THE EVOLUTION OF ACTIVE GALACTIC NUCLEI AND THEIR SPINS

M. VOLONTERI1,2, M. SIKORA3, J.-P. LASOTA1,4, AND A. MERLONI3
1 Institut d’Astrophysique de Paris, 98bis Bd. Arago, F-75014 Paris, France
2 Astronomy Department, University of Michigan, 500 Church Street, Ann Arbor, MI 48109, USA
3 Nicolaus Copernicus Astronomical Center, Bartycka 18, 00-716 Warszawa, Poland
4 Astronomical Observatory, Jagiellonian University, ul. Orla 171, 30-244 Kraków, Poland
5 Max-Planck-Institut für Extraterrestrische Physik, Giessenbachstr., D-85741 Garching, Germany

Received 2012 October 3; accepted 2013 July 30; published 2013 September 11

ABSTRACT

Massive black holes (MBHs), in contrast to stellar mass black holes, are expected to substantially change their properties over their lifetime. MBH masses increase by several orders of magnitude over a Hubble time, as illustrated by Sołtan’s argument. MBH spins also must evolve through the series of accretion and mergers events that increase the masses of MBHs. We present a simple model that traces the joint evolution of MBH masses and spins across cosmic time. Our model includes MBH–MBH mergers, merger-driven gas accretion, stochastic fueling of MBHs through molecular cloud capture, and a basic implementation of accretion of recycled gas. This approach aims at improving the modeling of low-redshift MBHs and active galactic nuclei (AGNs), whose properties can be more easily estimated observationally. Despite the simplicity of the model, it does a good job capturing the global evolution of the MBH population from $z \sim 6$ to today. Under our assumptions, we find that the typical spin and radiative efficiency of MBHs decrease with cosmic time because of the increased incidence of stochastic processes in gas-rich galaxies and MBH–MBH mergers in gas-poor galaxies. At $z = 0$, the spin distribution in gas-poor galaxies peaks at spins $0.4–0.8$ and is not strongly mass dependent. MBHs in gas-rich galaxies have a more complex evolution, with low-mass MBHs at low redshift having low spins and spins increasing at larger masses and redshifts. We also find that at $z > 1$ MBH spins are on average the highest in high luminosity AGNs, while at lower redshifts these differences disappear.

Key words: black hole physics – galaxies: active – galaxies: nuclei

Online-only material: color figures

1. INTRODUCTION

Astrophysical black holes (BHs) span a large range of masses, from the remnants of stellar evolution to monsters weighing by themselves almost as much as a dwarf galaxy. Notwithstanding the several orders of magnitude difference between the smallest and the largest BH known, all of them can be described by only two parameters: mass and spin. So, besides their masses, $M$, astrophysical BHs are completely characterized by their dimensionless spin parameter, $a \equiv J/J_{\text{max}} = c J_h/G M^2$, where $J_h$ is the angular momentum of the BH and $0 \leq a \leq 1$.

Many theoretical efforts have focused on the mass growth of massive black holes (MBHs) and on their feedback onto their host galaxies (see, e.g., Haehnelt & Rees 1993; Haiman & Menou 2000; Kauffmann & Haehnelt 2000; Wyithe & Loeb 2003; Volonteri et al. 2003; Hopkins et al. 2006; Croton et al. 2006). Spin has received less attention in the cosmological context (Moderski et al. 1998; Volonteri et al. 2005; Shapiro 2005; Lagos et al. 2009; Fanidakis et al. 2011; Barausse 2012), but it has great relevance for the overall growth of MBHs, as follows.

First, geometrically thin and optically thick accretion disks radiate with an efficiency $\epsilon_{\text{rad}}$ that is almost equal to the mass-to-energy conversion efficiency, $\epsilon_{\text{rad}} \approx \epsilon \equiv 1 - \epsilon_\text{ISCO}$, where $E_{\text{ISCO}} = \sqrt{1 - 2/(3\epsilon_{\text{ISCO}})}$ is the specific energy of the gas particle (in rest mass-energy units) in the innermost stable ($= $ marginally stable) circular orbit (ISCO) and $\rho_{\text{ISCO}}$ is the radius of this orbit in $GM/c^2$ units. This radius, and therefore the radiation efficiency, depend solely on the BH spin parameter $a$. Maximal efficiency ($\epsilon \approx 0.42$) is achievable by disks rotating around maximally spinning BHs; the efficiency drops to $\approx 0.06$ for non-spinning BHs and to $\approx 0.05$ for maximally counter-rotating BHs. This result entails a dependence of the BH mass-growth rate on the spin value, implying the longest growth time scales for larger positive spins. More precisely, for a BH accreting at the Eddington rate, the BH mass increases with time as

$$M(t) = M(0) \exp\left(\frac{1 - \epsilon}{\epsilon} \frac{t}{t_{\text{Edd}}^4}\right),$$

where $t_{\text{Edd}} = M_{\text{BH}} c^2 / L_{\text{Edd}} = (\sigma T c) / (4\pi G M_{\text{BH}}) = 0.45$ Gyr. The higher the spin, the higher $\epsilon$, implying longer timescales to grow the MBH mass by the same number of $e$-foldings.

The radiative efficiency is also the fundamental free parameter in the Sołtan argument (Sołtan 1982) and, more recently, in synthesis models (e.g., Merloni & Heinz 2008), which relate the integrated MBH mass density to the integrated emissivity of the AGN population via the integral of the luminosity function of quasars. If the average efficiency of converting accreted mass into luminosity is $\epsilon = L / M c^2$, then the MBH will increase its mass by $M = (1 - \epsilon) M_{\text{in}}$, accounting for the fraction of the inflowing mass, $M_{\text{in}}$, that is radiated away. Applying this argument to the whole MBH population, the MBH mass density, $\rho_{\text{MBH}}$, can be related to the integral of the luminosity function of quasars, $\Psi(L, z)$, with the radiative efficiency being a free parameter:

$$\rho_{\text{MBH}}(z) = \int \frac{dt}{dz} \int_0^\infty \frac{(1 - \epsilon) L}{\epsilon c^2} \Psi(L, z) dL. \quad (2)$$

Recent results suggest that this approach might be too simplistic, as the radiative efficiency evolves with the cosmic time. Wang...
et al. (2009), for instance, suggest that quasars at the peak of their activity ($z \sim 2$) have high radiative efficiencies and hence high spins. At later times ($z < 1$), however, the average radiative efficiency decreases, hinting at lower spins. This paper addresses these issues by providing spin distributions of MBHs as a function of cosmic epoch.

MBH spins also affect the incidence of MBHs in galaxies via the “gravitational recoil” mechanism. When the members of a BH binary coalesce, the center of mass of the coalescing system recoils due to the non-zero net linear momentum carried away by gravitational waves in the coalescence. If this recoil were sufficiently violent, the merged hole would break loose from the host and leave an empty nest. Recent breakthroughs in numerical relativity have allowed reliable computations of BH mergers and recoil velocities, taking the effects of spin into account. Non-spinning MBHs, or binaries where MBH spins are aligned with the orbital angular momentum, are expected to recoil with velocities below $200 \text{ km s}^{-1}$. The recoil is much larger, up to thousands of $\text{km s}^{-1}$, for MBHs with large spins in non-aligned configurations (Campanelli et al. 2007; González et al. 2007; Herrmann et al. 2007).

Finally, the spin of a BH might determine how much energy is extractable from the BH itself (Blandford & Znajek 1977; Tchekhovskoy et al. 2011; McKinney et al. 2012). The so-called “spin paradigm” asserts that powerful relativistic jets are produced in AGNs with fast rotating BHs (Blandford et al. 1990), implying that MBHs rotate slowly in radio-quiet quasars, which represent the majority of quasars (Wilson & Colbert 1995). Sikora et al. (2007) proposed a “spin-accretion paradigm,” suggesting that the production of powerful relativistic jets is dependent on the presence of fast rotating holes, while it also depends on the accretion rate and on the presence of the disk magneto-hydrodynamical winds required to provide the initial collimation of the central Poynting flux dominated outflow, as in, e.g., the Blandford–Znajek process. Recently, Sikora & Begelman (2013) proposed that the magnetic flux threading the BH, rather than BH spin or the Eddington ratio, is the dominant factor in launching powerful jets.

As described above, MBH spins directly determine the mass-to-energy conversion efficiency of quasars. On the other hand, accretion determines the evolution of MBH spins. A BH that is initially non-rotating gets spun up to a maximally rotating state ($a = 1$) after increasing its mass by a factor $\sqrt{\delta} \approx 2.4$. A maximally rotating hole is spun down by retrograde accretion to $a = 0$ after growing by a factor $\sqrt{3/2} \approx 1.22$. Different modes of MBH feeding imply different spin histories. Spin-up is a natural consequence of prolonged disk-mode accretion: any BH that increases its mass substantially by capturing material with a constant angular momentum axis would end up spinning rapidly (“coherent accretion”). Spin-down occurs when counter-rotating material is accreted, if the angular momentum of the accretion disk is strongly misaligned with respect to the direction of the MBH spin. It has been suggested that accretion may also proceed via small (i.e., the accreted mass is a very small fraction of the MBH mass: $\sim 1\%$ or less) and short uncorrelated episodes (“chaotic accretion”; Moderski & Sikora 1996; King & Fringle 2006), where accretion of co-rotating material (causing spin-up) and counter-rotating material (causing spin-down) is equally probable. As the ISCO for a retrograde orbit is at a larger radius than for a prograde orbit, the transfer of angular momentum is more efficient in the former case. Accretion of counter-rotating material therefore spins MBHs down more efficiently than co-rotating material spins them up. King et al. (2008) considered a MBH evolution scenario in which chaotic accretion very rapidly adjusts the BH’s spin parameter to average values $a \sim 0.1–0.3$ from a broad range of initial conditions, only weakly dependent on the overall angular momentum distribution of the accreting gas parcels.

MBH–MBH mergers also influence the spin evolution. Berti & Volonteri (2008) consider how the dynamics of BH mergers influence the final spin. Except in the case of aligned mergers, a sequence of BH mergers can lead to large spins $> 0.9$ only if MBHs start already with large spins and they do not experience many major mergers. Therefore, the common assumption that mergers between MBHs of similar mass always lead to large spins needs to be revised.

The focus of this paper is the cosmic evolution of spins of MBHs with $M \sim 10^6$–$10^{10} M_\odot$ (Richstone et al. 1998; Ferrarese & Ford 2005), specifically how accretion and MBH–MBH mergers determine the magnitude of spins. Very few studies thus far have investigated the joint MBH mass and spin coevolution (Moderski et al. 1998; Volonteri et al. 2005; Shapiro 2005; Lagos et al. 2009; Fanidakis et al. 2011; Barausse 2012). As with previous efforts, we adopt a semi-analytical approach in order to capture both the cosmic evolution of structures and the processes that occur near MBHs. This approach allows us to model accretion processes using an analytical formalism that in principle has unlimited spatial resolution. This unlimited spatial resolution is particularly relevant as the physical processes that influence spin evolution occur near the MBH and, unfortunately, direct cosmological simulations at sub-parsec resolution are still infeasible. The other advantage of this approach is that each assumption is clearly described mathematically, making the calculation easily reproducible, modifiable, and testable under different assumptions by scientists with different theoretical stances. Finally, one should appreciate that our formalism does not have many more “cranks” than sub-grid prescriptions adopted in numerical simulations, while offering a clear framework that can be replicated or modified in a very economical way using a standard desktop by any scientist who decides so.

It is important to note that our model reproduces a large number of observational constraints (the luminosity function of AGNs, the mass function of MBHs, the relation between MBHs and their hosts, and the mass density in MBHs at low and high redshift). Since we are comparing our models with a large number of observables, there is not much leverage for the model parameters or assumptions to be varied. In Section 5, we discuss how the model’s parameters can be changed (and which cannot be modified). Of course, since there are not many observational constraints on MBH spins (however, see Sections 6 and 7), the possibilities to compare our models with observations are rather limited. Within the assumptions made, and the observational constraints used to anchor our calculation, the model is robust.

Since this investigation is theoretical, we present a framework that predicts a set of properties for the MBH population. In contrast to previous investigations that focused on high-redshift quasars (e.g., Volonteri et al. 2005; Shapiro 2005), our main interest here is to study the populations of low-redshift MBHs and AGNs whose spins may be directly measured through X-ray spectroscopy or indirectly estimated through their average radiative efficiency. In particular, our models aim at translating the theoretical expectations in a framework that can be directly applied to observational samples, for instance by casting our results in terms of AGN luminosity rather than MBH mass, as
is typically done in the literature (as only a small subsample of AGNs have mass measurements).

The outline of this paper is as follows. In Section 2, we describe the basic infrastructure that we use to model the cosmic evolution of structures. In Section 3, we summarize how we model spin evolution in MBH–MBH mergers, while in Section 4 we describe how different phases of accretion, related to the cosmic evolution of galaxies and the MBHs they host, influence MBH spins. In Section 5, we consider a series of observational constraints that we adopt to anchor our model. In Section 6 we discuss our results and we present our conclusions in Section 7.

2. THE BACKBONE: DARK MATTER HALOS AND GALAXIES

We investigate the evolution of MBHs via cosmological realizations of the merger hierarchy of dark matter halos from early times to the present in a ΛCDM cosmology (WMAP5; Komatsu et al. 2009). We track the merger history of 300 parent halos with present-day masses in the range $10^{11} < M_h < 10^{15} M_\odot$ with a Monte Carlo algorithm (Volonteri et al. 2003). The mass resolution of our algorithm reaches $10^6 M_\odot$ at $z = 20$ and the most massive halos are split into up to 600,000 progenitors.

We wish to keep our models as simple as possible, while making sure that the properties of the MBHs we study are correctly determined through the cosmic evolution of their hosts. We do not explicitly model the evolution of the baryonic component of the host galaxies through cooling, star formation, and various feedback mechanisms (see Lagos et al. 2009; Fanidakis et al. 2011; Barausse 2012 and references therein for semi-analytical models that treat in detail the baryonic component of galaxies and its link to MBH evolution). In our models, we use only one parameter to link the host halo to the central MBH: the host’s central velocity dispersion. We link the central stellar velocity dispersion of the host to the asymptotic virial velocity ($V_c$) assuming a spherical, isothermal halo such that $\sigma = V_c/\sqrt{2}$. We calculate the circular velocity from the mass of the host halo and its redshift. A halo of mass $M_h$ collapsing at redshift $z$ has a circular velocity:

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

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

At high redshift, we need dark matter halos with MBHs created by gas collapse. Specifically, we adopt here the formation model detailed in Natarajan & Volonteri (2012) based on Toomre instabilities ( Lodato & Natarajan 2006). The Toomre parameter is defined as $Q = (c_s c)/\kappa (\pi G \Sigma)$, where $\Sigma$ is the surface mass density, $c_s$ is the sound speed, $\kappa = \sqrt{2V_c}/R$ is the epicyclic frequency, and $V_c$ is the circular velocity of the disk. When $Q$ approaches a critical value, $Q_c$, of the order of unity, the disk is subject to gravitational instabilities. If the destabilization of the system is not too violent, instabilities lead to mass infall instead of fragmentation into bound clumps and global star formation in the entire disk ( Lodato & Natarajan 2006). This process stops when the amount of mass transported to the center is sufficient to make the disk marginally stable. The mass that has to be accumulated in the center to make the disk stable, $M_{\text{inf}}$, is obtained by requiring that $Q = Q_c$. This condition can be computed from the Toomre stability criterion and from the disk properties, determined from the dark matter halo mass, via $T_{\text{vir}} \propto M_h^{2/3}$ and angular momentum, via the spin parameter, $\lambda_{\text{spin}}$:

$$M_{\text{inf}} = f_d M_{\text{halo}} \left[ 1 - \frac{8\lambda_{\text{spin}}}{f_d Q_c} \left( \frac{j_d}{T_{\text{vir}}} \right) \left( \frac{T_{\text{gas}}}{T_{\text{vir}}} \right)^{1/2} \right],$$

for $\lambda_{\text{spin}} < \lambda_{\text{max}} = f_d Q_c/8(f_d/j_d)(T_{\text{vir}}/T_{\text{gas}})^{1/2}$. Here, $\lambda_{\text{max}}$ is the maximum halo spin parameter for which the disk is gravitationally unstable, $f_d = 0.05$ is the gas fraction that participates in the infall, and $j_d = 0.05$ is the fraction of the halo angular momentum retained by the collapsing gas. We further assume $T_{\text{gas}} = 5000 \text{ K}$ and $Q_c = 2$ (see Volonteri et al. 2008 for a validation of these parameter choices). Given the mass and spin parameter of a halo, the mass that accretes to the center in order to make the disk stable, $M_{\text{acc}}$, is an upper limit on the mass that can go into MBH formation. We assume here $M_{\text{acc}} = M_{\text{inf}}$. We refer the reader to Natarajan & Volonteri (2012) and references therein for details on the MBH formation process.

Our model for MBH growth requires only the knowledge of whether a galaxy is gas rich (we typically refer to gas-rich galaxies as “disks” in this paper) or gas-poor (“spheroids”). Since forming galaxy disks even in high-resolution cosmological simulations is extremely challenging (Governato et al. 2007), we keep our model for galaxy morphology as simple as possible. Morphology is related to the merger history, using a three-parameter model, where spheroid formation depends on both the halo mass ratio and the absolute halo mass, and a spheroid can reacquire a disk through cold flows and mergers with gas-rich galaxies. Kodama et al. (2007) show that the fraction of disk- versus spheroid-dominated galaxies is well explained if the only merger events that lead to spheroid formation have mass ratios >0.3 and virial velocities >55 km s$^{-1}$; also, the merger timescale must be shorter than the time between when the merger starts and today, $z = 0$. We assume that spheroids form after a merger that meets these requirements. We additionally allow a disk to reform after 5 Gyr in galaxies with virial velocities <300 km s$^{-1}$ where no major mergers occurred, in order to include the effects of cold flows. In Section 5, we discuss the rationale for this approach.

3. SPIN EVOLUTION DUE TO MBH–MBH Mergers

We assume that when two galaxies hosting MBHs merge, the MBHs themselves merge within the merger timescale of the host halos ( Sesana et al. 2007; Dotti et al. 2007 and references therein). We adopt the relations suggested by Boylan-Kolchin et al. (2008) for the galaxy merger timescale. We model MBH spin changes due to mergers, adopting an analytical scheme similar to that described in Berti & Volonteri (2008), based on simulations of BH mergers in full general relativity (Rezzolla et al. 2008; Lousto et al. 2010a). Kesden et al. (2010) has validated the consistency of different fitting formulae for calculating the spin of MBH remnants, and we refer the reader to Lousto et al. (2010b) for the most comprehensive fitting formulae. Due to computational constraints, we adopt here the fitting formulae of Rezzolla et al. (2008) for their easy
implementation:

\[
\begin{align*}
|a_{\text{fin}}| &= \frac{1}{(1 + q^2)^2} (|a_1|^2 + |a_2|^2 q^2 + 2|a_2||a_1| q^2 \cos \alpha \\
+ 2 (|a_1| \cos \beta + |a_2|^2 q^2 \cos \gamma) (|q| + |q|^2)^2),
\end{align*}
\]

where \( q \equiv M_2/M_1 \leq 1 \) is the mass ratio between the two MBHs; the three (cosine) angles \( \alpha, \beta, \gamma \) are defined by

\[
\begin{align*}
\cos \alpha &= \hat{a}_1 \cdot \hat{a}_2, \quad \cos \beta = \hat{a}_1 \cdot \hat{L}, \quad \cos \gamma = \hat{a}_2 \cdot \hat{L}.
\end{align*}
\]

where \( \hat{L} \) is the magnitude of the orbital angular momentum and

\[
\begin{align*}
|\hat{L}| &= \left( \frac{s_4}{1 + q^2} \right) (|a_1|^2 + |a_2|^2 q^2 + 2|a_2||a_1| q^2 \cos \alpha) \\
+ \left( \frac{s_5 v^2 + t_0 + 2}{1 + q^2} \right) (|a_1| \cos \beta + |a_2|^2 q^2 \cos \gamma) \\
+ 2\sqrt{3 + t_2 v^2 + t_3 v^2},
\end{align*}
\]

where \( v \) is the symmetric mass ratio \( v \equiv M_1M_2/(M_1 + M_2)^2 \) and the coefficients take the values \( s_3 = -0.129 \pm 0.012, s_5 = -0.384 \pm 0.261, t_0 = -2.686 \pm 0.065, t_2 = -3.454 \pm 0.132, \) and \( t_1 = 2.353 \pm 0.548. \)

We model the spin–orbit configuration differently depending on the properties of the host galaxies. During a gas-rich merger, large amounts of gas are driven toward the centers of the two interacting galaxies (Downes & Solomon 1998) and form a dense circumnuclear disk in which the MBHs settle. In this phase, the MBHs accrete in a coherent manner at a rate sufficient to align their spins, initially oriented at random, to the angular momentum of the nuclear disk (Liu 2004; Bogdanović et al. 2007; Dotti et al. 2010). In response to the Bardeen–Petterson (Bardeen & Petterson 1975) warping of the small-scale accretion disks grown around each MBH, total angular momentum conservation imposes fast \((\lesssim 1 \text{ Myr})\) alignment of the BH spins with the angular momentum of their orbit and therefore the angular momentum of the large-scale circumnuclear disk (Scheurer & Feiler 1996; Natarajan & Pringle 1998; Natarajan & Armitage 1999; Nelson & Papaloizou 2000; Volonteri et al. 2007; Martin et al. 2009; Lodato & Price 2010; Ulubay-Siddiki et al. 2013), unless one considers a vertical viscosity equal to the radial one, as early works did (Papaloizou & Pringle 1983; Kumar & Pringle 1985) or disks with low viscosities, such as protoplanetary disks or thick disks (Lubow et al. 2002). See Appendix B for additional information.

Thereafter, accretion remains prograde until coalescence, with no major changes in the MBH spin orientation. Under these circumstances, \( \cos \alpha = \cos \beta = \cos \gamma = 1 \), and the MBH remnant retains the spin direction of the parent MBH spins, both oriented parallel to the angular momentum of their orbit: the post-coalescence MBH may thus acquire a large spin \( > 0.7–0.9 \) (Berti & Volonteri 2008), the sum of the internal and orbital spins.

In the case of gas-poor mergers, instead, i.e., when MBHs do not accrete during mergers and evolve solely via stellar dynamical processes, there is no reason to expect any symmetry or alignment (Bogdanović et al. 2007), so isotropy should be a good assumption in the absence of accretion disks or gas; \( \cos \alpha, \cos \beta, \) and \( \cos \gamma \) are isotropically distributed. Berti & Volonteri (2008) show that for isotropic configurations, mergers tend to “spin-down” a fast-spinning BH (see also Hughes & Blandford 2003). For intermediate to large mass ratios (mass ratio \( q = M_{\text{BH},2}/M_{\text{BH},1} \approx 1 \) between 0.1 and 1), mergers tend to produce MBHs with average spins very close to the value \( \approx 0.7 \) resulting from equal-mass, non-spinning mergers. For smaller mass ratios, the larger MBH dominates the dynamics and the final spin can be substantially larger or smaller than this value.

4. SPIN EVOLUTION DUE TO ACCRETION

We discuss here the feeding of MBHs in the quasar phase and its aftermath, in addition to MBHs in more quiescent Seyfert galaxies. Simulations of galaxy mergers and MBH activity (Di Matteo et al. 2005; Hopkins et al. 2006) show that for every accretion episode triggered by a galaxy merger, a three-phase picture can be drawn. At the beginning, the MBH has a “healthy diet,” with \( f_{\text{Edd}} \equiv L/L_{\text{Edd}} \leq 1 \). When the MBH mass reaches the “M-σ” relation, the MBHs feedback can be sufficient to unbind and “blowout” the gas feeding it (Hopkins & Hernquist 2006), causing a final “starvation,” when \( f_{\text{Edd}} \) rapidly decreases until no more gas is available to feed the MBH. We argue that during the healthy diet and the blowout phases MBHs gain a high spin while accreting efficiently and coherently (Dotti et al. 2010), building the population of high-\( z \) quasars. Coherent accretion ensues because MBHs in merger remnants are expected to be surrounded by dense circumnuclear disks (Sanders & Mirabel 1996). Maio et al. (2013) studied the evolution of the angular momentum of material feeding MBHs embedded in circumnuclear disks and found that coherence of the accretion flow near each MBH reflects the large-scale coherence of the disk’s rotation.

After the “blowout” phase, starving MBHs are no longer surrounded by a thick gas disk that determines the angular momentum of the material ending up in the accretion disk. During this last phase, we do not expect accretion to necessarily proceed coherently any longer. During the starvation phase, the accretion rate decreases rapidly. The depletion of gas in galaxies and the decrease in the galaxy interaction rate at late cosmic times therefore causes a widespread “famine” in low-\( z \) ellipticals (i.e., the merger remnants), where AGNs with low accretion rates dominate (the “radio”; mode, see, e.g., Croton et al. 2006; Churazov et al. 2005).

We investigate the evolution of MBH spins during the quasar phase by expanding previous work (Volonteri et al. 2005, 2007) to more realistic models. We model the joint mass and spin evolution by coupling the results of the mass accretion rate as a function of time in simulations (Hopkins et al. 2005; Volonteri et al. 2006) with the spin evolution due to disk accretion (Volonteri et al. 2007), thus solving a system of two coupled differential equations \( f_{\text{Edd}} \) as a function of \( M \) and time; spin \( a \) as a function of \( f_{\text{Edd}} \) and \( M \). The framework has been derived in Volonteri et al. (2005, 2006, Hopkins & Hernquist 2006, and Volonteri et al. 2007). We remind the reader here of the relevant information.

4.1. Quasar Phase

After a halo merger with mass ratio larger than 3:10, in which at least one of the two participants is a disk galaxy (hence, with a conspicuous cold gas content), we assume that a merger-driven accretion episode is triggered.\(^6\) After a dynamical timescale

\(^6\) Note that this assumption is at variance with previous models of MBH cosmic evolution within our framework (e.g., Volonteri et al. 2003, 2008; Volonteri & Natarajan 2009), where the threshold was set to a lower value of 1:10. The reason for increasing the threshold necessary for the quasi-stellar object (QSO) phase to occur is due to the addition of avenues for MBH growth other than merger-driven accretion and MBH–MBH mergers.
has elapsed (roughly, after the first pericentric passage; cf. Van Wassenhove et al. 2012), accretion starts. If at that point the MBH mass lies below the $M-\sigma$ relation, accretion occurs at the Eddington rate ($f_{\text{Edd}} = 1$). Additionally, during this early phase, the MBH is nested into a nuclear disk that feeds the MBH coherently. The following scheme is applied to the joint evolution of mass, spin, and radiative efficiency in this phase. Let us define $M$ and $a$ as the BH mass and spin parameter at the beginning of the timestep and $\mu$ as the cosine of the angle between the MBH spin and the inner accretion disk angular momentum. Irrespective of the infalling material’s original angular momentum vector, Lense–Thirring precession imposes axisymmetry close in, with the gas accreting on either prograde ($\mu = 1$) or retrograde ($\mu = -1$) equatorial orbits. In natural units, where $c = G = 1$:

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

is the radius of the ISCO, where $Z_1$ and $Z_2$ are functions of $a$ only (Bardeen et al. 1972):

$$Z_1 \equiv 1 + (1 - a^2)^{1/3}[(1 + a)^{1/3} + (1 - a)^{1/3}],$$

$$Z_2 \equiv [3a^2 + Z_1^2]^{1/2},$$

and the upper (lower) sign refers to prograde (retrograde) orbits. We calculate the accretion efficiency as

$$\epsilon = 1 - E_{\text{ISCO}},$$

$$E_{\text{ISCO}} = \left(1 - \frac{2}{3r_{\text{ISCO}}}\right)^{1/2},$$

which is also a plausible assumption for the radiative efficiency ($\epsilon$, mass-to-energy conversion) for thin-disk accretion occurring at large fractions of the Eddington rate (see below for the case of radiatively inefficient flows). We calculate self-consistently the radiative efficiency from the MBH spin and the location of the ISCO (i.e., taking into consideration the direction of the relative angular momentum of spin and disk, co- or counter-rotating).

Assuming that during a timestep $\Delta t \sim 10^4-10^6$ yr the radiative efficiency and Eddington rate remain constant ($\epsilon = \text{const}$ and $f_{\text{Edd}} = \text{const}$), from the derivation shown in Appendix A, one obtains that the MBH mass grows as

$$M(t + \Delta t) = M(t) \exp \left( f_{\text{Edd}} \frac{\Delta t}{f_{\text{Edd}}} \frac{1 - \epsilon(t)}{\epsilon(t)} \right),$$

where $f_{\text{Edd}} = (\sigma \epsilon G m_p) / (4\pi G m_p) = 0.45$ Gyr and $f_{\text{Edd}}$ represents the Eddington fraction. We update the magnitude of the MBH spin through:

$$a(t + \Delta t) = \frac{r_{\text{ISCO}}(t)^{1/2}}{3} \frac{M(t)}{M(t + \Delta t)}$$

$$\left[4 - \left(\frac{3M(t)^2}{M(t + \Delta t)^2}r_{\text{ISCO}}(t)^2 - 2\right)^{1/2}\right]^2$$

for $M(t + \Delta t) / M(t) \leq r_{\text{ISCO}}^{1/2}(t)$,

$$a(t + \Delta t) = 0.998 \text{ for } \frac{M(t + \Delta t)}{M(t)} \geq r_{\text{ISCO}}^{1/2}(t)$$

(Bardeen 1970). After updating the spin magnitude, we also update the mass-to-energy conversion efficiency by determining the new ISCO corresponding to $a(t + \Delta t)$, and therefore $\epsilon(t + \Delta t)$ to be used at the successive timestep iteratively. We assume here fast alignment between the accretion disk and the spin (see Natarajan & Pringle 1998; Volonteri et al. 2005, 2007; Perego et al. 2009), as the alignment timescale is $\sim$ Myr, so that accretion is prugrade during the quasar phase.

### 4.2. Decline Phase

When a MBH reaches a mass close to the value corresponding to the $M-\sigma$ correlation ($M_{\text{BH, } \sigma}$) for its host, we assume, following Hopkins et al. (2006) and Hopkins & Hernquist (2006), that self-regulation ensues and the MBH feedback unbinds the gas closest to the MBH, thus reducing its feeding. For simplicity, we further assume that the BH–$\sigma$ ($M-\sigma$) scaling is:

$$M = 10^8 \left(\frac{\sigma}{200 \text{ km s}^{-1}}\right)^4 \frac{M_\odot}{25}$$

(Tremaine et al. 2002). To match the luminosity function of quasars, we start the decline phase when the MBH mass is $0.25 \times M_{\text{BH, } \sigma}$. The MBH continues accreting during the decline of accretion and it typically reaches a value closer to the $M_{\text{BH, } \sigma}$ by the end of the accretion episode. Note that our model does not necessarily imply that the $M_{\text{BH, } \sigma}$ relation is a tight correlation. We assume only that the feedback during a high-accretion rate quasar phase establishes at that time for that object an $M-\sigma$ relation (cf. Silk & Rees 1998; Fabian 1999). Additional processes, such as MBH–MBH mergers (Section 3), accretion during the “decline phase” (Section 4.2), accretion of recycled gas (Section 4.3), and accretion of gas stolen from molecular clouds (MCS; Section 4.4) do not have any limit imposed by the $M_{\text{BH, } \sigma}$ relation. In fact, these processes produce scatter by pushing the MBHs above or below the relationship, depending also on the galaxy history (see Volonteri & Ciotti 2013).

We model the decrease of the accretion rate through the analytical formula from Hopkins & Hernquist (2006):

$$f_{\text{Edd}}(t) = \left(\frac{t + t_{\text{fsh}}}{t_{\text{fsh}}}\right)^{-\eta_L},$$

where $\eta_L \simeq 2$, $t_{{fsh}} \simeq 4 \times 10^6 (M_{\text{BH, } \sigma}/10^8) M_\odot$ yr, and $t = 0$ (where $f_{\text{Edd}} = 1$) represents the time when the MBH reaches the threshold $(0.25 \times M_{\text{BH, } \sigma})$. Since $M_{\text{in}}(t) = f_{\text{Edd}}(t) M(t) / f_{\text{Edd}}$ depends on time, the accretion rate must be integrated self-consistently, and the mass now grows with time as

$$M(t) = M(0) \exp \left(\frac{1}{\eta_L - 1} \left[\frac{1}{1 - \eta_L} - \frac{1}{(t + t_{\text{fsh}})^{1-\eta_L}}\right] \frac{t_{\text{fsh}}}{t_{\text{fsh}} - \eta_L} \frac{1 - \epsilon(t)}{\epsilon(t)} \right).$$

7 We limit the MBH spin to $a = 0.998$ following the calculation of Thorne (1974) that showed that the radiation emitted by the disk and swallowed by the BH produces a counteracting torque, which prevents spin up beyond this value. We note that magnetic fields connecting material in the disk and the plunging region may further reduce the equilibrium spin by transporting angular momentum outward in non-geometrically thin disks. Fully relativistic magnetohydrodynamic simulations for a series of thick accretion disk models show that spin equilibrium is reached at $a = 0.93$ (Gammie et al. 2004), while in slim disks accretion can (for low viscosities) increase the spin up to $a = 0.9994$ (Sadowski et al. 2011).
We again use Equations (8)–(15) to model spin evolution, however, we explore two possible scenarios. In our reference case, we assume that the “outflow” causes some stirring of the angular momentum of the gas within the central region. We therefore explore a “chaotic” case where in the decline phase we pick a new random \( \mu = 1 \) or \( \mu = -1 \) at each timestep (\( \sim 10^3 \)–\( 10^4 \) years) to mimic the lack of coherence in the accretion flow after MBH feedback has blown away the surrounding gas. In a second model, we assume instead that the “blow-out” of gas does not strongly affect the angular momentum of the material near the MBH and keep \( \mu = 1 \). Given that the timescale for decline is longer for larger MBHs (cf. Equation (17)), larger MBHs have longer phases at relatively high accretion rates and the rate of decrease of MBH spins increases, as more mass is accreted in a non-coherent fashion.

When the accretion rates become very sub-Eddington, we assume that the accretion flow becomes optically thin and geometrically thin. In this state, the radiative power is strongly suppressed (e.g., Narayan & Yi 1994; Abramowicz et al. 1995) so that the radiative efficiency differs from the mass-to-energy conversion efficiency, \( \epsilon \), that depends on the location of the ISCO only. Indeed, the radiative efficiency becomes very model-dependent and uncertain. In order to estimate the effect of radiatively inefficient accretion on the MBH population, we adopt here a specific functional form for the radiative efficiency. Following Merloni & Heinz (2008), we write the radiative efficiency, \( \epsilon_{\text{rad}} \), as a combination of the mass-to-energy conversion, \( \epsilon \), and a term that depends on the properties of the accretion flow itself. Merloni & Heinz (2008) suggest that the transition in the disk properties occurs\(^8\) at \( f_{\text{edd}} < f_{\text{edd,cr}} = 3 \times 10^{-2} \) and that \( \epsilon_{\text{rad}} = \epsilon (f_{\text{edd}} / f_{\text{edd,cr}}) \). This specific choice allows us to estimate qualitatively the impact of radiatively inefficient sources on AGN populations and on the inferences that one can (or cannot) make from the observables.

4.3. Quiescent Elliptical Phase

After the formation of an elliptical galaxy, the feeding of the MBH can be sustained by the recycled gas (primarily from red giant winds and planetary nebulae) of the evolving stellar population (Ciotti & Ostriker 1997, 2001, 2007; Ciotti et al. 2010). As shown by Ciotti & Ostriker (2012), the behavior of the accretion rate is similar to what we describe above: at early times the evolution is characterized by major, albeit intermittent, accretion episodes, while at low redshift the accretion rate is smooth. In a second model, we assume instead that the “blow-out” of gas forms around the MBH corresponds to the material originally contained in a cylinder with radius \( R_{d} \). If \( R_d < 0.1 \) pc, the spin evolution would then resemble the “chaotic accretion” scenario. This argument was discussed only qualitatively by Volonteri et al. (2007). We now wish to provide quantitative statistical predictions for the distribution of MBH spins in different hosts. We follow here Sanders (1981) and Hopkins & Hernquist (2006) to determine the event rate and Volonteri et al. (2007) to couple accretion episodes to spin evolution.

In a disk galaxy, at each timestep, \( \Delta t \), we determine the probability of a MC accretion event as

\[
P = \frac{\Delta t}{t_{\text{MC}}} \simeq 10^{-3} \frac{\sigma}{R_{d}} \left( \frac{M}{10^{7} M_{\odot}} \right),
\]

where \( R_d \approx 10 \) pc (Hopkins & Hernquist 2006). As in Volonteri et al. (2007), we further assume a lognormal distribution (peaked at \( \log (M_{MC} / M) = 4 \), with a dispersion of 0.75) for the mass function of MCs close to galaxy centers (based on the Milky Way case; e.g., Perets et al. 2007).

We model accretion of MCs through a description inspired by Bottema & Sanders (1986) and Wardle & Yusef-Zadeh (2008). We assume that the MBH captures only material passing within the Bondi radius, \( R_{B} \), and we also assume that specific angular momentum is conserved, so that the outer edge of the disk that forms around the MBH corresponds to the material originally at the Bondi radius:

\[
R_d = 2 \lambda^2 R_B = 8.9 \text{ pc} \lambda^2 \left( \frac{M}{10^{7} M_{\odot}} \right) \left( \frac{\sigma}{100 \text{ km s}^{-1}} \right)^{-2},
\]

where \( \lambda \) is the fraction of angular momentum retained by gas during circularization. The maximum captured mass will be contained in a cylinder with radius \( R_{d} \) and length \( 2 \times R_{d} \), the MC diameter. If \( k \) is the ratio of the mass going into the disk with respect to the whole mass in the cylinder, then:

\[
M_{d, \text{max}} = k R_B M_{MC} = 4.7 \times 10^{4} k \left( \frac{M}{10^{7} M_{\odot}} \right)^{2} \left( \frac{\sigma}{100 \text{ km s}^{-1}} \right)^{-4} M_{\odot}.
\]

The inflow time will be of order the viscous timescale for the disk:

\[
\tau_{\text{visc}} = \left( \frac{R_{d}^3}{\alpha^2 G M_{d}} \right)^{1/2} = 3.4 \times 10^{5} \lambda^{3} \left( \frac{M}{\alpha_{d} / 10^{7} M_{\odot}} \right) \left( \frac{\sigma}{100 \text{ km s}^{-1}} \right)^{-3} \text{ yr},
\]

\(^{8}\) Note that Merloni & Heinz (2008) use a different notation and terminology. Their \( \mu \) is our \( f_{\text{edd}} \equiv L / L_{\text{edd}} \equiv \epsilon_{\text{rad}} M_{\text{eff}} / L_{\text{edd}} \) and their \( \mu \) is \( f_{\text{edd}} / \epsilon_{\text{rad}} \). As long as the accretion flow is optically thick and geometrically thin, i.e., before the transition to very sub-Eddington flows, \( \dot{m} = f_{\text{edd}} \).
Yusef-Zadeh (2012), we set the probability of MC accretion increases with galaxy mass. However, the most massive galaxies are spheroids and they therefore do not have a population of MCs available.

(A color version of this figure is available in the online journal.)

so that for the whole disk to be consumed we can calculate an upper limit to the mean accretion rate and luminosity:

\[
\dot{M}_{\text{max}} = \frac{M_{d,\text{max}}}{t_{\text{visc}}} = 0.1 \frac{\alpha_v \kappa}{\lambda^3} \frac{M}{10^7 M_\odot} \times \left( \frac{\sigma}{100 \text{ km s}^{-1}} \right)^{-1} N_{23} M_\odot \text{yr}^{-1},
\]

where \(N_{23}\) is the column density in the MC in units of \(10^{23} \text{ cm}^{-2}\) and \(\alpha_v = 0.1\) is the viscosity parameter. Following Wardle \\& Yusef-Zadeh (2012), we set \(\lambda = 0.3\) and \(\kappa = 1\). If \(M_{\text{max}}\) is less than the Eddington rate (assuming a radiative efficiency of 10%), we let the MBH accrete the whole \(M_{d,\text{max}}\) over a time \(t_{\text{visc}}\), otherwise we treat accretion similarly to the “decline” phase of quasars (Equations (17) and (18)) since feedback from the high luminosity produced by accreting the cloud will limit the amount of material the MBH can effectively swallow.

From Equation (21), it is evident that the mass accreted in one of these episodes is typically much less than the mass of the MBH (usually between \(10^{-2}\) and \(10^{-3}\) of the MBH mass). Therefore, we assume that no alignment between the accretion disk and the MBH spin can occur and that retro- and prograde accretion are equally probable, i.e., we assign \(\mu = 1\) or \(\mu = -1\) with equal probability at each event and keep \(\mu\) constant over the accretion phase, using Equation (15) to evolve the spin magnitude. As shown by Volonteri et al. (2007), these assumptions result in a spin down in a random walk fashion that depends on the mass of the MBH and on the number of events. We note that this is an extreme condition of randomness in the orbits and distribution of MCs, as any common sense of rotation caused by the presence of disk-like structures in the host would decrease the degree of anisotropy (Dotti et al. 2013).

Equation (19), instead, shows that accretion of MCs is more probable in large galaxies, since the accretion probability is directly proportional to the velocity dispersion of the galaxy. Large galaxies, however, are more likely to be gas-poor spheroids.

Large galaxies therefore have a lower probability of being gas-rich and hosting a population of MCs, while at the same time their MBHs have a higher probability of capturing MCs, if clouds are present. Thus, this type of accretion events occurs typically in galaxies hosted in halos with masses \(\sim 10^{11}–10^{12} M_\odot\) and fuels MBHs with masses \(\sim 10^{5}–10^{7} M_\odot\). In Figure 1, we show the distribution of the masses of MBHs and their host halos where accretion of MCs takes place according to our scheme.

5. ANCHORING THE MODEL

Here, we present the constraints that we use to “anchor” our model to the observed properties of galaxies and AGNs. After validating our scheme for accretion and host evolution, we will discuss what its implications are for the still unknown distribution of MBH spins.

First, in Figure 2 we compare our model to constraints at \(z = 0\). In the bottom panel, we show the morphological fraction as a function of galaxy stellar mass. We scale from halo mass to stellar mass through the data described in Figure 1 of Hopkins et al. (2010), assuming that a fixed fraction (10%–50%) of the baryons is in stars (open red points and filled orange points, respectively). We compare the fraction of spheroids to the result of Conselice (2006).

(A color version of this figure is available in the online journal.)

(A color version of this figure is available in the online journal.)

Figure 1. Properties of all MBHs fed by MCs at all redshifts: normalized distribution of the masses of MBHs (top) and their host halos (bottom). The probability of MC accretion increases with galaxy mass. However, the most massive galaxies are spheroids and they therefore do not have a population of MCs available.

Figure 2. Top panel: relationship between MBH mass and circular velocity. The circles are model MBHs at \(z = 0\) and the error bars show the datapoints collected in Kormendy et al. (2011) and the best fit derived in Volonteri et al. (2011c). Bottom panel: fraction of spheroids as a function of stellar mass. We scale from halo mass to stellar mass through the data described in Figure 1 of Hopkins et al. (2010), assuming that a fixed fraction (10%–50%) of the baryons is in stars (open red points and filled orange points, respectively). We compare the fraction of spheroids to the result of Conselice (2006).

(A color version of this figure is available in the online journal.)
Our scheme produces an AGN population in good agreement with the observational constraints and they are shown in Figure 3. Poissonian statistics, and the fraction of absorbed AGNs (La Franca et al. 2005). Orange (45° hatching): coherent accretion during the decline of the quasar phase. Gray (horizontal hatching): chaotic accretion during the decline of the quasar phase.

(A color version of this figure is available in the online journal.)

Figure 3. Luminosity function at different redshifts. We show here the minimum and maximum values, considering both 1σ statistical uncertainties, using Poissonian statistics, and the fraction of absorbed AGNs (La Franca et al. 2005). Orange (45° hatching): coherent accretion during the decline of the quasar phase. Gray (horizontal hatching): chaotic accretion during the decline of the quasar phase.

suppressed growth is that no MC accretion can happen in our scheme in gas-poor galaxies, and we have not implemented a time-dependent accretion of recycled gas. As noted by Ciotti & Ostriker (2007), the behavior of a MBH fueled by stellar mass loss is self-regulated between the “on” and “off” phases. Our model includes only the “off” (quiescent) phase and therefore underestimates the growth due to this fueling channel.

Our second anchor is the luminosity function of AGNs in the redshift range 0.5 ≤ z ≤ 3. Luminosity function strongly constrain our accretion scheme and they are shown in Figure 3. Our scheme produces an AGN population in good agreement with the observational results at z = 0.5, z = 2, and z = 3, while we slightly underproduce AGNs at z = 1. The lack of high luminosity quasars at z = 2 and z = 3 is due to our merger trees not including halos massive enough to host MBHs with masses above a few 10^8 M⊙ at those redshifts. Overall, however, we obtain the correct trends. This figure shows that there is little difference in the two models we explore in terms of their effects on feedback over the angular momentum of the nuclear gas (“chaotic” or “coherent” decline phases). The difference between chaotic and coherent decline (the “quasar” phase is coherent in all cases) reflects only on spin and, as a consequence, on radiative efficiency, not on the Eddington ratio. In the following, we will distinguish between the two models only when the results are significantly different, otherwise we will show only the reference case (the chaotic decline phase).

The primary parameters influencing the performance of the model against the constraints are the mass ratio above which a merger can trigger quasar activity and the fraction of MBHs when the decline phase starts. These two parameters are weakly degenerate. The former parameter is set to > 3:10 in the present model to match as closely as possible the luminosity function at the bright end without overestimating MBH masses at a given circular velocity at z = 0, as happens when a lower threshold is selected. A much lower threshold would also be in disagreement with simulations of galaxy mergers that study merger-drive AGN activity; these simulations show that with a mass ratio of 1:6, a high level of AGN activity does not occur (S. Van Wassenhove et al., in preparation). We tested a case where we instead increased the threshold to 1:2. In this case, the bright end of the luminosity function would disappear at z > 2, while little to no change would occur at the faint end. For the latter parameter, we adopted a value of 0.25 × MBH,σ. We tested a case that brought the MBHs exactly on MBH,σ, but this situation leads to strongly overestimating the MBH masses at z = 0 at a given Vc. We also tested a case with 0.5 × MBH,σ. In that case, we still overestimated MBH masses at z = 0 at a given Vc (the masses are high by a factor of two overall in this case, over the best fit relationship). Decreasing the parameter value to 0.125 × MBH,σ instead underestimates MBH masses at z = 0 at a given Vc, by a factor of 2.25 overall, over the best fit relationship. Finally, we tested a case where we decreased the mass ratio threshold for merger-driven AGN activity to 1:10 and, at the same time, we decreased the mass limit to 0.125 × MBH,σ to compensate. In this case, the luminosity function is similar to the case with 1:10 and 0.25 × MBH,σ, but the relationship between MBH mass and Vc is tilted, with a shallower slope that underestimates the real MBH masses at the high Vc end. If we were to choose 1:10 and 0.25 × MBH,σ, then the relationship between MBH mass and Vc would overall be overestimated by more than a factor of two. In summary, we have run several tests to limit the space of free parameters and to disentangle weakly degenerate parameters until we found the set that best matches the observational constraints.

Accretion of MCs affects the faint end of the luminosity function. One could in principle boost the probability of MC accretion by assuming more compact clouds, however, the mass gained through this process is constrained by the faint end of the luminosity function of AGNs (Figure 3). Furthermore, only small variations can be tolerated by our model, as the current implementation gives a good match with observations. We note that MC accretion accounts for almost all sources up to L * ∼ 10^{12} L⊙ at z = 0.5–1 and L ∼ 10^{11} L⊙ at z = 2–3.

Finally, we generate ad hoc merger trees of Mh = 2 × 10^{13} M⊙ halos at z = 5–6 to check that our model reproduces the existence of powerful quasars at z ∼ 6 and that the mass density we obtain at z > 5, 6 × 10^3 M⊙Mpc^−3, does not overproduce the X-ray background (upper limit of 10^4 M⊙Mpc^−3; Salvaterra et al. 2012). Figure 4 compares the theoretical mass function of MBHs that power quasars with bolometric luminosities larger than 10^{45} erg s^−1 at z = 6 to the empirical mass function derived by Willott et al. (2010) from a sample of z = 6 quasars in the Canada–France High-z Quasar Survey.

As discussed above, our model, while far from being able to explain every single detail of the MBH population and its growth, qualitatively grasps most of the global behavior. We therefore consider our attempts to model the spin evolution also of a qualitative nature. Regardless of the simplified nature of our models, we can learn how different patterns influence the evolution of MBH spins.

6. ACCRETION AND SPIN EVOLUTION: RESULTS

In Figure 5, we show examples of spin evolution of an MBH hosted in a large spheroid today along its cosmic history. Most of the accreted MBH mass is accumulated during episodes of efficient growth at early times (up until z ∼ 2). At lower
redshift, the MBH grows mostly through MBH–MBH mergers. While different prescriptions for “chaotic” or “coherent” post-feedback phases lead to different histories for the MBH spin, the final spin is set by a MBH merger between two roughly equal mass systems occurring at late cosmic times. This fact is noticeable as a small jump from $a = 0.6$ to $a = 0.5$ at $z = 0.3$ in the top panel of Figure 5. In the case of disk galaxies, late phases of MC accretion at substantial accretion rates ($> 10^{-3}$ in Eddington units) contribute to setting the final spin of the MBH (Figure 6).

Statistically, Figure 7 shows the evolution of the logarithmic Eddington ratio as a function of redshift. The top panel shows a sample selected on the basis of MBH mass ($10^6 M_\odot < M < 10^8 M_\odot$ and $M > 10^8 M_\odot$) showing classic “anti-hierarchical” behavior. The bottom panel focuses instead on luminosity-selected AGNs, for which the typical accretion rates are much higher. (A color version of this figure is available in the online journal.)
typical Eddington ratio decreases slightly at late times (bottom panel of Figure 7), tracking instead the overall increase in the mass of MBHs.

In Figure 8, we compare the spins of MBHs hosted in disks and spheroids in different redshift bins. At low MBH mass ($M \sim 10^6 M_\odot$), MBHs hosted in gas-rich galaxies tend to have low spins. The spin distribution in gas-rich galaxies tends to move toward higher spins as mass increases. This fact is mostly due to MC accretion. For the most massive MBHs, most of the growth occurs earlier through merger-driven accretion that tends to spin-up MBHs. Accretion of MCs does not significantly modify the spins of the MBHs because the total angular momentum accreted through MCs is less than the total angular momentum the MBH has (i.e., the total mass accreted by the MBH through MCs is much less than the mass of the MBH). On the other hand, for low-mass MBHs, the mass accreted in MCs is of the same order as the MBH mass. In this case, MCs have a stronger effect on the spin distribution, lowering the typical spin of low-mass MBHs (we remind the reader that we have assumed that MCs accrete isotropically onto MBHs). The distribution of spins of MBHs hosted in gas-poor galaxies has little dependence on mass and redshift. In these galaxies, in general, most MBHs have spins $a \sim 0.4$–0.8. Spins tend to slightly decrease as MBH mass increases. We find no strong dependence on whether accretion occurs mostly chaotically or coherently after “feedback” effects take place, except at the highest masses. We have run a test case where we have artificially “turned-off” spin evolution via MBH–MBH mergers (while keeping the mass increase through mergers). In general, the effect of MBH–MBH mergers is to decrease the spins of the most massive MBHs in the case of the coherent post-feedback phase, while mergers increase MBH spins in the case of the chaotic post-feedback phase.

In Figure 9, we focus on active MBHs. MBHs accreting at high rates, $f_{\text{Edd}} > 0.1$, have very large spins at all $z > 2$. These are, for the most part, MBHs in the “quasar” phase. At $1 < z < 2$, more systems are caught in the decline phase and here, especially at high MBH masses, it becomes crucial whether accretion occurs mostly chaotically or coherently after the quasar phase (cf. left and right columns). This mass ($>10^8 M_\odot$) and redshift ($0.5 < z < 1$) range is the most suitable to probe how feedback affects the angular momentum of nuclear gas. For low-mass BHs, the spin distribution is mostly insensitive to the chaotic and coherent models (green triangles in Figure 9), while there is a stronger impact on high-mass BHs (orange squares in Figure 9). Therefore, the changes most strongly affect the high-luminosity end of the luminosity function. Finally, at $z < 0.5$, differences tend to disappear as MC accretion becomes the dominant feeding mechanism. The behavior of lower accretion rate systems is similar, although more and more systems are in the “decline” phase and lower spins become more common if accretion occurs chaotically during this phase.

The spin distribution, however, does not necessarily map the radiative efficiency distribution (Figure 10; also note that the mean efficiency differs, mathematically, from the efficiency corresponding to the mean spin), if a large population of sources has geometrically thick and optically thin accretion disks where the radiative efficiency may be suppressed with respect to the mass-to-energy conversion efficiency. Especially at low luminosity and low redshift, many AGNs are radiatively inefficient sources and the MBH spin is not relevant in determining their radiative efficiency. At low luminosity and high redshift, AGNs are a mixture of low-mass MBHs accreting at high rates in a radiatively efficient fashion and higher mass MBHs accreting at low rates. At low redshift, on the other hand, most low

![Figure 8](image1.png)

![Figure 9](image2.png)
luminosity sources are genuinely inefficient accretors. At high luminosity, instead, the signature of spin evolution with redshift is evident. Figures 9 and 10 show that our model is in very good agreement with the evolution of radiative efficiency with redshift derived observationally by Li et al. (2012; see their Figure 7). Our model also agrees with the suggestion of Li et al. and Shankar et al. (2013) that radiative efficiency may increase with BH mass.

Finally, in Figure 11, we show the distribution of MBH spins before and after a MBH–MBH coalescence. As expected, if MBHs have large spins prior to coalescence, MBH mergers tend to spin down the systems. Likewise, if MBHs have low spins prior to coalescence, MBH mergers tend to increase spins. We here show the spin of the primary MBHs in a binary prior to coalescence, as this quality is what gravitational wave observatories such as eLISA can measure.

7. DISCUSSION AND CONCLUSIONS

We developed a model for the evolution of MBHs that takes into account several physical mechanisms of MBH growth: MBH–MBH mergers, merger-driven accretion, stochastic accretion, and accretion of recycled gas. This model, however, does not include MBH feeding through disk instabilities, nor the burst phase of recycled gas feeding in elliptical galaxies. Under a series of plausible assumptions, we have derived the growth of MBHs, the properties of the AGN population, and the evolution of MBH spins. Our approach produces a population of MBHs and AGNs consistent with observations one in terms of, e.g., the luminosity function of AGNs, the relationship between MBHs and their hosts, and high-redshift quasars.

The main results of our models of MBH evolution can be summarized as follows.

1. At high redshift, MBHs grow mostly by merger-driven accretion, while at later times other channels become more important. In gas-rich galaxies, MC accretion dominates the growth of low-mass ($<10^7 M_\odot$) MBHs at $z < 2$. In gas-poor galaxies, MBH–MBH mergers are the main growth channel, especially at high MBH masses ($>10^7 M_\odot$).

2. The mass of most active BHs decreases with increasing cosmic time, in an anti-hierarchical fashion. Sustained accretion grows the most MBHs since early times without overproducing the MBH population as a whole.

3. MBH spins tend to be larger at redshifts $z > 2$, typically $a \gtrsim 0.8$. These large spins result from massive, coherent accretion events triggered by major mergers. This result is in general agreement with the trend found by Barausse (2012) using a complementary, more refined approach to modeling galaxy evolution.

4. A significant drop in the average value of MBH spins takes place at $z < 2$. This result is caused by the increasing number of dry MBH–MBH mergers at lower redshifts in the case of spheroids. Additionally, a dramatic drop is predicted at $z < 0.5$ for low-mass MBHs in gas-rich galaxies. This effect is due to low-mass, chaotic accretion events involving the capture of MCs.

5. In general, in gas-rich galaxies at $z < 1$, low-mass MBHs tend to spin slightly less rapidly than high-mass MBHs. The difference is less pronounced in gas-poor galaxies. As MBH mass increases, the distributions become more similar. At the highest masses ($>10^8 M_\odot$), the statistics are poor in the case of disk galaxies (between a couple and ~50 objects per bin), as the most massive MBHs tend to reside in gas-poor galaxies.

6. If outflows do not affect the angular momentum of nuclear disks and accretion proceeds coherently in both the quasar and decline phase, the spin distribution and its evolution do not differ very much for highly accreting MBHs and more quiescent MBHs. Differences are clearer in the population of the most massive BHs ($>10^8 M_\odot$).
7. If quasar feedback disrupts the nuclear disk feeding, the MBH and accretion proceeds chaotically in the decline phase. The spin distribution and its evolution shows a stronger dependence on mass, but not on morphology, again, except at the highest masses ($>10^8 M_\odot$) and $z < 1$.

The same comment on statistical significance discussed above applies here as well.

Qualitatively similar results have been obtained by Li et al. (2012) on observational grounds. They inferred the MBH spin evolution by tracing the evolution of the radiative efficiency of accretion flows, using the continuity equation for the MBH number density. Both our result and those of Li et al. seem to contradict the predictions of the “spin paradigm” scenario, according to which the jet production efficiency—and therefore, the radio loudness of AGNs—should reflect the MBH spin distribution and its evolution (Wilson & Colbert 1995; Hughes & Blandford 2003). Applying such a scenario to our results, one should expect the radio-loud fraction of AGNs to be much larger at high redshifts than in the present epoch. At least in the case of quasars, an opposite trend has been inferred, i.e., such a fraction has been suggested to decrease with redshift (Jiang et al. 2007), although Volonteri et al. (2011b) find that the radio-loud fraction is roughly constant with redshift for the most luminous sources ($L > 10^{47}$ erg s$^{-1}$). For high-redshift blazars powered by $M > 10^8 M_\odot$ MBHs, activity seems to peak around $z \sim 4$ (Volonteri et al. 2011b). Based on these indications, it may be that spin plays a more important role in powering jets at high accretion rates. Intermittent jet production in high accretion-rate AGNs (Sikora et al. 2007) may occur, similar to that observed in the Galactic micro-quasar GR5 1915+105 (Fender et al. 2004). However, this fact does not explain the finding that radio-loud quasars reside in more massive and denser environments than the radio-quiet quasars (Shen et al. 2009; Donoso et al. 2010).

Regarding low-accretion rate AGNs, given the similar MBH mass density in units of $4\pi \rho_{BH,z=0}/0.2 \times 10^5 M_\odot$ Mpc$^{-3}$ is the local ($z = 0$) MBH mass density in units of $4\pi \rho_{BH,z=0}/0.2 \times 10^5 M_\odot$ Mpc$^{-3}$ (Marconi et al. 2004). Using the integrated bolometric luminosity function from Hopkins et al. (2007), M. Gilfanov & A. Merloni (2013, in preparation) obtained $R \sim 0.075/\xi_{CT}$. Here, $\xi_{CT}$ is the mass density of BHs at the highest redshift probed by the bolometric luminosity function, $z \approx 6$, in units of the local mass density, $\xi_i$ encapsulates uncertainties on the process of MBH formation; $\xi_{CT}$ is the fraction of supermassive BH mass density (relative to the local density) grown in unseen, heavily obscured, Compton thick AGNs still missing from our census. Finally, $\xi_{lost}$ is the fraction of the BH mass density contained in “wandering” objects that have been ejected from a galaxy nucleus following, for example, a merging event and the subsequent production of gravitational waves, the net momentum of which could induce a kick capable of ejecting the BH from the host galaxy. The model presented in this paper allows us to provide an estimate of two of the unknowns in Equation (24): $\xi_i$ and $\xi_{lost}$. In our models, these unknowns are both of the order of 0.1 and they roughly cancel each other out. We can also directly estimate $\langle \epsilon \rangle = \langle \sum_{j=1}^{n} \epsilon_i \Delta M_{Y,j} / \sum_{j=1}^{n} \Delta M_{Y,j} \rangle = 0.13-0.18$ (the sum is done for all accreting MBHs starting from $z = 20$ down to $z = 0$, so it is an average over mass and time), leading to an estimate of $\xi_{CT} \sim 0.5$. Current estimates of the Compton thick AGN fraction (that are, however, not “mass weighted” as in the formalism of Equation (24); see also the note in Appendix A) based on local universe and models of the X-ray background range between 20% and 50% (e.g., Akylas et al. 2012).

We thank M. Dotti for helpful discussions and comments. M.V. acknowledges funding support from NASA through award ATP NNX10AC84G; from the NSF through award TM1-12007X, from the NSF through award AST 1107675, and from a Marie Curie Career Integration grant (PCIG10-GA-2011-303609). J.P.L. was supported in part by a grant from the French Space Agency CNES, by the Polish Ministry.
of Science and Higher Education within the project N N203 380336, and by the Polish National Science Centre through grant DEC-2012/04/A/ST9/00083.

APPENDIX A
BLACK HOLE GROWTH

We recall here how one can describe the growth of a MBH as a function of a constant or variable Eddington rate. We start by defining \( f_{\text{Edd}} = L/L_{\text{Edd}} \) and \( L_{\text{Edd}} = M c^2 / \dot{M}_{\text{Edd}} \), where \( \dot{M}_{\text{Edd}} = (\sigma_T c) / (4\pi G m_p) = 0.45 \) Gyr and \( f_{\text{Edd}} \) represents the Eddington fraction. Therefore, if the accretion rate is \( M_{\text{in}} \) and \( M \) is the mass that goes into increasing the MBH mass:

\[
L = \epsilon M_{\text{in}} c^2 = f_{\text{Edd}} L_{\text{Edd}} c^2
\]

and \( dM = (1 - \epsilon) dM_{\text{in}} \) and

\[
\frac{dM}{dt} = \frac{1 - \epsilon}{\epsilon} L_{\text{Edd}} f_{\text{Edd}} c^2 = \frac{1 - \epsilon}{\epsilon} f_{\text{Edd}} M c^2 / \dot{f}_{\text{Edd}}. \tag{A2}
\]

One therefore obtains

\[
\frac{dM}{M} = \frac{1 - \epsilon}{\epsilon} f_{\text{Edd}} \Delta t_{\text{Edd}} dt. \tag{A3}
\]

If \( \epsilon \) and \( f_{\text{Edd}} \) are constant over the time of integration, then

\[
\int M^{-1} dM = \int \frac{1 - \epsilon}{\epsilon} f_{\text{Edd}} \Delta t_{\text{Edd}} dt \tag{A4}
\]

and the MBH mass grows as

\[
M(t + \Delta t) = M(t) \exp \left( f_{\text{Edd}} \frac{\Delta t}{\dot{f}_{\text{Edd}}} \left( 1 - \frac{\epsilon(t)}{\epsilon(t)} \right) \right). \tag{A5}
\]

On the other hand, if for instance \( f_{\text{Edd}} \) is a function of time, one has to self-consistently integrate:

\[
\int M^{-1} dM = \int \frac{1 - \epsilon}{\epsilon} \Delta t_{\text{Edd}} f_{\text{Edd}}(t) dt, \tag{A6}
\]

as shown in Section 4.2.

Note that this mathematical formalism differs from the approximate form \( M_{\text{BH}} = (1 - \epsilon) M_{\text{in}} \Delta t \) (the two expressions agree in the limit \((1 - \epsilon) / \epsilon f_{\text{Edd}} \Delta t / \dot{f}_{\text{Edd}} \rightarrow 0\) and the mass-to-energy conversion efficiency appears within an exponential. This fact causes some inconsistency with the “standard” formalism used to evaluate the mass-to-energy conversion efficiency in Sołtan’s argument that adopts the simplified expression. At fixed \( \epsilon, f_{\text{Edd}}, \) and \( \Delta t \), the approximate expression underestimates the mass growth and, therefore, statistically, Sołtan’s argument tends to underestimate \( \epsilon \) with respect to our formalism.

APPENDIX B
ALIGNMENT OF BLACK HOLE SpINS IN ACCRETION DISKS

In this paper, we have assumed that most MBHs evolve in thin accretion disks where the importance of jets and magnetic fields is limited. In this case, warp propagation occurs diffusively (Bardeen & Petterson 1975; Papaloizou & Pringle 1983). In thick accretion disks \((H/R > \alpha)\), warp propagation occurs instead through bending waves (Nelson & Papaloizou 2000), while in magnetized disks with jets a “magneto-spin alignment” mechanism has been recently discovered in numerical simulations (McKinney et al. 2012).

We refer the reader to Nelson & Papaloizou (2000), Sorathia et al. (2013), and references therein for a full discussion of the mathematical treatment and the differences between diffusive and wave propagation; we summarize here the relevant information. Bardeen & Petterson (1975) showed that a viscous disk would be expected to relax to a form in which the inner regions become aligned with the equatorial plane of the BH (Lense–Thirring precession) out to a transition radius, beyond which the disk remains aligned with the outer disk. This fact is because the Lense–Thirring precession rate drops off sharply as the radius increases. The transition radius \( r_{\text{tr}} \) is expected to occur approximately where Lense–Thirring precession is balanced by the rate at which the warps are diffused or propagated away.

In the diffusive regime, the warping of the disk is counteracted by the diffusion of the warp, which acts over a diffusion timescale \( \tau_{\text{diff}} \sim r^2 c / \epsilon(H c^2/\Omega) \). In the bending wave regime, warps evolve on a sound crossing time \( \tau_{\text{wave}} \sim r/c_\epsilon \). The critical variable for determining the timescales over which (anti)alignment occurs is therefore the transition radius between the inner equatorial disk and the outer tilted disk. Nelson & Papaloizou (2000) performed a numerical parametric study of both regimes and they concluded that although the processes differ, the typical (anti)alignment timescale is well described by the formalism introduced by Rees (1978) and subsequently studied in greater detail by Scheuer & Feiler (1996) and Natarajan & Pringle (1998). Similar conclusions were recently obtained by Sorathia et al. (2013). We note, however, that Fragile & Anninos (2005), Fragile et al. (2007), and Dexter & Fragile (2011) find no alignment in their three-dimensional general relativistic magneto-hydrodynamic simulations of tilted disks. After Lense–Thirring precession causes the disk to warp, the propagation of the warp stops at the disk radius where the sound crossing time becomes shorter than the precession time and the disk remains tilted.

Finally, we remark that McKinney et al. (2012), also using three-dimensional general relativistic magneto-hydrodynamic simulations, recently proposed that in magnetized disks the frame-dragnning forces first cause the MBH magnetosphere to align with the MBH spin axis and then the disk loses the misaligned component of its angular momentum and reorients with the magnetosphere, at least at small radii (the outer disk may remain tilted; see their Table 2).

REFERENCES

Abramowicz, M. A., Chen, X., Kato, S., Lasota, J.-P., & Regev, O. 1995, ApJL, 438, L37
Akylas, A., Georgakakis, A., Georgantopoulos, I., Brightman, M., & Nandra, K. 2012, A&A, 546, A98
Barausse, E. 2012, MNRAS, 423, 2533
Bardeen, J. M. 1970, Natur, 226, 64
Bardeen, J. M., & Petterson, J. A. 1975, ApJL, 195, L65
Bardeen, J. M., Press, W. H., & Teukolsky, S. A. 1972, ApJL, 178, 347
Berti, E., & Volonteri, M. 2008, ApJL, 684, 822
Blandford, R. D., & Znajek, R. L. 1977, MNRAS, 179, 433
Bottema, R., & Sanders, R. H. 1986, A& A, 158, 297
Boylan-Kolchin, M., Ma, C.-P., & Quataert, E. 2008, MNRAS, 383, 93
Broderick, J. W., & Fender, R. P. 2011, MNRAS, 417, 184

Volonteri et al.
