refers to the process by which the stellar disk becomes massive enough to be prone to non-axisymmetric instabilities which ultimately lead to the formation of a central ellipsoidal component via the buckling of nuclear stellar orbits (see references in Mo et al. 2010). During this process, the result is the formation of a bar and/or pseudobulge structure (Combes & Sanders 1981; Pfenniger & Norman 1990; Athanassoula 2005; Sellwood 2016; Spinoso et al. 2017). The L-Galaxies model accounts for galactic DI with a simple analytic stability criterion based on the Efstathiou et al. (1982) and Mo et al. (1998) simulations:

\[
\frac{V_{\text{max}}^2}{(GM_{*d}/R_{*d})^{1/2}} \leq \epsilon_{\text{DI}},
\]

(1)

where \( V_{\text{max}} \) is the maximum circular velocity of the host dark matter, \( R_{*d} \) and \( M_{*d} \) are respectively the exponential scale-length and stellar mass of the stellar disc and \( \epsilon_{\text{DI}} \) a parameter which determines the importance of the disk self-gravity (set to 1.5 as in Izquierdo-Villalba et al. 2019). If the stability criterion of Eq. (1) is met, an amount

\[
\Delta M_{\text{DI}} = M_{\text{disk}} - \left( \frac{V_{\text{max}}^2 R_{*d}}{G M_{*d}} \right),
\]

(2)

do of the disk stellar mass is transferred to the bulge in order to restore the disk (marginal) stability. Following the history and the physical conditions of the galaxy in which a DI takes place, the model distinguishes between instabilities that are merger-induced and the ones that are a consequence of the slow, secular evolution of galaxies. On one hand, merger-induced DIs are produced as a consequence of the fast increase of stellar disk mass after the collisional starburst or smooth satellite galaxy accretion. On the other hand, secular DIs result from the slow, but continuous, mass growth of the disk, playing an important role in galaxies evolving in isolation. Under the assumption that bars and pseudobulges are a consequence of the secular evolution of galaxies (Debattista et al. 2006; Méndez-Abreu et al. 2010; Kormendy & Ho 2013; Moledzian et al. 2017; Zana et al. 2018a; Rosas-Guevara et al. 2019) we link secular DIs to the formation of galactic bars and pseudobulges, whereas merger-induced DIs contribute to the formation of classical bulges. For more details we refer the reader to Izquierdo-Villalba et al. (2019).
Finally, in order to distinguish between bulge morphologies in the SAM, we follow the definition of Izquierdo-Villalba et al. (2019). In brief, we assume that a galaxy hosts a pseudobulge when the fraction of bulge formed via secular induced DI is at least 2/3. Galaxies display a classical bulge or elliptical structure when the fraction of bulge formed via secular induced DI is smaller than 2/3 of the total bulge mass and its bulge-to-total stellar fraction \( B/T \) is respectively 0.01 < \( B/T < 0.7 \) and \( B/T > 0.7 \). Finally, galaxies with \( B/T < 0.01 \) are considered bulgeless galaxies.

### 2.2.3 Improving the galaxy initialization in Millennium trees

As we discussed, the SAM populates each new resolved halo with an amount of hot gas in form of a diffuse atmosphere. From that moment, different analytical recipes are included to deal with the baryonic physical processes that lead to a mature galaxy. The halo mass resolution of the Millennium simulation used in this work \( M_{\text{halo}}^{-\text{Res}} \sim 10^9 M_\odot/h \) imposes a clear limit on the galaxy and BH evolutionary pathway that can be followed. Nevertheless, new resolved halos could already host a mature galaxy and its evolved BH. To improve this aspect we follow the approach of Bonoli et al. (2014) and Angulo et al. (2014) initializing each MS halo with random galaxies (hosted in the same redshift and halo mass) extracted from a run of L-Galaxies on top of the high resolution Millennium II merger trees \( M_{\text{halo}}^{-\text{Res}} \sim 10^8 M_\odot/h \). The galaxy grafting includes all the properties of the galaxy and its respective central black hole (including wandering BHs). We stress that this does not imply for our model to be able to follow halos below the resolution limit of MS, as we do not have access to their merger history and cosmological evolution. The aim of this procedure is to better match the DM mass of the newly resolved halo to the properties of the newly initialised host galaxy.

### 3 A COMPREHENSIVE MODEL OF BLACK HOLE GROWTH

Astrophysical black holes can be fully determined by two quantities: mass and spin. In this section we present the implemented physic to track their evolution along the cosmic history.

#### 3.1 Black hole seeding and spin initialization

Even though the existence of BHs is now well established, their origin is still an open issue. Up to date, several theories have been proposed and explored (see Volonteri 2010; Mayer & Bonoli 2019, for a review). The most studied one is the light seed scenario where BHs of mass between \( 10^{-9} \) and \( 10^{-4} M_\odot \) are formed at \( z \gtrsim 20 \) as remnants of Population III stars (Pop III, Madau & Rees 2001; Heger & Woosley 2002). Another complementary formation channel is the heavy seed scenario where much more massive black holes \( \sim 10^4 - 10^5 M_\odot \) are formed after the direct collapse of massive gas clouds, either in pristine primordial halos or in gas-rich merging galaxies (Koushiappas et al. 2004; Volonteri & Stark 2011; Mayer et al. 2010; Bonoli et al. 2014). In this work we simply assume that each newly resolved halo (independently of redshift and halo properties) is seeded with an initial BH mass of \( 10^3 M_\odot \). This initial seed mass is a reasonable assumption given the minimum mass of new resolved halos in MS and MSH \( \sim 10^6 M_\odot \) and \( \sim 10^8 M_\odot \), respectively. In future works, a proper BH seeding in L-Galaxies will be tackled (Spinosa et al. in prep.).

Besides the initial mass, the BH angular momentum, \( |J_{\text{BH}}| \), has to be set. The exact value of \( |J_{\text{BH}}| \) can be expressed as:

\[
|J_{\text{BH}}| = \frac{1}{\sqrt{2}} M_{\text{BH}} a (G M_{\text{BH}} R_{\text{Sch}}) \frac{1}{2} = \frac{a G^2 M_{\text{BH}}^2}{c},
\]

where \( R_{\text{Sch}} = 2 G M_{\text{BH}} / c^2 \) is the Schwarzschild radius, \( G \) the gravitational constant, \( c \) the light speed, and \( a \) the dimensionless spin parameter spanned between \( 0 < |a| < 0.998 \) (Thorne 1974). While \( M_{\text{BH}} \) and \( R_{\text{Sch}} \) are fully determined by the seed mass, the value of \( a \) does not have to depend on the initial mass. Indeed, no observational constraints and theoretical predictions about the initial BH spin exist to date (see e.g. Battaglia et al. 2017; Fuller & Ma 2019; Roulet & Zaldarriaga 2019; Zackay et al. 2019). Here we assume that the initial value of \( a \) is a random number selected between \( 0 < |a| < 0.998 \). We have tested that the initial selection of \( a \) does not affect the results presented in this work. In fact, a BH looses any memory of its initial spin when it has accreted in a coherent way a gas mass comparable to its own (see for instance Pandakakis et al. 2011, Figure 1).

#### 3.2 Black hole gas accretion

In this section we summarize the different mechanisms that feed central black holes with gas: (i) galaxy mergers, (ii) galaxy disk instabilities and, (iii) gas accretion from the hot gas atmosphere.

##### 3.2.1 Cold phase accretion

The cold gas accretion is triggered by two different processes: galaxy mergers and disk instabilities.

i) Gas accretion after mergers. Simulations have shown that gravitational torques during galaxy mergers are able to drive cold gas towards the galaxy inner regions, triggering the black hole accretion (Di Matteo et al. 2005; Springel 2005; Hopkins et al. 2009). Following Kaufmann & Haehnelt (2000), Croton (2006) and Bonoli et al. (2009) we assume that the fraction of cold gas accreted by the BH after a galaxy merger is:

\[
\Delta M_{\text{gas}} = \frac{m_{\text{gas}}}{1 + (V_{\text{BH}} / V_{200c})^2} M_{\text{Cold}}.
\]

where \( \Delta M_{\text{gas}} \) includes a dependence on the galaxy baryonic merger ratio, \( m_{\text{gas}} \leq 1 \), the virial velocity of the host DM halo, \( V_{200c} \), the redshift of the merger, \( z_{\text{merge}} \); and two adjustable parameters \( V_{\text{BH}} \) and \( z_{\text{merge}} \) set to \( 280 \text{ km/s} \) and \( 0.034 \), respectively.

ii) Gas accretion after disk instabilities. Non-axisymmetric instabilities of galactic disks have a deep impact on the morphology of the nuclear parts. They are
able to modify the gas disk structure via gravitational torques (see Shlosman & Begelman 1989; Hopkins & Quataert 2010). These processes produce strong gas inflows that lead to nuclear bursts of star formation and can drive part of the gas also in the BH surroundings (Shlosman et al. 1989; Hopkins et al. 2009; Hopkins & Quataert 2010; Fanali et al. 2015; Du et al. 2017; Spinoso et al. 2017). Consequently, each time a galaxy undergoes a disk instability episode we allow the central black hole to accrete an amount of gas given by:

\[
\Delta M_{\text{BH}}^{\text{gas}} = \frac{\Delta M_{\text{BH}}^{\text{DI}}}{1 + (V_{\text{BH}}/V_{200c})^2},
\]

(5)

where \(V_{\text{BH}}\) and \(V_{200c}\) are the same parameters described before, \(\Delta M_{\text{BH}}^{\text{DI}}\) the redshift in which the DI takes place and \(\Delta M_{\text{BH}}^{\text{DI}}\) = the mass of stars that triggers the stellar disk instability (see Eq. 2). \(\Delta M_{\text{BH}}^{\text{DI}}\) is a free parameter that takes into account the accretion efficiency. As we discussed in Section 2.2.2, we assume DIs to be the consequence of either the galaxy secular evolution or of galaxy minor mergers. In the former case, we set the efficiency \(\eta_{\text{DI}}\) = 0.008. Instead, for DI merger induced which are causally connected with the last galaxy minor merger (\(\lesssim 10^9 M_{\odot}\), Izquierdo-Villalba et al. 2019) we assume \(\eta_{\text{Di}}^{\text{merger}} = 0.08\). By using Eq. (6), Eq. (7) and Eq. (8), we refer to other works such as Pacucci & Ferrara (2015) or Volonteri et al. (2015) in which the same distinction is performed.

By definition of Eddington limit, the first phase is characterized by \(f_{\text{Edd}}(t) = 1\). The BH grows in this phase for the times necessary to reach the mass \(M_{\text{BH, peak}}\) defined by the moment in which the BH consumed a fraction \(\gamma\) (free parameter set to 0.7) of its whole gas mass reservoir (see Marulli et al. 2008; Bonoli et al. 2009; Hirschmann et al. 2012). Once this mass is reached, the black hole enters in a self-regulated or quiescent growth regimen characterized by small \(f_{\text{Edd}}\) values. Following Hopkins et al. (2006b) we assume that \(f_{\text{Edd}}\) in this phase is parameterized as:

\[
f_{\text{Edd}}(t) = \frac{1}{1 + (t/t_0)^{1/2}} \eta_{\text{Edd}}.
\]

(9)

where \(t_0 = t_0 \Theta(t/\beta)\ln(10)\), being \(t_0 = 1.26 \times 10^8 \text{yr}\), \(\beta = 0.4\) and \(\xi = 0.3\) (see the discussion of these values in Hopkins et al. 2006b). We highlight that \(t\) is relative to the moment in which the BH finishes the Eddington phase.

As pointed out by Hopkins et al. (2006b) and Merloni & Heinz (2008) we also take into account the fact that the radiative efficiency has a dependence with the accretion efficiency and the nature of the accretion flow:

\[
\epsilon(t, f_{\text{Edd}}) = \begin{cases} \eta_{\text{Edd}} & f_{\text{Edd}} > f_{\text{Edd}}^{\text{crit}} \\ \frac{f_{\text{Edd}}}{f_{\text{Edd}}^{\text{crit}}} & f_{\text{Edd}} \leq f_{\text{Edd}}^{\text{crit}} \end{cases}
\]

(10)

where \(f_{\text{Edd}}^{\text{crit}}\) (Merloni & Heinz 2008; Volonteri et al. 2013, set to 0.03) is the Eddington factor by which below it the accretion disc becomes radiatively inefficient.

Finally, the black hole mass evolution during any accretion event can be obtained from the black hole growth rate, \(M_{\text{BH}} = dM_{\text{BH}}/dt\). By using Eq. (6), Eq. (7) and Eq. (8), \(M_{\text{BH}}\) can be computed as:

\[
M_{\text{BH}} = M_{\text{BH}}^{\text{peak}} = \frac{1 - \eta(t)}{\epsilon(t)} \frac{M_{\text{BH}}(t)}{\epsilon(t)},
\]

(11)

As we will see in Section 3.3, \(\epsilon\) and \(\eta\) experience a time variation given their dependence with the black hole spin, \(a\), which varies with time as a consequence of BH-BH mergers and gas accretion. In order to ease notation, instead of denoting the accretion and radiative efficiency as \(\eta(a(t))\) and \(\epsilon(a(t))\) we will use just the notation \(\eta\) and \(\epsilon\).

### 3.2.2 Hot phase accretion

This mechanism of black hole growth is triggered by the accretion of the hot gas which surrounds the galaxy. Following Henriques et al. (2015) the BH mass accretion rate during this mode is determined by:

\[
M_{\text{BH}} = k_{\text{AGN}} \left( \frac{M_{\text{BH}}}{10^{11}M_{\odot}} \right) \left( \frac{M_{\text{BH}}}{10^8M_{\odot}} \right).
\]

(12)
Thus, for most of the accretion event, the inner part of the disk and its angular momentum. This process is called Bardeen-Petterson effect (Bardeen & Petterson 1975a). For large-viscosity disks this misalignment configuration is unstable making the orbital plane of the inner parts of the disk align (prograde orbit) or counter-align (retrograde orbit) to the black hole angular momentum. This process is called Bardeen-Petterson effect (Bardeen & Petterson 1975b). The criterion of counter-alignment was established by King et al. (2005) and it takes into account the ratio between the black hole and disk angular momentum and the angle \( \theta \) formed between them:

\[
\cos \theta < -\frac{|J_d|}{2|J_{BH}|}.
\]

Thus, for most of the accretion event, the inner part of the accretion disk is aligned if i) \( |J_d| > 2|J_{BH}| \) or ii) \( |J_d| < 2|J_{BH}| \) and \( \theta < \pi/2 \). On the contrary, when \( |J_d| < |J_{BH}| \) and \( \theta > \pi/2 \), the (anti-)alignment depends if the ratio of \( |J_d|/|J_{BH}| \) is small enough with respect to exact value of \( \cos \theta \).

We model this in our semi-analytical framework following the spin model presented in Dotti et al. (2013) and Sesana et al. (2014). The model assumes that the gas available during an accretion event (given by \( M_{BH} \), see Section 3.2) is accreted in chunks of transient accretion disks limited by self-gravity, instead of being consumed in a single episode (King et al. 2008). In an \( \alpha \)-disk the self-gravity radius, \( R_{\text{disk}}^{\alpha} \), is determined by the distance to the BH at which the Toomre parameter \( Q \) is less than unity: \( Q \sim c_s \Omega / \pi G \Sigma \lesssim 1 \),

where \( \Omega \) is the Keplerian angular velocity, \( c_s \) sound speed and \( \Sigma \) the gas surface density. Thus, the value of \( R^{\alpha}_{\text{disk}} \) is:

\[
R^{\alpha}_{\text{disk}} \approx \frac{55^{22/45} \alpha^{28/45} (10^{9} M_{\odot})^{-52/45} (\eta/0.1)^{22/45}}{R_{\text{Sch}}},
\]

(14)

where \( R_{\text{Sch}} \) is the Schwarzchild radius and \( \alpha \) the radial shear viscosity parameter (set hereafter to 0.1). Concerning the self-gravity mass, \( M^{\alpha} \), the value is expressed as:

\[
M^{\alpha} \approx 2 \times 10^{45} f_{\text{Edd}}^{-1/45} \left( \frac{\alpha}{0.1} \right)^{-1/45} \left( \frac{M_{BH}}{10^{9} M_{\odot}} \right)^{34/45} M_{\odot},
\]

(15)

which was computed by solving the expression \( M_{\text{cld}} = 2 \pi \Sigma(R) R^2 dR \) with a surface density:

\[
\Sigma(R) = 7 \times 10^{-7} \left( \frac{\alpha}{0.1} \right)^{-4/5} \left( \frac{M_{BH}}{10^{9} M_{\odot}} \right)^{19/20} \left( \frac{\eta}{0.1} \right)^{-7/10} \left( \frac{R}{R_{\text{Sch}}} \right)^{-3/4} \text{g cm}^{-2}.
\]

(16)

To avoid unrealistic accretion episodes, Dotti et al. (2013) set the maximum mass of the transient accretion disk to the minimum between \( M^{\alpha} \) and a fixed molecular cloud mass, \( M_{\text{cld}} \), set in our case to \( 7 \times 10^{9} M_{\odot} \). As pointed out by Sesana et al. (2014), fixing the value of \( M^{\alpha}_{\text{cld}} \) is an idealized case, as molecular clouds might be characterized by a broad mass spectrum \((\lesssim 10^{9} M_{\odot})\) eventually dependent on the host galaxy properties. However, since the cloud mass function is not trivial, we follow Dotti et al. (2013) keeping \( M^{\alpha}_{\text{cld}} \) constant over the whole BH life. The value of \( M^{\alpha}_{\text{cld}} \) plays also a relevant role in the model, determining the mass above which a BH changes form maximally to not maximally spinning. Specifically, the transition happens when \( M^{\alpha} = M^{\alpha}_{\text{cld}} \) with a critical BH mass of \( 4 \times 10^{9} (0.1 f_{\text{Edd}}/\eta)^{-2/27} (M^{\alpha}_{\text{cld}}/10^{9} M_{\odot})^{45/34} M_{\odot} \) (Dotti et al. 2013). For a given cloud mass, the radius of the accretion disk is:

\[
\frac{R^{\alpha}_{\text{disk}}}{R_{\text{Sch}}} = 4 \times 10^{-4} f_{\text{Edd}}^{-14/25} \left( \frac{M_{\text{cld}}}{10^{9} M_{\odot}} \right)^{4/5} \alpha^{16/25} \left( \frac{\eta}{0.1} \right)^{14/25}.
\]

(17)

Each transient accretion disk has associated an angular momentum \( |J_d| \) determined by Perego et al. (2009):

\[
|J_d| = \frac{8}{21} R^{7/4}_{\text{disk}} (GM_{BH})^{1/2} M_{\text{BH}},
\]

(18)

where

\[
R^{\text{disk}} = \frac{9.14 \times 10^{6} f_{\text{Edd}}^{3/10}}{f_{\text{Edd}}} \alpha^{4/5} \left( \frac{M_{BH}}{10^{9} M_{\odot}} \right)^{1/20} \left( \frac{\eta}{0.1} \right)^{-3/10} \text{cm}^{5/4} / \text{s},
\]

(19)

and i) \( R_d = R^{\alpha}_{\text{disk}} \) if \( M^{\alpha}_{\text{cld}} < M^{\alpha} \) or ii) \( R_d = R^{\alpha}_{\text{disk}} \) if \( M^{\alpha} < M^{\alpha}_{\text{cld}} \).

\footnote{The value of \( R^{\alpha}_{\text{disk}} \) is computed by solving \( M_{\text{cld}} = 2 \pi \Sigma(R) R^2 dR \).}
Therefore, the BH spin evolution during any gas accretion episode is followed by using Eq.(14)-(18) and checking Eq.(13). If $|\tilde{J}_d| > 2|\tilde{J}_{BH}|$, the fraction of transient disks consumed in a prograde accretion, $n_{Pa}$, is always equal to 1. On contrary, when $|\tilde{J}_d| < 2|\tilde{J}_{BH}|$ the scenario is more complex and the exact value of $\cos \theta$ is needed to check the (anti)alignment. To determine $\cos \theta$ in the regime of $|\tilde{J}_d| < 2|\tilde{J}_{BH}|$, previous works have assumed the chaotic accretion presented in King et al. (2005, 2008), consisting in choosing randomly the value of $\cos \theta$ (Faniakis et al. 2011; Volonteri 2008). Here, instead, following the approach of Dotti et al. (2013) and Sesana et al. (2014), we assume a more general scenario, allowing an asymmetry in the $n_{Pa}$ value. According to Sesana et al. (2014) the fraction of transient disks accreted in a prograde accretion when $|\tilde{J}_d| < 2|\tilde{J}_{BH}|$ is given by:

$$n_{Pa} = F + \frac{|\tilde{J}_d|}{2|\tilde{J}_{BH}|}(1 - F), \quad (20)$$

where $F$ is an isotropy parameter which takes into account the fraction of accretion events with an initial angular momentum with $\theta < \pi/2$.

Following Sesana et al. (2014) we link $F$ to the bulge morphology. This connection attempts to address the fact that the gas driven towards the BH surroundings retains certain memory of both the event which triggered its nuclear in-flow and the bulge assembly (Hopkins et al. 2009; Hopkins & Quataert 2010). In particular, in Sesana et al. (2014) the exact value of $F$ is associated to the ratio between the bulk rotation velocity of the bulge, $\nu$, and its velocity dispersion, $\sigma$. This ratio gives an idea of the coherence motion of the galactic bulge. If $v/\sigma > 1$ then $F = 1$, if $v/\sigma = 0$ then $F = 0.5$ and if $0 < v/\sigma < 1$ then $F = (1 + v/\sigma)/2$. Thus, bulges characterized by a large coherent motion (e.g. bars) feed the BHs with a larger number of transient chunks of co-rotating accretion disks than bulges dominated by velocity dispersion (e.g. ellipticals).

Since the exact value of $v/\sigma$ is shaped by a complex galaxy dynamics which $L$-Galaxies does not address, we decide to take the $v/\sigma$ values from the observed probability distribution function (PDF) of pseudobulges, classical bulges and ellipticals presented in Eq. 14-18 of Sesana et al. (2014). In brief, while elliptical structures have an asymptotic PDF around 0, classical bulges have a PDF centred at $\approx 0.55$ (see Eq.15 and 16 of Sesana et al. 2014). For pseudobulges, we follow the hybrid model of Sesana et al. (2014) assuming a log-normal PDF with 0.34 dex of standard deviation and an average value given by its Eq.14. Following Izquierdo-Villalba et al. (2019) we assume that in $L$-Galaxies, DI secular events build-up the pseudobulge/bar structure, minor mergers and DIs merger-induced assembly the classical bulge component, and major mergers shape an elliptical galaxy. Therefore, each time that a galaxy undergoes one of these events we extract a random number form its $v/\sigma$ out-coming bulge PDF and we perform a mass-weighted average between that value and the one which characterizes the bulge at that moment. In the case of major mergers, we reset the value of $v/\sigma$ with a value randomly selected from the elliptical-bulge PDF. For the special case of smooth accretion events which bring gas onto the BH but do not build-up a bulge structure, we assume that the $v/\sigma$ of the gas brought towards the BH surroundings is linked to the kinematics of the galaxy gaseous disk. For that, as we did with the pseudobulges, we assume a log-normal PDF with 0.34 dex of standard deviation and an average value given by Eq.14 of Sesana et al. (2014).

Following Thorne (1974) we can determine the evolution of the BH spin by:

$$\dot{a} = \left[ \frac{L_{\text{ISCO}}}{a} - 2aE_{\text{ISCO}}(a) \right] \left[ \frac{M_{\text{BH}}}{M_{\text{BH}}} \right]$$

(21)

where $M_{\text{BH}}$ is the black hole mass, $M_{\text{BH}}$ the accretion rate onto the black hole and $L_{\text{ISCO}}$, $E_{\text{ISCO}}$ quote respectively the specific angular momentum and energy in the innermost stable circular orbit (ISCO). Since an accretion event is composed by a fraction of $n_{Pa}$ prograde and $(1-n_{Pa})$ retrograde orbits, Eq. (21) can be rewritten as (Sesana et al. 2014; Barausse 2012):

$$\dot{a} = \left[ \left( n_{Pa}a_{\text{ISCO}}^{\text{pro}}(a) + (1-n_{Pa})a_{\text{ISCO}}^{\text{retro}} \right) - 2a \left[ n_{Pa}a_{\text{ISCO}}^{\text{pro}}(a) + (1-n_{Pa})a_{\text{ISCO}}^{\text{retro}} \right] \right] \frac{M_{\text{BH}}}{M_{\text{BH}}}$$

(22)

where $a_{\text{ISCO}}^{\text{pro}}$ and $a_{\text{ISCO}}^{\text{retro}}$ are, respectively, the specific angular momentum and energy in a prograde (retrograde) ISCO (Bardeen et al. 1972).

Finally, the accretion efficiency, $\eta$, it is computed as (Bardeen 1970):

$$\eta = 1 - E_{\text{ISCO}} = 1 - \sqrt{1 - \frac{2}{3} \frac{1}{1 + n_{\text{ISCO}}^{\text{pro}}}}$$

(23)

where $n_{\text{ISCO}}^{\text{pro}}$ is the radius of the ISCO and its exact value depends in the spin value and the (anti)alignment of the transient accretion disk (Bardeen et al. 1972; Barausse et al. 2012):

$$n_{\text{ISCO}} = 3 + Z_2 = a \sqrt{(3 - Z_1)(3 + Z_1 + 2Z_2)}$$

(24)

where the upper and lower sign refers respectively to a prograde and retrograde orbits and the values of $Z_1$ and $Z_2$ are determined by:

$$Z_1 = 1 + (1-a^2)^{1/3} \left[ (1+a)^{1/3} + (1-a)^{1/3} \right]$$

(25)

$$Z_2 = \sqrt{3a^2 + Z_1^2}$$

Therefore, the final value of $\eta$ after a fraction of prograde $n_{Pa}$ and $(1-n_{Pa})$ of retrograde orbits is:

$$\eta(a_{\text{BH}}) = n_{Pa} \eta_{\text{pro}}(a) + (1-n_{Pa})\eta_{\text{retro}}(a)$$

(26)
3.3.2 Spin evolution after a BH-BH coalescence

The last stage of a BH binary system is the coalescence. The final spin of the remnant BH is shaped by the initial spin configuration of its progenitors and their initial masses. In this work we follow the results of Barausse & Rezzolla (2009) which found that the spin modulus of the remnant black hole, $|a_f|$, is determined by:

$$|a_f| = \frac{1}{(1+q)^2} \left[ |a_1|^2 + |a_2|^2 + 2 |a_1| q^2 \cos \alpha \right. + 2 \left(|a_1| \cos \beta + |a_2| q^2 \cos \gamma \right)|\ell| + |\ell|^2 q^2 \right]^{1/2},$$  

(27)

with

$$|\ell| = 2 \sqrt{3} + t_2\nu + t_3\nu^2 +$$

$$\frac{s_4}{(1 + q^2)^2} \left[ (|a_1|^2 + |a_2|^2) q^4 + 2 |a_1| |a_2| q^2 \cos \alpha \right. +$$

$$\left. + \frac{s_5 \nu + t_0 + 2}{1 + \nu^2} (|a_1| \cos \beta + |a_2| q^2 \cos \gamma), \right.$$  

(28)

where $q = M_{BH1}/M_{BH2} \leq 1$ is the binary mass ratio, $\nu = q/(1 + q^2)$ the symmetric mass ratio, $|a_1|$ and $|a_2|$ are the initial spin magnitudes, $\alpha$ is the angle formed between the two spins and $\beta, \gamma$ the angle formed between the spin (1 and 2) with the orbital angular momentum, respectively. The other parameters are set to the following values: $s_4 = 0.1229 \pm 0.0075$, $s_5 = 0.4537 \pm 0.1463$, $t_0 = -2.8904 \pm 0.0359$, $t_1 = 2.5763 \pm 0.4833$ and $t_2 = -3.51714 \pm 0.1208$ (Barausse & Rezzolla 2009; Barausse 2012).

Given that in the SAM we only track the evolution of the BH spin modulus we have to establish the direction between the two black hole spins and their directions with respect to the orbital angular momentum, i.e the values of $\alpha, \beta$ and $\gamma$ of Eq.(27) and Eq.(28). To do that, we use the standard approach of dividing the BH-BH mergers in two types (Barausse 2012; Volonteri et al. 2013): wet and dry. While the former is characterized by $(M_{Res,1}+M_{Res,2}) > (M_{BH1}+M_{BH2})$, the latter is characterized by $(M_{Res,1}+M_{Res,2}) < (M_{BH1}+M_{BH2})$. In the case of wet mergers, it has been shown that the dissipative dynamics between black holes and the massive gas disc produces a torque with the effect of aligning both spins to the orbital angular momentum. However, Dotti et al. (2010) showed that this alignment is not perfect and there is a residual offset in the spin direction relative to the orbital angular momentum at the level of $10^\circ$. Therefore for the wet BH mergers we assume that $\cos \alpha, \cos \beta$ and $\cos \gamma$ are random numbers between $\cos 0^\circ - \cos 10^\circ$. For dry mergers, we assume random values of $\cos \alpha, \cos \beta$ and $\cos \gamma$.

3.4 Black hole coalescence

3.4.1 Black hole binaries: merger delay

In the standard hierarchical paradigm galaxies grow mainly through mergers with smaller or comparable mass companions (White & Rees 1978). This fact, together with recent observational results confirming the existence of BHs being hosted in the center of galaxies (Haehnelt & Rees 1993), suggest that binary BH (BBH) systems have been formed and merged during the whole lifetime of the universe. The evolutionary pathway of BBHs is described by three different stages (Begelman et al. 1980): i) an initial pairing phase in which dynamical friction exerted by the galaxy gas and stars drives both BHs individually to the nucleus of the merger remnant eventually forming a binary; ii) a following hardening phase in which the BBH orbital separation shrinks due to three-body slingshots and/or interaction with a massive gaseous disk; iii) a final gravitational wave inspiral phase which drives the binary to the final coalescence (see Colpi 2014, and references therein for a review). Building on this scenario, several studies have tackled the evolution of BHs (see for instance Quinlan & Hernquist 1997; Sesana 2013; Kelley et al. 2017a,b; Bonetti et al. 2018b,a). On the semi-analytical modelling side, up to date, most of the studies presented in the literature have assumed an instantaneous coalescence right after a galaxy-galaxy merger (see the few exceptions of Volonteri et al. 2003; Tanaka & Haiman 2009; Antonini et al. 2015; Bonetti et al. 2018a,b). This might be an acceptable assumption at high-$z$, where galaxy interactions are gas rich and dynamical friction can efficiently reduce the angular momentum of both BHs and speed up the hardening and final coalescence (Mayer 2013; del Valle et al. 2015). However, at low-$z$ galaxies are generally poorer in gas and the timescales for merger can become very large.

In this work we relax the assumption of instantaneous BH-BH coalescence by allowing for a time delay between the galaxy merger and the BH binary final coalescence. Therefore, we introduce the possibility for galaxy-merger remnants of hosting a binary BH system for a given time interval $\tau_{\text{delay}}$. Following the results of several dedicated hydrodynamical simulations (Callegari et al. 2011a,b; Fiacconi et al. 2013; Mayer 2013; del Valle et al. 2015), we implement a simple prescription to determine the lifetime of a BH binary system:

$$\tau_{\text{delay}}^\text{BH} = 0.01 \left( \frac{0.1}{q} \right) \left( \frac{0.3}{f_{\text{gas}}} \right) \frac{1}{F(e)} \text{Gyr},$$  

(29)

where $q (< 1)$ is the mass ratio between the two BHs at the moment of the galaxy-galaxy merger, $f_{\text{gas}}$ is the galaxy gas fraction and $F(e)$ is a factor which takes into account the dependence of $\tau_{\text{delay}}^\text{BH}$ with the binary eccentricity $e$ (for simplicity, we assume $F(e) = 1$). Even though the hydro-dynamical simulations of Escala et al. (2004, 2005), Dotti et al. (2007) and Cuadra et al. (2009) showed that dense gaseous regions are very effective in hardening BBHs, promoting their coalescence in less than $\lesssim 10^7$ yr, the dependence with the gas fraction might be much more complex. For instance, Tamburello et al. (2017) showed that the final BH-BH coalescence can be speeded-up or delayed depending on the lumpiness of the circumnuclear...
disks around the BHs. In a future work we plan to explore how different assumptions for the merger delays affect the global BH population.

We allow both BHs to accrete their respective gas reservoir carried before the merger while they are in a binary system. We also assume that any new amount of gas accreted after the galaxy merger (see Section 3.2.1) is added to the reservoir of the most massive BH. Still, the life of a galaxy is rather complicated and it can undergo multiple mergers in short time scales. If a merger happens in a moment in which the galaxy is still hosting a BBH, a third BH (or even another BBH) will perturb the system (Hut & Rees 1992). The final configuration is not trivial and few studies have addressed this (see e.g. Bonetti et al. 2018a,b, and references therein). Here we use a simplified approach allowing only the two most massive BHs to remain in the galaxy (see Volonteri et al. 2003; Volonteri & Perna 2005). The lightest BH suffers a scattering event and it is ejected from the galaxy in the system. We highlight that we do not follow the wandering phase of these scattered BHs since we do not track the binary system separation (as, for instance, Volonteri et al. 2003 do) and therefore we cannot set the initial velocity of the ejection (see Section 3.5). In Appendix A the number density of both gravitational and 3-body scattering ejections is compared. We show that at any redshift and halo mass, the former is $\gtrsim 2\text{dex}$ more frequent. Finally, if the two initial BHs forming the binary system are the most massive ones, we eject the intruder without changing the $t_{\text{BH}^\text{delay}}$ value. Conversely, if there was an exchange of BHs in the system, we reset the value of $t_{\text{BH}^\text{delay}}$ with the new initial conditions. The distribution of BH-BH merger delays, $t_{\text{BH}^\text{delay}}$, is presented in Fig. 1. As shown, the earlier is the redshift of binary formation, $t_{\text{BH}^\text{form}}$, the shorter is the time spent by the two BHs in a binary system. While at $t_{\text{BH}^\text{form}} > 3$ BHs merge in $\lesssim 100\text{Myr}$, at $t_{\text{BH}^\text{form}} < 3$ a considerable fraction of BHs spend more than $1\text{Gyr}$ orbiting each other before the final coalescence. Besides, at $t_{\text{BH}^\text{form}} < 3$ a sizeable fraction of BBHs survive for times larger than the lookback time corresponding to their $t_{\text{BH}^\text{form}}$, and therefore still inhabit galaxy nuclei today. The increase of the binary system lifetime towards low-$z$ is caused principally by the dependence of the gas fraction in Eq. (29).

\[ t_{\text{BH}^\text{delay}} = 100\text{Myr} \]

where $m_0 = 0.9515 \pm 0.001$, $m_1 = -0.013 \pm 0.007$.

3.5 The population of wandering black holes

Black holes outside of a galaxy in bound orbits within the dark matter halo are defined as wandering black holes (wBHs). In this section we study two types of events responsible of the formation of the wBH population: (i) disruption of satellite galaxies and (ii) gravitational recoils after a BH-BH merger.

3.5.1 Orphan black holes: a disruption event of its host galaxy

In L-Galaxies, as soon as two dark matter halos merge their host galaxies do it as well in a time scale given by the dynamical friction formula of Binney & Tremaine (1987). The smallest galaxy starts to sink towards the inner parts of its new host halo. During this process, if the galaxy is not compact enough it can be disrupted before reaching the halo center, losing the possibility to merge with the central galaxy. When this happens, the central black hole (or binary system) of the satellite galaxy is deprived of its host and is incorporated in the dark matter halo as a wBH. The disruption process implemented in L-Galaxies is presented in Guo et al. (2011). We follow the orbit of the black hole using the formalism presented in Section 3.5.3. As initial conditions, we consider the position in which the satellite galaxy is disrupted and its infalling-velocity modulus evaluated at the same instant. Since L-Galaxies do not track the 3D components of halo infalling velocities, we resort to assuming the initial velocity as radially oriented and track only the 1D equations of motion of the wBHs. Additionally, we assume that the black hole completely loses its accretion disk. From here onwards, we tag this type of wBHs as...
orphan black holes. In the upper panel of Fig. 2 we present a cartoon showing the formation of the orphan black hole population.

3.5.2 Ejected black holes: Kicks due to gravitational recoils

During a BH merger, the binary system emits gravitational waves able to carry away energy and angular momentum. Due to conservation of linear momentum, in the instant of the merger, the emission of gravitational waves imparts a kick to the remnant black hole. The velocity characterizing this kick is called recoil velocity, \( \vec{V}_{\text{recoil}} \). Using numerical simulations of general relativity, it has been shown that recoil velocities can take a wide range of values: from small ones which just displace the remnant BH from the galaxy center, to larger values big enough to exceed the escape velocity of the host galaxy or even of the host halo (Baker et al. 2008; Lousto & Zlochower 2008; van Meter et al. 2010). Here we use the fitting formula established by Lousto et al. (2012):

\[
\vec{V}_{\text{recoil}} = v_m \hat{e}_1 + v_\perp (\cos \xi \hat{e}_1 + \sin \xi \hat{e}_2) + v_l \hat{n}_l.
\]

\[
v_m = A_m \sqrt{\frac{(1 - q)}{(1 + q)} (1 + B_m v)}.
\]

\[
v_\perp = H \sqrt{\frac{v^2}{1 + q}} \left( a^2 - q a_0 \right).
\]

\[
v_l = \frac{16v^2}{(1 + q)} \left| V_{1,1} + V_A \hat{S}_A + V_B \hat{S}_B + V_C \hat{S}_C \right| |q_a^2 - qa_0|^2 \cos(\phi_\Lambda - \phi_1).
\]

where, \( v_m \) is the velocity due to the mass-asymmetry contribution and, \( v_\perp \) and \( v_l \) are spin contribution producing kicks parallel and perpendicular to the orbital angular momentum respectively. As in the previous section, \( q = m_1/m_2 \leq 1 \) is the binary mass ratio, \( v = \sqrt{(1 + q)^2 - 1} \) is the symmetric mass ratio, \( |a_\phi| \) is the initial spin magnitude of the black hole \( i \), || and \( \perp \) refer respectively to components parallel and perpendicular to the orbital angular momentum. \( \hat{e}_1 \) and \( \hat{e}_2 \) are orthogonal unit vectors in the orbital plane. \( \xi = 145^\circ \) is the angle between the unequal mass and spin contribution (González et al. 2007; Lousto & Zlochower 2008), \( S = 2(a_2 + q^2a_0)/(1 + q)^2 \). \( \phi_\Lambda \) is the angle between the two line components \( \hat{S}_A = (M_{BH1} + M_{BH2})\hat{S}_2^1/M_{BH1} - \hat{S}_2^2/M_{BH2} \) and the infall direction at merger. The values of the other coefficients are obtained numerically: \( A_m = 1.2 \times 10^{4} \), \( B_m = -0.93 \), \( H = 6.9 \times 10^{5} \) (González et al. 2007; Lousto & Zlochower 2008) and \( V_{1,1} = 3677.76 \text{ km/s} \), \( V_A = 2481.21 \text{ km/s} \), \( V_B = 1792.45 \text{ km/s} \) and \( V_C = 1506.52 \text{ km/s} \) (Lousto et al. 2012).

If the modulus of this velocity is larger than the escape velocity of the galaxy (\( V_{\text{esc}} \)) the kicked BH is ejected from the host galaxy, starting its wandering phase.
halo. This happens more frequently at high-$z$, where the halos are smaller and their potential wells shallower. This scenario is summarized in the lower cartoon displayed in Fig. 2. Hereafter, we call this type of wBHs ejected black holes. As we did for orphan BHs, we assume a radial velocity for the orbital integration (see Section 3.5.3). Finally, we neglect the fact that the ejected BH can retain an accretion disk. It is not clear how massive is the disk that can be retained and subsequently accreted by a BH after a recoil (see Blecha et al. 2011). Loeb (2007) showed that in an $a$-disk the maximum amount of the disk material that a recoiled BH can carry out is, at most $\sim M_{\text{BH}}$ which account for a relatively small AGN lifetime ($\sim 4.5 \times 10^6$ years, Volonteri & Madau 2008). In this work we neglect the accretion disk of ejected BHs, but we plan to include it in a future work, focused on the observability of AGN pairs.

### 3.5.3 Tracking the orbits of wandering black holes

Once in the wandering phase, the BH starts to orbit within the halo potential (i.e. around its centre $r_0$). Assuming a radial motion along the direction $\hat{r}$, the equation of motion can be expressed as:

$$\ddot{r} = \left(\frac{G M(<(r-r_0))}{(r-r_0)^2} + \frac{a_{\text{df}}^{\text{class}} + a_{\text{df}}^{\text{gas}}}{v_{\text{BH}}^2}\right) \hat{r},$$

where the first term refers to the gravity acceleration caused by a mass $M(<(r-r_0))$ of dark matter, stars and gas (in both galaxy and inter-cluster medium) within a radius $r-r_0$. The second and third terms are related to the dynamical friction caused by collisionless (stars and dark matter) and gas components, respectively. Since in this work we are interested in the position of wBHs, we track their orbital evolution by solving numerically Eq. (32) using a $4^{th}$ order Runge-Kutta integrator with time step of $1\,\text{Myr}$. We have also used integration intervals of $0.1\,\text{Myr}$, obtaining nearly identical results.

Assuming a Maxwellian velocity distribution and a black hole velocity $v_{\text{BH}}$, the value of $a_{\text{df}}^{\text{class}}$ is determined by (Chandrasekhar 1943):

$$a_{\text{df}}^{\text{class}} = -\frac{4\pi G^2 M_{\text{BH}} \rho(r-r_0) \ln \Lambda}{v_{\text{BH}}^2} \left[ err f(I) - \frac{2 I}{\sqrt{\pi}} e^{-I^2} \right],$$

with $I = |v_{\text{BH}}|/\sqrt{2} \sigma$ where $\sigma$ is the dark matter velocity dispersion computed as $\sigma = (GM_{\text{halo}}/2\,\text{R}_{200})^{1/2}$. $err f$ is the error function and $\ln \Lambda$ is the Coulomb logarithm fixed to 3.1 (Escala et al. 2004; Gualandris & Merritt 2008; Blecha & Loeb 2008a). Finally, $\rho(r-r_0)$ is the mass density of the collisionless enclosed within $r-r_0$. In our case $\rho(r-r_0) = \rho_{\text{DM}}(r-r_0) + \rho_{\text{ICM}}(r-r_0)$, i.e. the mass of dark matter and the mass of stars in the inter-cluster medium (ICM or stellar halo). For the case of dark matter we assume a NFW profile (Navarro et al. 1996) while for ICM an isothermal one.

Since galaxies are surrounded by hot gas which fills their host DM halos, we also take into account the dynamical friction caused by that gas. Escala et al. (2004) determined that the damping of the oscillation amplitude for an object moving inside a gas follows the equation:

$$a_{\text{df}}^{\text{gas}} = \frac{4\pi G^2 M_{\text{BH}} \rho_{\text{gas}}(r-r_0)}{v_{\text{BH}}^2} f(M),$$

where $\rho_{\text{gas}}(r-r_0)$ is the mass density of the hot gas at position $r-r_0$, $\ln \Lambda$ is again the Coulomb logarithm fixed to 3.1 and, $M$ the Mach number equal to $|v_{\text{BH}}|/c_s$ with $c_s$ the sound speed computed as in Tanaka & Haiman (2009) and Choksi et al. (2017): $c_s \approx (1+z)^{1/2} (M_{\text{halo}}/10^7 M_{\odot})^{1/2} (\Omega_M h^2/0.14)$ km/s.

The exact value of $f(M)$ is determined by:

$$f(M) = \begin{cases} \frac{1}{2} \ln \Lambda - \frac{\sqrt{\pi}}{2} M e^{-M^2/2} & \text{if } M \leq 0.8 \\
\frac{1}{2} \ln \Lambda - \frac{\sqrt{\pi}}{2} M e^{-M^2/2} & \text{if } 0.8 \leq M \leq 1.5 \\
\frac{1}{2} \ln (1 - M^{-2}) + \ln \Lambda & \text{if } M > 1.5 \end{cases}$$

The integration of the wBH orbit stops when i) the black hole is re-incorporated in the galaxy (see Section 3.5.4), ii) the recoil velocity of ejected BHs is larger than the halo escape velocity or iii) the black hole position exceed $3 \times R_{200}$ ($R_{200}$ is the halo virial radius) and it is still moving away from the galaxy.

### 3.5.4 Recomposition of wBHs

We assume that a BH is reincorporated in a galaxy when it passes thought the galactic center, $r_0$, with a velocity ($v_{\text{BH}}$) smaller than the galaxy escape velocity ($v_{\text{esc}}$), i.e. $v_{\text{BH}} < v_{\text{esc}}$. The exact value of $v_{\text{esc}}$ is computed taking into account both bulge and disk (stellar and cold gas):

$$v_{\text{esc}} = \sqrt{v_{\text{esc,Bulge}}^2 + v_{\text{esc,Stellar}}^2 + v_{\text{esc, Cold disk}}^2}.$$  

In the case of the bulge component we assume a Hernquist profile (Hernquist 1990) with an escape velocity at $r_0$:

$$v_{\text{esc,Bulge}}^2 = 2 \left(1 + \sqrt{2}\right) \frac{GM_{\text{Bulge}}}{r_{\text{Bulge}}},$$

where $M_{\text{Bulge}}$ is the total bulge mass and $r_{\text{Bulge}}$ the bulge half-mass radius. Although, this profile can approximate classical bulges, it is a rough approximation for pseudobulges.

For stellar and cold gas disk we assume an exponential profile. In this case, the escape velocity at $r_0$ is determined by:

$$v_{\text{esc, Stellar/Cold disk}}^2 = 3.36 \frac{GM_{\text{Disk}}}{r_{\text{disk}}}.$$  

Note that the sub-steps that L-Galaxies does between DM snapshots are $\sim 2-20\,\text{Myr}$, depending on redshift.

\footnote{Here we assume that the galaxy position is always at the center of the host halo, i.e. $r_0$}

\footnote{Note that the sub-steps that L-Galaxies does between DM snapshots are $\sim 2-20\,\text{Myr}$, depending on redshift.

\footnote{The concentration parameter as a function of redshift and halo mass has been computed by using the fits of Dutton & Macciò (2014).}
where $M_{\text{disk}}$ is the total stellar/cold gas disk mass and $r_{\text{disk}}$ its half-mass radius.

### 3.5.5 Changing reference system during merger events

Even though the previous formalism describes reasonably well the orbital evolution of wBHs in galaxies and halos evolving in isolation, the scenario complicates when halos merge. During the interaction, the satellite galaxy is deprived of its host DM halo because of either merger or DM stripping. In the following, we will refer to the halo merger as the instant in which the satellite halo is no more identified within the underlying merger tree. As a consequence of the satellite DM halo loss, the motion of its associated wBHs gets modified, starting to oscillate according to the potential generated by the central galaxy and DM halo\(^8\). Therefore, the position of the recently accreted wBHs has to be referred to the central halo (galaxy): $(x_{\text{BH}}, y_{\text{BH}}, z_{\text{BH}})$. Nevertheless, since we only follow the radial coordinate $r$ of the wandering BH (see Eq. 32), they can be placed at any angular position $(\theta, \varphi)$ in a sphere of radius $r$. We draw values for $\theta$ and $\varphi$ from a random distribution within a sphere. Once $r$, $\theta$ and $\varphi$ are fixed, we can determine the Cartesian coordinates of the wBHs, $(x_{\text{wBH}}, y_{\text{wBH}}, z_{\text{wBH}})$:

\[
\begin{align*}
  x_{\text{wBH}} &= x_{\text{BH}} + r \sin \theta \cos \varphi, \\
  y_{\text{wBH}} &= y_{\text{BH}} + r \sin \theta \sin \varphi, \\
  z_{\text{wBH}} &= z_{\text{BH}} + r \cos \theta,
\end{align*}
\]

(39)

where $(x_{\text{BH}}, y_{\text{BH}}, z_{\text{BH}})$ is the positions of the satellite galaxy at the moment of merger. Therefore, the radial coordinate of the wBHs within the new host halo, $r_{\text{new}}$, is:

\[
r_{\text{new}} = \left[ (x_{\text{wBH}} - x_{\text{BH}})^2 + (y_{\text{wBH}} - y_{\text{BH}})^2 + (z_{\text{wBH}} - z_{\text{BH}})^2 \right]^{1/2} / \hat{r}.
\]

(40)

Finally, the new initial velocity condition for the just acquired wBHs is set as the maximum between the velocity of the BH at the moment of the merger and the galaxy infall-velocity.

### 4 NUCLEAR BLACK HOLES

In this section we compare the predictions for nuclear black holes (i.e BHs located at the center of the host galaxy at the analyzed redshift) with a variety of available observational results. By checking the black hole mass function, the bolo- metric luminosity functions and spin values we prove that our black holes form a reliable population at $z \leq 4$. We do not present results beyond $z = 4$ since the model predictions are very sensitive to the exact seed mass assumed. In a future work, where a more careful seeding in L-Galaxies is modeled, we will analyze the high-$z$ black hole population (Spinoso et al. in prep.). On top of this, we show only the results for BHs whose mass is larger than $10^6 M_\odot$, as at smaller masses the resolution of Millennium DM simulation does not allow us to draw reliable conclusions.

#### 4.1 Black hole mass assembly

The evolution of the black hole mass function (BHMF) between $z = 0$ and $z = 4$ is shown in the upper panel of Fig.3. There is a significant evolution form $z \sim 4$ up to $z \sim 2$, where the BHMF amplitude increases up to $2 \text{dex}$. Conversely, at $z < 1$ the BHMF does not evolve significantly. Such small evolution at low-$z$ indicates that most of the local BHs were already assembled by $z \sim 1$. Similar behavior is found in the theoretical work of Merloni & Heinz (2008) where a very weak evolution in the BHMF is seen below $z \sim 1$ and, as we find here, most of it happens at small masses ($\lesssim 10^7 M_\odot$). On the semi-analytical side, works such as Fanidakis et al. (2011); Griffin et al. (2018) or Marshall et al. (2019) show a stronger redshift evolution at $z \leq 1$ where the normalization of the BHMF increases relatively fast at $> 10^6 M_\odot$ and decreases at $\lesssim 10^8 M_\odot$. On the other hand, the results of Hirschmann et al. (2012) show a similar weak evolution in the BHMF at $z \sim 1$. In the same figure, we have compared

---

\(^8\) The orbits of the wBHs moving around the central halo (galaxy) are also affected by the change of the dark matter potential, see Eq (32).
the predictions at $z = 0$ with the observational constrains provided by Marconi et al. (2004) and Shankar et al. (2004, 2009). Even though the BHMF of the model is compatible with the observations, mainly with Marconi et al. (2004), it displays a smaller amplitude when compared with Shankar et al. (2004, 2009). As pointed out by Shankar et al. (2016), this discrepancy might be caused by biases effects affecting the observations. Shankar et al. (2016) showed that, because of selection effects, the normalization of the scaling relations used to link the BH mass with galaxy properties (bulge velocity dispersion, total stellar mass and bulge mass) is increased by a factor $\sim 3$ (see also Bernardi et al. 2007; Shankar et al. 2019). The lower amplitude in the empirical relations pointed out by Shankar et al. (2019) yields smaller BH masses, lower black hole mass density and higher radiative efficiencies which would support rapidly spinning BHs and lower levels of gravitational wave emission, consistent with the current non-detection of this signal by pulsar timing array experiments (see Sesana et al. 2016).

The lower panel of Fig. 3 presents the BHMF of active and inactive BHs at $z \sim 0$ and 3. We consider as threshold between active and non-active the Eddington ratio $f_{\text{Edd}} \sim 10^{-4}$ (Rosas-Guevara et al. 2016), i.e. below this value the emission of the BH is essentially undetectable against the emission of the host galaxy. Even though this threshold is quite arbitrary, we did not find significant differences in the results assuming a threshold between $10^{-3} - 10^{-5}$. The inactive population increases form $z \sim 3$ to $z \sim 0$, with the most massive BHs becoming non-active at an earlier epoch: while at $z \sim 3$ inactive BHs have typical $M_{\text{BH}} > 10^{8} M_{\odot}$, at $z \sim 0$, black holes with masses $M_{\text{BH}} > 10^{8} M_{\odot}$ are inactive. This behaviour is usually referred to as downsizing, and it has been reported by observational works (Merloni et al. 2003; Heavens et al. 2004; Hasinger et al. 2005), semi-analytical models (Bonoli et al. 2009; Fanidakis et al. 2012; Hirschmann et al. 2012) and cosmological hydro-simulations (Stiavelli et al. 2009; Rosas-Guevara et al. 2016; Thomas et al. 2019). In the same figure we divided the active population by accretion geometry. We assume that the emission of active BHs with $f_{\text{Edd}} > f_{\text{Edd}}^\text{crit}$ can be described by the thin and radiatively-efficient Shakura-Sunyaev disk model (Shakura & Sunyaev 1973, thin disk geometry, hereafter). On the other hand, the emission of active BHs with $10^{-4} < f_{\text{Edd}} \leq f_{\text{Edd}}^\text{crit}$ are associated to a thick and radiatively-inefficient accretion disk to which we will refer as a Advection Dominated Accretion Flow geometry (or just ADAF geometry, Rees et al. 1982; Narayan & Yi 1994). As shown, regardless of the BH mass, $z \sim 3$ active BHs are characterized by a thin disk geometry fuelled by Eddington limited accretion flows (see Fig. 4). This picture changes at $z \sim 0$, where ADAF becomes the main accretion geometry for BHs with $M_{\text{BH}} > 10^{9} M_{\odot}$. In this case, BHs are powered by the quiescent phase of cold gas accretion and the consumption of hot gas which surrounds its host galaxy (see right panel of Fig. 4). We refer the reader to Section 3.2, where a detailed description of how $f_{\text{Edd}}$ values are computed in our BH growth model.

Finally, we have also explored the evolution of the black hole masses in the chaotic scenario proposed by King et al. (2005)\(^{10}\). For the sake of the present discussion, we only compare the main predictions of this chaotic model to our findings. We have seen that, as a consequence of the different outcomes in the spin distribution, there is a faster BH assembly in the chaotic model than in the one explored in this work. Whereas the former display typically spin values of $a \gtrsim 0.1$, the latter is characterised by $a \gtrsim 0.7$. As shown by Eq. (23), the spin has an impact in the mass-to-energy conversion, implying longer timescales in the BH growth at larger spin parameters (see Eq. (11)). We have also checked that the fast growth undergone by the BHs in the chaotic scenario has an imprint in the faint end of $z \lesssim 1$ bolometric luminosity functions (LFs). Since the BH growth in the chaotic scenario happens on a shorter time-scale than in the model presented here, the BHs tend to consume the gas reservoir faster, hence entering to the quiescent phase earlier. This implies that BHs at low-$z$ do not have enough gas in order to effectively contribute to the faint-end of the bolometric luminosity function, reducing its amplitude. In order to contrast this effect and match the observed faint-end of the AGN bolometric LF, the parameters $f_{\text{merge}}^\text{BH}$ and $f_{\text{BH}}^\text{DI}$ (Eq. (4)-(5)) controlling the amount of gas accreted during mergers and DI, had to be increased. This resulted in a significant excess on the predicted BHMF at $M_{\text{BH}} > 10^{8} M_{\odot}$, making difficult its comparison to observations.

4.2 The evolution of bolometric luminosity

Since the discovery of the first quasar (Schmidt 1963), it has been widely accepted that the quasar phase is triggered by the gas accretion onto BHs. The understanding of the quasar luminosity function at different redshifts is a crucial point for inferring the assembly history of BHs. They provide us with information about the BH growth rate, the nature of accretion disks and fundamental quantities such as BH spins and radiative efficiencies. In the left panel of Fig. 4 we present the evolution of the bolometric luminosity ($L_{\text{bol}}$) function predicted by our model (in Appendix B it is presented the same but for hard and soft X-rays luminosity). The predictions of our model are compared to the observational work of Hopkins et al. (2007); Aird et al. (2015) and Ueda et al. (2014). A good agreement is achieved at any redshift bin. However, at $z < 0.5$ our model slightly underestimates the number of objects in the range of $10^{45} \lesssim L_{\text{bol}} \lesssim 10^{47}$ erg/s. This might be caused by other feeding processes which are not accounted by our model. For instance, Volonteri et al. (2013) discussed that substantial growth of BHs in ellipticals galaxies was due to the consumption of recycled gas of the evolving stellar population. In future works, we plan to include this extra growth channel in both elliptical and spiral galaxies.

9 For instance, they reported that local galaxies with black hole mass estimate with a dynamical approach are a biased subset of all galaxies, i.e. at fixed stellar mass BHs display larger velocity dispersion that the bulk of the population, regardless of the exact morphological type.

10 Following Sesana et al. (2014), the chaotic scenario presented in King et al. (2005) is obtained fixing $F = 1/2$ in Eq. (20).
the same figure, we have divided the luminosity functions in three types of bulge population: galaxies hosting a pseudobulge, classical bulge and elliptical structure. While at high-$z$ ($z>2$) the bolometric luminosity function is dominated by BHs accreting in pseudobulge structures, at lower redshifts classical bulges and elliptical galaxies are the main structures hosting active BHs. At $z<1$, pseudobulges host only faint AGNs ($L_{\text{bol}}\lesssim10^{44}$ erg/s). Since pseudobulges form in galaxies experiencing a quiet merger history (see Izquierdo-Villalba et al. 2019), our model points out that high-$z$ AGNs ($z\gtrsim2$) and intermediate luminous AGNs ($L_{\text{bol}}\sim10^{44-45}$ erg/s) at $z\lesssim1$ are mainly triggered by secular processes rather than galaxy encounters. This is in agreement with recent observational and theoretical results. For instance, the observational work of Allevato et al. (2011) pointed out that major or minor mergers alone are not able to reproduce the high bias factors of X-ray AGN selected sample. Similar results were obtained in Marian et al. (2019) for X-ray and optical selected AGNs at $z\sim2$. On the theoretical side, Martin et al. (2018), by using the Horizon-AGN hydro-dynamical simulation, found out that only 35% of today's BH mass is directly attributable to merger-driven gas accretion. Similar results were found by Steinborn et al. (2018), which using large-scale cosmological hydrodynamic simulations from the Magneticum Pathfinder set, concluded that mergers could not be the statistically prevalent fueling mechanism for nuclear activity at $z\sim0-2$, except for very luminous AGNs, with $L_{\text{bol}}>10^{46}$ erg/s. From the semi-analytical models perspective, recent works of Lagos et al. (2008); Fanidakis et al. (2011); Griffin et al. (2018) and Marshall et al. (2019) found that, regardless of redshift, DIs are the main drivers of BH growth and AGN activity. Even though all these works use the same analytical prescription to detect instabilities in the galactic disk (see Eq. 1), their implementation is slightly different from the one used here. For instance, the model of Fanidakis et al. (2011) (and Griffin et al. 2018) assumes that any disk instability event is able to destroy the galactic disk, transforming the galaxy into a pure spheroidal structure and triggering a burst of star formation. On the other hand, the works of Lagos et al. (2008) and Marshall et al. (2019) follows the approach presented here transferring to the galaxy spheroidal component only the amount of stellar mass needed to restore the disk stability. However, they do not distinguish between their triggering mechanism, as we do here.
To directly explore the relation between bolometric luminosity and BH mass, in the right panel of Fig.4 we show the black hole mass-bolometric luminosity plane at six different redshifts. At $z \geq 3$ there is a tight correlation between black hole mass and luminosity, as almost all BHs are accreting at $f_{\text{edd}} = 1$. On the contrary, at $z < 2$ most BHs are far from growing at the Eddington limit, and diverse combinations between BH mass and $f_{\text{edd}}$ are present. For instance, whereas at $z \sim 4$ all the quasars shining at $L_{\text{bol}} \sim 10^{45}$ erg/s are triggered by BHs of mass $M_{\text{BH}} \sim 10^6 M_{\odot}$ with $f_{\text{edd}} = 1$, at $z \lesssim 3$ the same luminosity is triggered by a wide range of BHs ($M_{\text{BH}} \sim 10^7 - 10^8 M_{\odot}$) accreting at $f_{\text{edd}} \sim 1 - 0.01$. Only the highest luminosity ($L_{\text{bol}} \gtrsim 10^{46}$ erg/s) are reached always by the same type of objects: BHs of mass $M_{\text{BH}} \gtrsim 10^8 M_{\odot}$ accreting at the Eddington limit. Interestingly, the plane $L_{\text{bol}} - M_{\text{BH}}$ displays two branches: one at $f_{\text{edd}} \gtrsim 10^{-3}$ and the other at $f_{\text{edd}} \lesssim 10^{-3}$. While the former is caused by cold gas accretion, the latter is originated by hot gas consumption. At $z > 3$ these two regimes are clearly separated, whereas at $z \approx 2$ the branches start to blend. This is because BHs accreting via the cold gas phase enter in the quiescent (or self-regulated) phase, characterized by low $f_{\text{edd}}$. Eventually, as we discussed in Fig.4 left panel, at $z = 0$ the number density of active BHs is dominated by low $f_{\text{edd}}$ values (or ADAF geometries) caused by both hot and cold gas accretion.

4.3 The evolution of black hole spin

Apart from mass, the spin parameter is the other fundamental property of BHs that has to be taken into account to build a complete picture of how black holes assemble their masses across cosmic time, given also the dependence of the radiation efficiency on the spin value. From an observational point of view, spin measurements are still a challenge and the current estimates display large errors. Reynolds (2013) present a compilation of few BHs with reliable measurements of both spin and mass. However, all the spin values have been computed via X-ray spectroscopy of iron $K\alpha$ line, biasing the results towards AGNs with bright hard X-rays luminosity ($> 10^{42}$ erg/s, see Table 2 in Sesana et al. 2014). As a check of the spin predictions in L-Galaxies, we present in Fig.5 a comparison of the model with Reynolds (2013) data. In order to perform a fair comparison we have only selected in our sample the BHs with hard X-ray luminosity larger than $10^{42}$ erg/s (see Appendix B for hard and soft X-rays LFs). We see that the model predictions are overall consistent with the observations.

In the left panel of Fig.6 we present the cosmological evolution of BH spin. The model generally predicts a rapidly spinning super-massive black hole population. While at $M_{\text{BH}} < 10^6 M_{\odot}$ BHs tend to be maximally spinning, at $M_{\text{BH}} > 10^6 M_{\odot}$ BHs display lower spin values (see also Sesana et al. 2014), with average spin values decreasing with increasing BH mass. Also, the median spin values show a modest decrease with decreasing redshift, but only for BHs with $M_{\text{BH}} > 10^7 M_{\odot}$. The different spins of small and high-mass BHs come from the ratio $|\bar{J}_{\text{bol}}|/2|\bar{J}_{\text{BH}}|$ during the gas consumption. At $M_{\text{BH}} < 10^6 M_{\odot}$, independently of the redshift and the nature of the accretion episode, the ratio $|\bar{J}_{\text{bol}}|/2|\bar{J}_{\text{BH}}|$ is always larger than one, resulting in a BH feeding characterized by $n_{\text{p}} a = 1$ (see Section 3.3).

Conversely, for $M_{\text{BH}} > 10^6 M_{\odot}$, the value of $|\bar{J}_{\text{bol}}|/2|\bar{J}_{\text{BH}}|$ is not necessary larger than one and the precise number of $n_{\text{p}} a$ depends on the galaxy bulge assembly. Indeed, these bulge dependence is seen in Fig.6 when the BH population is divided according to the host bulge type (pseudobulge, classical bulge and elliptical). Regardless of the redshift, at fixed mass, BHs hosted in pseudobulges display larger spin values ($a \gtrsim 0.9$) than the ones in classical bulges ($a \approx 0.7$) and ellipticals structures ($a \lesssim 0.4$). This is in agreement with the work of Orban de Xivry et al. (2011), which concluded, by observing local AGN, that BHs in local Narrow-line Seyfert 1 galaxies (with masses $\sim 10^6 M_{\odot}$) hosted in pseudobulge structures need to be rapidly spinning in order to explain their duty cycles. Notice that the model predicts a slightly spin-up of BHs hosted in elliptical galaxies form $z \approx 1$ ($a \approx 0.3$) to $z \approx 0$ ($a \approx 0.4$). This is because after the formation of the elliptical structure the galaxy undergoes successively minor mergers/smooth accretions (a median value of 5) which carry gas with large $v/\sigma$ towards the BH, ultimately causing its spin-up (see Figure 7 and 16 in Izquierdo-Villalba et al. 2019 where we show the number density of minor/smooth accretion events and the typical redshift of the last minor merger). This is along the line of recent observational work, pointing out that minor mergers ($\sim 8$ with mass ratio 1:10) contribute significantly to the build-up of massive elliptical galaxies at $z \lesssim 1$ (Boylan-Kolchin et al. 2006; Trujillo et al. 2011). The dependence on the bulge type is somehow blurred for BHs with $> 10^7 M_{\odot}$ inhabiting pseudo- and classical- bulge structures. This is because such massive BHs are hosted in massive galaxies ($M_{\text{stellar}} > 10^{11} M_{\odot}$) whose bulge assembly is complicated, shaped by both mergers and disk instabilities. This intricate bulge assembly results in a complex evolution of the bulge $v/\sigma$, significantly deviating form the $v/\sigma$ values produced by DI or minor mergers alone. For instance, the bulge

![Figure 5](image_url)
During the last years several studies have addressed the evolution of BH spin by using semi-analytical models. For instance, Fanidakis et al. (2011) and Griffin et al. (2018) found out that the chaotic scenario suggested by King et al. (2005) applied on GALFORM SAM (run on top of the Millennium and P-Millennium merger trees) yields a population of BHs characterized by low spin values ($a \lesssim 0.4$) at any mass and redshift (see Figure 9 of Fanidakis et al. 2011 and Griffin et al. 2018). Even though the chaotic scenario was able to reproduce the population of radio-loud AGNs in the local Universe, its outcome spin distribution makes difficult reconciling the model predictions with the recent claims of BHs with $a > 0.9$, the observed LFs are systematically above the observed one and its consistency with radio-loud AGNs in the local Universe is worse than in the chaotic scenario (see Figure 10 of Fanidakis et al. 2011 and Figure 16 of Griffin et al. 2018). Other attempts have been done by Barausse (2012) employing a SAM build on top of Press-Schechter merger trees. By using a more sophisticated approach consisting in feeding the BH respectively in a prograde and chaotic way during gas rich and gas poor mergers, they found a dichotomy in the spin distribution between rapidly ($a \gtrsim 0.9$) and low ($a \sim 0.2$) spinning BHs (see Figure 14 and 15 in Barausse 2012). Sesana et al. (2014) updated the spin evolution of Barausse (2012) by linking the BH growth with the assembly of the galactic bulge. As in this work, they found different spinning black holes hosted in elliptical and spiral galaxies. Finally, Volonteri et al. (2013) also studied the spin with a SAM based on Press-Schechter merger trees. Similar what we do here, they assumed that the BH quasar phase takes place in two different phases. Whereas the former is described by an Eddington limited growth with a spin evolution characterized by a prograde accretion, the latter displays low Eddington ratios with a chaotic spin evolution (even though they explored a coherent mode as well during this quiescent phase). They found a strong spin redshift evolution whereby $z > 2$ BHs are characterized by large $v/\sigma$ in pseudobulges decreases after a minor merger while $v/\sigma$ in classical bulges increases after a DI. Other mechanisms that blur the spin-bulge dependence at $0 < z < 1$ are the gas-poor BH-BH coalescences. In these cases the remnant BHs do not accrete gas and the spin change is driven by the BH coalescence. Berti & Volonteri (2008) showed that the spin evolution under only BH-BH mergers is driven towards $a \sim 0.4 - 0.7$ (with some intrinsic dispersion).

**Figure 6. Left panel:** Predicted black hole spins, $a$, as a function of black hole mass ($M_{BH}$) and at different reshifts. Solid black lines represent the median value of the spin per black hole mass. Dark and light orange shaded areas display the $1\sigma$ and $2\sigma$ of the distribution. To guide the reader we have highlighted with dashed lines the value of $a = 0.6$ (corresponding to an accretion efficiency $\sim 0.1$). **Right panel:** The same as in the left panel but dividing the galaxies between classical bulges (green), pseudobulges (blue) and ellipticals (red). Shaded areas display the $1\sigma$ of the distribution. The large fluctuations at high masses are due to small sample statistics (in both panels dashed lines in the median relation corresponds to the black hole masses where the number of objects is smaller than 5).
spin values ($a > 0.8$) whereas $z < 1$ BHs display small spins ($a < 0.4$). However, the model did not include BH feeding through disk instabilities, a channel which seems to be fundamental in SAMs (Fanidakis et al. 2012; Menci et al. 2014; Marshall et al. 2019).

5 WANDERING BLACK HOLES

In this section we discuss our main results on the wandering black hole population. We first explore the frequency and location of the events at different cosmological times. We then focus on the local universe, investigating the characteristics of wBHs hosted in different kinds of galaxies.

5.1 Wandering black holes across cosmic time

The left panel of Fig. 7 shows the redshift evolution of the wBH number density ($n$). We show only results for BHs whose mass is larger than $10^8 M_\odot$ since at smaller masses we can not draw solid conclusions given the resolution of the MS simulation, as previously discussed. Regardless of the BH mass, the number density evolves with redshift, reaching a maximum of $\sim 10^{-4} Mpc^{-3}$ at $z \sim 0$. The lighter is the mass of the wBH the larger is the number density at high-$z$. For instance, while at $z > 4$ black holes of $M_{BH} > 10^7 M_\odot$ have $n \lesssim 10^{-7} Mpc^{-3}$, black holes of $10^5 < M_{BH} < 10^7 M_\odot$ present 1 dex larger number densities. For comparison, the figure displays the $n$ values of nuclear black holes. Regardless of mass and redshift, nuclear BHs are $1-3$ dex more numerous than wandering BHs, especially at $z \gtrsim 4$. Concerning the different contribution between orphan and ejected BHs, we find that at $M_{BH} > 10^7 M_\odot$ the former dominate the number density of wBHs. On the other hand, at $10^5 < M_{BH} < 10^7 M_\odot$ the relative contribution of the two types of wBHs evolves with redshift; while ejected wBHs display larger $n$ values at $z \gtrsim 1.5$, orphan ones dominate at lower redshifts.

In the two central panels of Fig. 7 the evolution of the median halo and stellar mass in which wandering black holes are hosted is presented. At $z > 3$ wBHs with $M_{BH} > 10^8 M_\odot$ inhabit very massive halos and galaxies, with
Figure 8. **Left panel:** Connection between the wandering black holes and their host halos at $z=0$. **Left upper panel:** Probability $P$ at a given halo mass bin of finding at least one wBH above the mass threshold, $\text{M}_{\text{BH}}$. Solid black line, short dashed orange line and long dashed purple line display three different $\text{M}_{\text{BH}}$ thresholds: $\text{M}_{\text{BH}}=10^5 \text{ } \text{M}_\odot$, $10^6 \text{ } \text{M}_\odot$, and $10^8 \text{ } \text{M}_\odot$, respectively. Horizontal dashed lines highlight the values $P=0.1, 0.3, 0.9$ the **Left lower panel:** Median number of wandering black holes, $N_{\text{wBH}}$, per halo mass bin for a given $\text{M}_{\text{BH}}$ value. Shaded areas represents the 3σ interval of the distributions. **Right panels:** Wandering black hole mass function at three different halo mass bins. Solid blue, short red dashed and long blue dashed lines represent, respectively, the mass function for nuclear BHs, orphan wBHs and ejected wBHs. **Right panel:** Same as in the left panel but for stellar masses.

The position of both ejected and orphans wBHs with $\text{M}_{\text{BH}}>10^5 \text{ } \text{M}_\odot$ as function of redshift is presented in the right panel of Fig.7. The population has been divided in 3 different bins of halo masses while the distances (referred to the halo center, $r_0$) have been normalized by the halo virial radius, $R_{200}$. As shown, regardless of halo mass, ejected and orphan wBHs at $z>0.8$ are located in different regions inside the halo. While the former reside at $\lesssim 0.1 R_{200}$ (with a decreasing trend towards low-$z$), the latter orbit at $-0.4 - 0.6 R_{200}$ radii. On the contrary, at $z<0.8$ there are differences from a halo mass bin to another. While for $\text{M}_{\text{halo}}<10^{13} \text{ } \text{M}_\odot$ ejected and orphans wBHs still inhabit different halo regions, at $\text{M}_{\text{halo}}>10^{13} \text{ } \text{M}_\odot$ both types of wBHs are at similar distances from the halo center, $\sim 0.3 R_{200}$. To understand this behavior, in Fig.7 we divided the population of ejected wBH in two sub-classes: the acquired and inborn ones. While the former have been incorporated from other galaxies after an halo merger, the latter were generated in-situ, expelled form the central galaxy of the halo after a recoil. A completely different behavior is seen for the two populations, regardless of the redshift. While inborn ejected BHs are closer to the halo center than orphan BHs, the acquired ejected wBHs populate similar regions. Consequently, the change of trend of the ejected population in massive halos at low-$z$ it is caused by the acquired ejected wBHs, whose number becomes dominant with respect to the inborn ones.

5.2 The environment of wandering black holes in the local universe

We now focus on the local universe, exploring the typical halo and galaxy masses hosting wandering black holes as well as the frequency of wandering BHs for different galaxy type. As discussed before, we assume that wBHs are not accreting, thus not observable as active sources. In future works we will explore how simple assumptions for the growth of wBHs could translate into their observability for different host properties.

The mass function of wBHs at different halo and galaxy masses is presented in the the small panels of Fig.8. For what concerns the halo mass, we note that only the more massive halos host a significant population of wBHs, including some with mass larger than $10^8 \text{ } \text{M}_\odot$. By dividing the wandering BHs by formation scenario we can see that for $M_{\text{halo}} \lesssim 10^{13} \text{ } \text{M}_\odot$ the ejected population reaches larger masses than the orphan one, whereas at larger halo masses
the trend is inverted. Indeed, the amplitude of the orphan wBHs mass function increases with halo mass due to the increase of the average number of satellite galaxies. Analogously, galaxies with the largest masses are the ones which host the most massive wandering black holes, but already galaxies with $M_{\text{stellar}} \sim 10^{10} M_\odot$ can be surmounted by massive wBHs.

The larger panels of Fig. 8 present both the probability $P$ of finding at least one wBH above a given mass threshold $M_{\text{BH}}$ (upper panel) and the median number of wBHs ($N_{\text{wBH}}$) at a given halo mass (lower panel). Independently of the $M_{\text{BH}}$, both $P$ and $N_{\text{wBH}}$ increase with the host halo and galaxy mass. However, the exact value of $M_{\text{BH}}$ has an important effect in the values of $P$ and $N_{\text{wBH}}$. For instance, at $M_{\text{BH}} = M_{\text{seed}}$ all halos above $10^{13} M_\odot$ (galaxies above $10^{10.75} M_\odot$) host at least one wBH, with a typical $N_{\text{wBH}} > 5$. However, only halos above $10^{14} M_\odot$ (galaxies above $10^{11.5} M_\odot$) host a wBH with $M_{\text{BH}} > 10^6 M_\odot$.

6 THE IMPRINT OF GRAVITATIONAL RECOIL ON THE BH GLOBAL PROPERTIES

In this section we investigate the effects of gravitational recoil in the properties of nuclear black holes across cosmic time. As we will show, long wandering phases can have a visible effect in the black hole occupation fraction and in the BH growth, in particular for certain galaxy types.

6.1 Black hole occupation fraction

The seeding procedure followed in this works (see Section 3.1) imply that all galaxies host a BH when they are initialized. But gravitational recoil after BH mergers can leave galaxies without a central BH, at least until the kicked BH looses the acquired energy and settles again in the galaxy center. Therefore, especially at later times, a fraction of galaxies is expected to be deprived of a central BH. In Figure 9 we show the nuclear BH occupation fraction as a function of galaxy stellar mass, and for different galaxy morphologies. Regardless of the bulge morphology and redshift, $f_{\text{BH}}$ decreases towards low stellar masses: from $\sim 80\%$ for $M_{\text{stellar}} \geq 10^{11} M_\odot$ down to $\sim 60\%$ for galaxies with $M_{\text{stellar}} \lesssim 10^{10} M_\odot$. This mass dependence is related to the potential well of the galaxy: the smaller the galaxy, the lower the escape velocity, thus the higher the ejection probability for the central BH. Moreover, as we will see at the end of this section, for galaxies with $M_{\text{stellar}} < 10^{9} M_\odot$, it is also much more probable the ejection not only from the galaxy, but also from the halo. By dividing the galaxy population in different bulge morphological types, we find that, in massive galaxies with $M_{\text{stellar}} > 10^{10} M_\odot$, pseudobulges and bulgeless galaxies display lower occupation fractions than both classical bulges and ellipticals.

is because pseudobulges and bulgeless galaxies experience very few mergers during their evolution, thus it is more difficult for them to replenish their empty bulge after an ejection. Indeed, as we will see in Section 6.2 (Fig. 12), the reincorporation of BHs in a pseudobulge galaxy is not very common ($\lesssim 10\%$ of probability at any stellar mass) making it even more difficult for such structures to increase their nuclear BHs occupation fraction. Besides mass and morphology, Fig. 9 shows a trend of decreasing $f_{\text{BH}}$ towards low-$z$. This is because BH-BH mergers are characterized by smaller values of $q$ ($\sim 0.01, 0.08, 0.1$ for pseudobulges, classical bulges and ellipticals, respectively) and poorer gas environments towards low-$z$. For instance, at $z \sim 2$, $80\%$ of BH-BH mergers in pseudobulges, classical bulges and elliptical galaxies are considered as wet mergers. On contrary, at $z \sim 1$, $\sim 90\%$, $60\%$, $80\%$ of the BH-BH mergers in pseudobulges, classical bulges and elliptical galaxies are classified as poor gas mergers. These conditions increase the modulus of the kick velocity making it easier for a BH to be ejected (see Eq. (31)).

To explore in which moment the BH recoil velocities had major importance in driving ejections, Fig. 10 shows the comoving number density of BH ejections. Clearly, at $z \sim 1.5$ the number density reaches a maximum, coinciding with the peak of the galaxy-galaxy merger frequency (see Figure 2 in Izquierdo-Villalba et al. 2019). By dividing the ejections in different bins of BH mass we find that the larger is the BH mass, the lower is the redshift of its ejection. While black holes of $< 10^6 M_\odot$ are typical ejected at $z \gtrsim 2.5$, the ones with $> 10^6 M_\odot$ were kicked out at $0.1 < z < 0.6$. This mass segregation is just a growth matter. At very high-$z$ we do not find massive BHs ($> 10^8 M_\odot$) being ejected, since they did not have enough time to accrete gas and grow.

Finally, in Fig. 10 we also show the number density of

---

11 Typically, the average number of satellite galaxies increases with the halo mass in a power-law shape (see e.g. Berlind et al. 2003; Contreras et al. 2017).
reincorporation as compared to the one of ejections. The latter is larger than former by a factor of $2 - 3$ depending on the exact redshift. This is because gravitational recoil would expel more easily a BH from both galaxy and halo, rather than only from the galaxy. In other words, gravitational kicks are likely to overcome at once both the halo and galaxy escape velocities rather than only the galaxy one. This directly reduces the effective number of BHs that can be reincorporated in the parent galaxy one. This directly reduces the effective number of BHs that can be reincorporated in the parent galaxy. For instance, regardless of redshift, we have found that galaxies with $M_{\text{stellar}} < 10^9 M_\odot$, $10^9 < M_{\text{stellar}} < 10^{10} M_\odot$, $10^{10} < M_{\text{stellar}} < 10^{11} M_\odot$ have, respectively, 2.5, 2 and 1.5 times more galaxy+halo ejections than only galaxy ejections. On the other hand, in more massive systems ($M_{\text{stellar}} > 10^{11} M_\odot$) the trend is inverted and the former is 2 times less common than the latter. A similar behavior is seen with halo mass.

### 6.2 Effects on the scaling relations

As we have seen in the previous section, the population of nuclear black holes is affected by the physics of gravitational recoil included in the model. In this section we explore how this translates into predictions for the bulge-black hole $z = 0$ relation.

The left panel of Fig.11 shows the bulge-black hole scaling relation for all galaxies when the SAM is run including or not both BH recoil velocities and wandering phase. Clearly, the scatter and the median relation are different in the two runs. The model displays a trend of hosting less massive BHs at fixed bulge mass in the run in which recoil velocities and wandering phase are taken into account. In the other two panels the population is divided by bulge morphological type: classical bulges and ellipticals (middle panel) and pseudobulges (right panel). While we still see some effects in the classical bulge population, the signatures of BH recoil are more pronounced for the pseudobulge population, where the median relation at $M_{\text{bulge}} \gtrsim 10^9 M_\odot$ changes up to $\sim 2$ dex and the scatter increases up to $4$ dex.

These results clearly show that the inclusion of gravitational recoil physics is important in the predictions for the BH scaling relations, in particular in pseudobulge galaxies. Similar effects were found in the hydro-simulations of Blecha et al. (2011) and the analytical model of Gerosa & Sesana (2015). By running over 200 simulations of gaseous galaxy mergers with 60 different BH-BH merger configurations Blecha et al. (2011) found out that the inclusion of gravitational recoils leaves an imprint in the correlation between bulge velocity dispersion and the mass of the host BH. At fixed BH mass, the run with gravitational recoils displays large bulge velocity dispersion values. This deviation is caused by the ejection interrupting the BH growth while the galaxy continues its evolution. On the other hand, the work of Gerosa & Sesana (2015) explored the repercussion of gravitational recoils in brightest cluster galaxies (BCGs). They pointed out that the ejections and subsequent replenishments of BHs after galaxy mergers increase the scatter of the BCGs black hole - bulge mass relation.

In Fig.12 we explore in more details the origin of the deviations of the recoiled population with respect to the median relation. We again present results for classical bulges (left panel) and pseudobulges separately (right panel). The median relation is divided into galaxies whose nuclear BHs never underwent an ejection nor a wandering phase ($G_0$), galaxies whose bulge did not host a central BHs but after a merger\footnote{major, minor or smooth accretion.} it was refilled by one ($G_{\text{refill}}$) and galaxies whose BHs were reincorporated after a wandering phase ($G_{\text{reincop}}$). Regarding pseudobulges, galaxies belonging to $G_0$ follow a very similar relation than the model without recoil velocities (see Fig.11). A similar trend is visible for the galaxies type $G_{\text{refill}}$ at $M_{\text{bulge}} \sim 10^9 M_\odot$. However, their trend breaks at larger bulge masses. Concerning galaxies of type $G_{\text{reincop}}$, they show a much more flattened relation compared to the previous ones, regardless of bulge mass. This is because the merger (principally smooth accretions) that filled the pseudobulge happened at relatively low-$z$, typically $z \lesssim 1.0$, with a galaxy hosting a BH that is $\sim 2$ dex ($q \sim 0.01$) lighter than the average BH that would inhabit the empty bulge. As shown in the upper panel of Fig.12, the relative contribution of $G_0$, $G_{\text{refill}}$ and, $G_{\text{reincop}}$ causes the change on the pseudobulge scaling relation with respect to the one without recoil velocities. Since galaxies $G_0$ are the dominant ones at $M_{\text{bulge}} < 10^{10} M_\odot$ we do not find any differences between the model with and without BH recoils. However, at $M_{\text{bulge}} > 10^{10} M_\odot$ the trend changes since galaxies type $G_{\text{reincop}}$ take the main importance, causing the flattening in the relation, seen in Fig.11.

Concerning classical bulges and ellipticals structures,
all the three types $G_0$, $G_{\text{refill}}$ and, $G_{\text{reincop}}$ have similar bulge-BH mass relation than the relation without recoil (even though $G_{\text{reincop}}$ is systematically below). This is caused by the fact that classical bulges and elliptical galaxies do not spend a lot of time without a BH; the reincorporation time of the BH is relatively low ($\lesssim 0.01\,\text{Gy}$) and since their formation is linked with the hierarchical growth of structures, they experience frequent mergers (with $q \gtrsim 0.2 - 0.3$) in short time-scales. As it is shown in the upper panel of Fig.12, galaxies $G_0$ dominate at any bulge masses. However, at $M_{\text{bulge}} > 10^{10.5}M_\odot$ the probability $P$ of finding a galaxy $G_{\text{refill}}$ increases up to $P \sim 0.4$. This causes the small change seen in the relation with and without recoil velocities in Fig.11. We have also explored the $G_0$, $G_{\text{refill}}$ and, $G_{\text{reincop}}$ population at $z = 2$ for both classical bulges (and ellipticals) and pseudobulges. In this case, at any mass and bulge type the $G_0$ population dominates. For instance, at $M_{\text{bulge}} \sim 10^{10.5}M_\odot$ $G_0$ counts for more than the 65% and 80% of the population of classical bulges and pseudobulges, respectively. This causes the median value in the runs with and without recoil velocities to be very similar, even though the population $G_{\text{refill}}$ and $G_{\text{reincop}}$ helps in increasing the scatter.

Finally, even though we have showed that reincorporated BHs have small impact in the bulge - black hole mass relation, in Fig.13 we have explored the effect of the BH reincorporation after a wandering phase in the scaling relations. In the figure it is shown the scaling relation for only classical bulges (including ellipticals) and pseudobulges whose central black hole was reincorporated after undergoing a wandering phase. Independently of the bulge morphology, there is a segregation by time in the wandering phase ($t_{\text{wand}}$): BHs which underwent longer wandering phase are systematically below the scaling relation, as growth in the model is only allowed when BHs are centrals (see Section 3.5). This trend is more clear in the pseudobulge population.
Figure 13. Bulge – black hole mass relation for classical bulges plus ellipticals (left panel) and pseudobulges (right panel) when only nuclear black holes which underwent a wandering phase are selected. The color map indicates the median value of the time spent in the wandering phase ($t_{\text{wand}}$) per bin of bulge and black hole mass. In each panel, the solid black line represents the scaling relation when no recoil velocities and wandering phase are included, whereas colored lines represent the median relation for nuclear BHs in a given bin of wandering phase, as labeled in figure.

7 SUMMARY AND CONCLUSIONS

In this paper we have studied the mass assembly, spin evolution and environment of both nuclear and wandering supermassive black holes. To do that, we have updated with new physical prescriptions the Izquierdo-Villalba et al. (2019) version of L-Galaxies semi-analytical model (SAM). The new prescriptions have been applied to the merger trees of the Millennium simulation, although galaxies have been initialized using the outputs of the higher-resolution simulation MillenniumII.

The starting point of our BH model is the gas accretion from a hot and cold phase. While the former is linked with the consumption of part of the hot gas atmosphere which surrounds the galaxy, the latter is triggered by cold gas accretion right after a galaxy disk instability (DI) and/or a galaxy merger. During both phases of gas accretion, we track the evolution of BH spin ($a$) using the approach presented in Dotti et al. (2013) and Sesana et al. (2014). We linked the number of prograde transient accretion disk (BH spin-up) with the degree of coherent motion in the bulge. In particular, we have assumed that DIs, which lead to the bar/pseudobulge formation in our model, are the processes that increase the coherence of the bulge kinematics. On the other hand, mergers, which lead to the assembly of classical bulges and ellipticals, bring disorder to the bulge dynamics. These assumptions result in a correlation between the predicted BH spin values and the host galaxy morphology.

Regardless of redshift and BH or bulge mass, galaxies with a pseudobulge structure host nuclear BHs with larger spin values than both classical and elliptical galaxies. In particular, at $M_{\text{BH}} > 10^6 M_\odot$ pseudobulges, classical bulges and elliptical galaxies host respectively BHs with a typical spin of $a \sim 0.9$, 0.7 and 0.4. At lower BH masses, as a consequence the implemented model, all the three bulge morphology host maximum spinning BHs. The model also follows the formation of binary black hole (BBH) systems by assuming that the lifetime of a BBH (from the formation to the coalescence) inversely correlates with the BH merger ratio and the galaxy gas fraction at the moment of the galaxy encounter. This simple assumption will be further improved in future works. After coalescence, the remnant has a spin value fully determined by the Barausse & Rezzolla (2009) analytic expression.

Due to conservation of linear momentum, in the instant of the BH-BH coalescence the propagation of gravitational waves imparts a recoil velocity to the remnant black hole. If the modulus of the recoil velocity, computed according to Lousto et al. (2012) equations, is larger than the escape velocity of the host galaxy, the model assumes that the black hole is kicked from its host and incorporated in the DM halo as a wandering black hole. Here we tag these BHs as ejected wBH. Since it is not clear which is the amount of mass in accretion disk that can be retained or accreted by a BH after the recoil, we have neglected the fact that the ejected BH can retain an accretion disk. The model also tracks the formation of another type of wandering black holes, orphan wBH, which are originated after the complete disruption of their host galaxy via tidal forces. Independently of their origin, wBH orbits are tracked by using numerical integration, taking into account the host halo and galaxy properties for the computation of gravitational acceleration and dynamical friction. In particular, the integration of the wBH orbit stops when i) the black hole is re-incorporated in the galaxy, i.e., when the BH passes through the galaxy center with a velocity smaller than its escape velocity, ii) the recoil velocity of ejected BHs is larger than the halo escape velocity or iii) the black hole position exceed 3 times the hosts halo virial radius and it is still moving away from the galaxy.

Turning to the model predictions, we find a good consistency with the observed local black hole mass function (BHMF), spin values, BH-bulge mass relation and quasar (bolometric, soft and hard X-rays) luminosity functions. The model predicts a stalling in the BHMF at $z \sim 1$, where no significant change up to $z=0$ is shown. This indicates that most of the local BHs were already assembled by $z \sim 1$. By dividing the BHs between inactive and active we have found that the former population increases form $z \sim 3$ to $z \sim 0$, with the most massive BHs becoming non-active earlier. The accretion geometry of the active population displays also a redshift evolution. Regardless of BH mass, $z \sim 3$ active BHs are characterized by a thin disk geometry fuelled by Eddington-limited accretion flows. ADAF, instead, becomes the main accretion geometry at $z \sim 0$ for BHs with $M_{\text{BH}} > 10^6 M_\odot$, powering BHs during the quiescent phase of cold and hot gas accretion. On top of this, the luminosity functions of the model indicate that accreting BHs are hosted in different bulge morphologies at different cosmological times. While at $z > 2$ the high end of the luminosity function is dominated by BHs accreting in pseudobulge structures, at lower redshifts classical bulges and elliptical galaxies are the main structures hosting the most powerful active BHs. Since in the SAM the pseudobulge formation is detached form the merger framework (see Izquierdo-Villalba
et al. 2019) our model points out that $z \gtrsim 2$ AGNs are mainly triggered by secular processes instead of galaxy encounters, in agreement with recent observational and theoretical results.

One of the main novelties of the present study is a thorough exploration of the population of wandering black holes (wBHs), either orphan or ejected. When looking at their mass density, we have found an increasing trend towards low-$z$, reaching a maximum of $10^{-3}$ Mpc$^{-3}$ at $z \sim 0$, which is considerable but still 1–3 dex smaller (depending on the BH mass) than that of nuclear black holes. Regarding the spatial distribution, ejected and orphan wBHs occupy different regions inside the halo. While the former reside at $\lesssim 0.1 R_{200}$, with a decreasing trend towards low-$z$, the latter inhabit the $0.2 - 0.4 R_{200}$ regions. Concerning the environments of wBHs in the local universe, the model predicts that halos of $M_{\text{halo}} < 10^{13} M_\odot$ rarely host a wBH with $M_{\text{BH}} > 10^6 M_\odot$. But if they do, the wBH is typically formed after a gravitational recoil, i.e., of the ejected type. The picture is inverted at $M_{\text{halo}} > 10^{13} M_\odot$, where it is relatively more common to find a wBH of $M_{\text{BH}} > 10^6 M_\odot$ (typical $\sim 5$) principally formed after the disruption of satellite galaxies, i.e. of the orphan type. We find the same tendency when looking at wBH statistics against the host stellar mass.

Besides being an important channel of wandering black hole formation, we have found that gravitational recoils also affect the co-evolution between the black hole and the host galaxy. In particular, they cause a systematic depletion of nuclear BHs towards low redshift and stellar mass. While the former dependence is because low-$z$ BH-BH coalescence are characterized by poorer gas environments than the ones at higher-$z$, condition which increase the modulus of the kick velocity, the latter is simply caused by the fact that low-mass galaxies have smaller potential wells than massive ones, making it more difficult for them to retain the BH after a kick. By dividing the galaxy population according to bulge type, we have found that at $M_{\text{stellar}} \gtrsim 10^{10} M_\odot$ both classical bulge and elliptical structures tend to display larger occupancy fractions than pseudobulges and bulgeless galaxies. This behavior is due to the fact that pseudobulges are detached form the merger framework, making more difficult for them to replenish their empty bulge after an ejection. On the other hand, at $M_{\text{stellar}} \lesssim 10^{10} M_\odot$ classical bulges are more affected by the gravitational recoils, having more than the 40% of empty bulges. By running the SAM with and without gravitational recoils and wandering phase we have also explored the imprints of gravitational recoils in the $z=0$ bulge-BH scaling relation. We have found that both the median of the relation and its scatter are different in the two runs. There is a trend of hosting less massive BHs at fixed bulge mass in the run in which recoil velocities and wandering phase are taken into account. A similar tendency is found when the population is divided into classical bulges, elliptical galaxies and pseudobulges. In particular, the latter structures are the ones which suffer a more pronounced effect, where the median relation at $M_{\text{bulge}} \gtrsim 10^{10} M_\odot$ changes $\sim 2$ dex and the scatter increases. Finally, we have also explored the effect of BH reincorporation after a wandering phase in the BH-bulge relation. Regardless of bulge morphology, the larger is the time of the BH in the wandering phase before its reincorporation, the larger is its offset form the median BH-bulge relation.

Despite the model presented here represents a considerable step forward in modeling the evolution of the population of black holes across cosmic time, still more effort is needed to construct a full coherent framework that can produce new quantitative predictions that can be tested against current and future observations. For example, including a physical model for the expected activity of wBH will lead to useful predictions about the number of active wBHs that may be observed in the local Universe and their more likely location with respect to their host galaxy/group/cluster. In the near future, new instrumentation will make it possible to constrain several aspects of the cosmic BH formation and evolution. The Laser Interferometer Space Antenna (LISA, eLISA Consortium et al. 2013) will constrain the nuclear BH merger rate and their spin properties, thus also providing precious information for the theoretical modeling of the kicked wBH population; James Webb Space Telescope (JWT, Gardner et al. 2006) and Advanced Telescope for High-Energy Astrophysics (ATHENA, Nandra et al. 2013) will allow to pierce deeper in the cosmos, probing progressively further and lower mass accreting BHs. Only a concerted interplay of theory and observations will unfold the full history of the cosmic, nuclear and wandering, supermassive black hole population.

ACKNOWLEDGEMENTS

The authors thank Lucio Mayer for useful discussions and comments. D.I.V. and D.S. acknowledge the support from project AYA2015-66211-C2-2 MINECO/FEDER, UE of the Spanish Ministerio de Economía, Industria y Competitividad and the Aragón Government through the Research Groups E16_17R. D.I.V. particularly thanks the grant Programa Operativo Fondo Social Europeo de Aragón 2014-2020 (Construyendo Europa desde Aragón), A.S. is supported by the ERC through the CoG grant ‘B Massive’, grant number 818691. Y.R.G acknowledges the support of the European Research Council through the grant ERCStG/716151. S.B. and Y.R.G. acknowledge PGC2018-097585-B-C22, MINECO/FEDER, UE of the Spanish Ministerio de Economía, Industria y Competitividad. This work used of CEFCA’s Scientific High Performance Computing system which has been funded by the Governments of Spain and Aragón through the Fondo de Inversiones de Teruel, and the Spanish Ministry of Economy and Competitiveness (MINECO-FEDER, grant AYA2012-30789).

REFERENCES

Aird J., Coil A. L., Georgakakis A., Nandra K., Barro G., Pérez-González P. G., 2015, MNRAS, 451, 1892
Allevato V., et al., 2011, ApJ, 736, 99
Amaro-Seoane P., et al., 2017, arXiv e-prints,
Angulo R. E., White S. D. M., 2010, MNRAS, 405, 143

MNRAS 000, 1–27 (2015)
APPENDIX A: EJECTIONS VIA GRAVITATIONAL RECOIL AND THREE BODY SCATTERING

As we have discussed in Section 3.4 nuclear BHs can be expelled from the galaxy nuclear region through both gravitational recoils and recoils after 3-body scattering. While the former is led by the final BH-BH coalescence, the latter is the consequence of a complex interaction between a binary BH system and an intruder BH which causes the ejection of the less massive BH. In this appendix, we show that the ejections after a gravitational recoil are more common that the ejected less massive BH. In this appendix, we show that the ejections consequence of a complex interaction between a binary BH pelled from the galaxy nuclear region through both gravitational recoils and three body scattering. As shown, at $n_{\text{BH}}$ is fully dominated by gravitational recoiled BHs. To ejection after 3-body scattering is dominated by low mass BHs ($\langle M_{\text{BH}} \rangle > 10^5 M_\odot$) at any redshift and halo mass. On the other hand, the ejection of BHs ($\langle M_{\text{BH}} \rangle > 10^8 M_\odot$) after a 3-body scattering is relatively rare, with less than 300 events at $z < 2$. These events happens mainly in $M_{\text{halo}} > 10^{13} M_\odot$, which are the ones hosting the galaxies with the most massive BHs in the SAM.

Figure A1. Number density ($n$) of nuclear black holes that are ejected from the nuclear part of the galaxy. Each panel represents a bin of halo mass ($M_{\text{halo}}$). While solid lines display the ejections caused by gravitational recoils, dashed ones display the same but for BHs that were ejected via 3-body scattering. Colors represents different bins of BH mass. With shaded areas it is highlighted the value of $n$ in which we have less than 100 (faint grey) and 10 (dark grey) events.

APPENDIX B: X-RAY LUMINOSITY FUNCTIONS

The redshift evolution of X-ray luminosity functions is presented in Fig.B1 and Fig.B2. Soft (0.5–2 keV) and hard (2–10 keV) X-ray luminosity have been computed by using the bolometric corrections derived in Marconi et al. (2004): $\log_{10}(L_{\text{Hx}}/L_{\text{bol}}) = -1.54 - 0.24 L - 0.012 L^2 + 0.0015 L^3$ (B1) $\log_{10}(L_{\text{Sx}}/L_{\text{bol}}) = -1.64 - 0.22 L - 0.012 L^2 + 0.0015 L^3$ (B2) where $L = \log_{10}(L_{\text{bol}}/L_{\odot})$. $L_{\text{Hx}}$ is the hard X-ray luminosity and $L_{\text{Sx}}$ soft X-ray luminosity. Predictions are compared to the observational works of Ueda et al. (2014) Aird et al. (2015) and Buchner et al. (2015). Given the uncertainty in modelling the fraction of obscured AGNs, we prefer to compare the simulations to Aird et al. (2015) observed soft X-ray luminosity functions for which the obscured fraction has already been taken into account. The fraction of obscured objects in the hard X-ray band is thought to be relatively small, so for this work we consider that there is no obscuration at hard X-ray wavelengths. As we can see, good agreement between observations and model is reached. In the same figures we have divided the population between elliptical, classical bulges and pseudobulges. As we found with the bolometric luminosity, at $z > 2$ the luminosity functions are dominated by BHs accreting in pseudobulges structures. On contrary, at $z \lesssim 1$ classical bulges and elliptical galaxies are the structures that preferentially host AGNs and quasars.

This paper has been typeset from a tex file prepared by the author.
Figure B1. X-ray luminosity functions at $z \approx 0.1, 0.5, 1.0, 2.0, 3.0, 4.0$. In solid black line we present the predicted hard X-ray (2-10 keV) luminosity functions. Red dashed-dotted line, green dashed line, and blue dotted line represents the luminosity functions for galaxies hosting, respectively, elliptical, classical bulge, and pseudobulge bulge structure. Each luminosity functions is compared with the compilation of Buchner et al. (2015); Aird et al. (2015); Ueda et al. (2014) (circle, square and, diamond points, respectively).

Figure B2. X-ray luminosity functions at $z \approx 0.1, 0.5, 1.0, 2.0, 3.0, 4.0$. In solid black line we present the predicted soft X-ray (0.5-2 keV) luminosity functions. Each luminosity functions is compared with the compilation of Buchner et al. (2015); Aird et al. (2015); Ueda et al. (2014) (circle, square and, diamond points, respectively).