Massive black holes and their hosts

M. Volonteri\textsuperscript{1}, S. Callegari\textsuperscript{2}, M. Colpi\textsuperscript{3}, M. Dotti\textsuperscript{1}, and L. Mayer\textsuperscript{2}

\textsuperscript{1} Astronomy Department, University of Michigan, 500 Church Street, Ann Arbor, MI, USA
\textsuperscript{2} Institute for Theoretical Physics, University of Z"urich, Winterthurerstrasse 190, CH-8057 Z"urich, Switzerland
\textsuperscript{3} Dipartimento di Fisica G. Occhialini, Universit`a degli Studi di Milano Bicocca, Piazza della Scienza 3, 20126 Milano, Italy

Abstract. Supermassive black holes are nowadays believed to reside in most local galaxies. I briefly review here some of the physical processes that are conducive to the evolution of the massive black hole population. I'll discuss black hole formation processes that are likely to take place at early cosmic epochs, and how massive black holes evolve in a hierarchical Universe. The mass of the black holes that we detect today in nearby galaxy has mostly been accumulated by accretion of gas. We discuss the timescales of accretion in galaxy mergers and how black hole growth relates to their hosts. While black hole–black hole mergers do not contribute substantially to the final mass of massive black holes, they influence the occupancy of galaxy centers by black hole, owing to the chance of merging black holes being kicked from their dwellings due to the “gravitational recoil”.

1. Introduction

Black holes (BHs), as physical entities, span the full range of masses, from tiny BHs predicted by string theory, to monsters weighing by themselves almost as much as a dwarf galaxy (massive black holes, MBHs). Notwithstanding the several orders of magnitude difference between the smallest and the largest BH known, all of them can be described by only three parameters: mass, spin and charge. Astrophysical BHs are even simpler system, as charge can be neglected as well. The interaction between astrophysical black holes and their environment is where complexity enters the game. I will focus here on the formation and evolution of MBHs, with masses above thousands solar masses, and how we believe their evolution is symbiotic with that of their host. Let’s start by recalling that MBHs in galaxy centers are far from being really “black”: we can easily trace the presence, as they are the engines powering the luminous quasars that have been detected up to high redshift. Nowadays we can detect in neighboring galaxies the dead remnants of this bright past activity. It is indeed well established that the centers of most local galaxies host MBHs with masses in the range $M_{\text{BH}} \sim 10^6 - 10^9 M_\odot$ (e.g., Ferrarese & Merritt 2000; Kormendy & Gebhardt 2001; Richstone et al. 1998). Observationally, the record for the smallest MBH belongs to the dwarf Seyfert 1 galaxy POX 52, which is thought to contain a MBH of mass $M_{\text{BH}} \sim 10^5 M_\odot$ (Barth et al. 2004). At the other end, however, the Sloan Digital Sky survey detected luminous quasars
at very high redshift, $z > 6$. At least some of these quasars are powered by supermassive black holes with masses $\sim 10^9 M_\odot$ (Barth et al. 2003). We are therefore left with the task of explaining the presence of very big MBHs when the Universe is less than 1 Gyr old, and of much smaller ones lurking in 14 Gyr old galaxies.

2. Scenarios for MBH formation

A single big galaxy can be traced back to the stage when it was split up in hundreds of smaller components with individual internal velocity dispersions as low as 20 km s$^{-1}$. Did black holes form with the same efficiency in small galaxies, or did their formation had to await the buildup of substantial galaxies with deeper potential wells? The formation of massive black holes is far less understood than that of their light, stellar mass, counterparts. The “flowing chart” presented by Rees 1978 still stands as a guideline for the possible paths leading to formation of MBH seeds in the center of galactic structures. One first possibility is the direct formation of a MBH from a collapsing gas cloud (Haehnelt & Rees 1993; Bromm & Loeb 2003; Begelman et al. 2006; Lodato & Natarajan 2006). In the most common situations, rotational support can halt the collapse before densities required for MBH formation are reached. Halos, and their baryonic cores, possess in fact angular momentum, believed to be acquired by tidal torques due to interactions with neighboring halos. The typical tidally induced angular momentum found in numerical simulations (e.g., Bullock et al. 2001; van den Bosch et al. 2002) would be enough to provide centrifugal support at a distance $\sim 20$ pc from the center, and halt collapse (Mo et al. 1998; Oh & Haiman 2002). Additional mechanisms inducing transport of angular momentum are needed to further condense the gas.

The loss of angular momentum can be driven either by (turbulent) viscosity or by global dynamical instabilities, such as the “bars-within-bars” mechanism (Shlosman et al. 1989; Begelman et al. 2006). The gas can therefore condense to form a central massive object, either a supermassive star, which eventually becomes subject to post-Newtonian gravitational instability and forms a seed MBH, or via a low-entropy star-like configuration where a small black hole forms in the core and grows by accreting the surrounding envelope. The masses of the seeds predicted by different models vary, but they are typically in the range $M_{\text{BH}} \sim 10^2 - 10^3 M_\odot$.

Alternatively, the seeds of MBHs can be associated with the remnants of the first generation of stars, formed out of zero metallicity gas. The first stars are believed to form at $z \gtrsim 10$ in halos which represent high-$\sigma$ peaks of the primordial density field. The main coolant, in absence of metals, is molecular hydrogen, which is a rather inefficient coolant. The inefficient cooling might lead to a very top-heavy initial stellar mass function, and in particular to the production of very massive stars with masses $> 100 M_\odot$. If very massive stars form above $260 M_\odot$, they would rapidly collapse to MBHs with little mass loss (Fryer et al. 2001), i.e., leaving behind seed MBHs with masses $M_{\text{BH}} \sim 10^2 - 10^3 M_\odot$ (Madau & Rees 2001; Volonteri et al. 2003).

2.1. Observational tests

What are the possible observational tests of MBH formation scenarios? Detection of gravitational waves from seeds merging at the redshift of formation (Sesana et al. 2007) is probably one of the best ways to discriminate among formation mechanisms. The planned Laser Interferometer Space Antenna (LISA) in principle is sensible to gravitational waves from binary MBHs with masses in the range $10^3 - 10^6 M_\odot$ basically at any redshift of interest. A large fraction of coalescences will be directly observable by LISA, and on the basis of the detection rate, constraints can be put on the MBH formation process. Theoretical models for the formation of MBH seeds and dynamical evolution of the binaries predict merger rates that largely vary one from the other (Sesana et al. 2007).

The imprint of different formation scenarios can also be sought in observations at
Fig. 1. The $M_{\text{bh}}$–velocity dispersion ($\sigma_c$) relation at $z = 0$. Every circle represents the central MBH in a halo of given $\sigma_c$. Observational data are marked by their quoted errorbars, both in $\sigma_c$, and in $M_{\text{bh}}$ (Tremaine et al. 2002). Left panel: direct collapse seeds, Population III star seeds. Top panels: fraction of galaxies at a given velocity dispersion which do not host a central MBH. Adapted from Volonteri, Lodato & Natarajan (2007).

3. MBH growth in galaxy mergers

Growth by accretion of gas is important in the build-up of the largest MBHs, though the relative importance of mergers and accretion will depend on the host galaxy properties come more pronounced. The comoving mass density, an integral constraint, is reasonably well determined out to $z = 3$ but is still poorly known at higher redshifts. The increasing area and depth of high redshift survey, especially in X-rays, will increase the strength of our constraints (Salvaterra et al. 2007).

In our neighbourhood, the best diagnostic of MBH formation mechanisms would be the measure of MBH masses in low-luminosity galaxies. This can be understood in terms of the cosmological bias. The progenitors of massive galaxies (or clusters of galaxies) have both a high probability of hosting MBH seeds (cfr. Madau & Rees 2001), and a high probability that the central MBH is not “pristine”, that is it has increased its mass by accretion, or it has experienced mergers and dynamical interactions. In the case of low-bias systems, such as isolated dwarf galaxies, very few of the high-$z$ progenitors have the deep potential wells needed for gas retention and cooling, a prerequisite for MBH formation. The signature of the efficiency of the formation of MBH seeds will consequently be stronger in isolated dwarf galaxies. Hence, MBH formation models are distinguishable at the low mass end of the MBH mass function, while at the high mass end the effect of initial seeds appears to be sub-dominant. The clearest signature of massive seeds, compared to Population III remnants, would be a lower limit of order the typical mass of seeds to the mass of MBHs in galaxy centers, as shown in Fig. 1. Additionally, the fraction of galaxies without a MBH increases with decreasing halo masses at $z = 0$. A larger fraction of low mass halos are devoid of central black holes for lower seed formation efficiencies. While current data in the low mass regime is still scant, future campaigns with the Giant Magellan Telescope or JWST are likely to probe this region of parameter space with significantly higher sensitivity.

lower redshifts (Volonteri, Lodato & Natarajan 2007). Since during the quasar epoch MBHs increase their mass by a large factor, signatures of the seed formation mechanisms are likely more evident at earlier epochs. A nice diagnostic compares the integrated co-moving mass density in MBHs to the expectations from Soltan-type arguments (F. Haardt, private communication), assuming that quasars are powered by radiatively efficient flows (for details, see Yu & Tremaine 2002; Elvis et al. 2002; Marconi et al. 2004). We compare the observational results to models that differ only with respect to the MBH formation scenario. We either assume that seeds are Population III remnants, or that seeds are formed via direct collapse with different efficiencies (Lodato & Natarajan 2006). While during and after the quasar epoch the mass densities in our theoretical models differ by less than a factor of 2, at $z > 3$ the differences become more pronounced. The comoving mass density, an integral constraint, is reasonably well determined out to $z = 3$ but is still poorly known at higher redshifts. The increasing area and depth of high redshift survey, especially in X-rays, will increase the strength of our constraints (Salvaterra et al. 2007).

In our neighbourhood, the best diagnostic of MBH formation mechanisms would be the measure of MBH masses in low-luminosity galaxies. This can be understood in terms of the cosmological bias. The progenitors of massive galaxies (or clusters of galaxies) have both a high probability of hosting MBH seeds (cfr. Madau & Rees 2001), and a high probability that the central MBH is not “pristine”, that is it has increased its mass by accretion, or it has experienced mergers and dynamical interactions. In the case of low-bias systems, such as isolated dwarf galaxies, very few of the high-$z$ progenitors have the deep potential wells needed for gas retention and cooling, a prerequisite for MBH formation. The signature of the efficiency of the formation of MBH seeds will consequently be stronger in isolated dwarf galaxies. Hence, MBH formation models are distinguishable at the low mass end of the MBH mass function, while at the high mass end the effect of initial seeds appears to be sub-dominant. The clearest signature of massive seeds, compared to Population III remnants, would be a lower limit of order the typical mass of seeds to the mass of MBHs in galaxy centers, as shown in Fig. 1. Additionally, the fraction of galaxies without a MBH increases with decreasing halo masses at $z = 0$. A larger fraction of low mass halos are devoid of central black holes for lower seed formation efficiencies. While current data in the low mass regime is still scant, future campaigns with the Giant Magellan Telescope or JWST are likely to probe this region of parameter space with significantly higher sensitivity.

3. MBH growth in galaxy mergers

Growth by accretion of gas is important in the build-up of the largest MBHs, though the relative importance of mergers and accretion will depend on the host galaxy properties
It is expected that quasar activity is triggered by galaxy mergers, via dynamical instabilities in the perturbed potential that cause gas to infall towards the galaxy centers and feed the central holes. During the black hole growth, the kinetic and radiative output from the AGN can become “too strong”, giving rise to “feedback”, that is, outflows that heat and energize the surrounding gas. The effect of “feedback” is twofold. First, it quenches star formation in the host. Second, it quenches accretion on the hole. Feedback is therefore expected to dictate the masses of black holes, the stellar content of galaxies, and the timescale over which this balance is achieved. Simulations of galaxy mergers including quasar activity probably grasp the fundamental link that ensues between holes and hosts during accretion episodes, the timescales and details are still unclear, though. We present a study on the dynamics of massive black holes in galaxy mergers, obtained from a series of high-resolution N-Body/SPH simulations, as an example of how the devil can be in the detail. The simulations cover a dynamical range of 5 orders of magnitude (in total) from the scale of 100 kpc down to 4 pc. In a first set of simulations, the merger is followed from the large scale \( \sim 100 \text{ kpc} \) down to \( \sim 200 \text{ pc} \) which is the force resolution limit. In a second run, the technique of particle splitting is used to follow the gas thermodynamics and the black hole evolution down to a force resolution of 4 pc.

The force limit is comparable to the gravitational sphere of influence of the MBHs only in the refined run: for this reason, the flow pattern around each MBH cannot be probed with enough accuracy to assess the magnitude of the accretion rate \( \dot{M}_{\text{BH}} \). Different recipes for \( \dot{M}_{\text{BH}} \) were applied, in order to extract information about the “possible” mass growth from the N-body/SPH run of a \( q = 1/4 \) merger with gas cooling and star formation.

Let’s first consider a standard Bondi–Eddington limited accretion recipe for \( \dot{M}_{\text{BH}} \). The density of the gas around each MBH is computed within a spherical radius twice our force softening (i.e., 200 pc in the less resolved simulations, and 4 pc in the high-resolution run). As reference speed for the gas in the Bondi formula we can consider either the thermal sound speed in the vicinity of the MBHs (\( \sim 15 \text{ k.ms}^{-1} \)), or a turbulent velocity for the gas (\( \sim 60 \text{ k.ms}^{-1} \)). Such turbulent motions are not resolved in the low resolution simulations and are thus simply considered as they are observed in circumnuclear discs of merger remnants Downes & Solomon 1998. Instead, in the high resolution simulation we measured turbulent velocities of such magnitude for the gas particles in the nuclear disc. They arise as a result of the final supersonic collision between the two gaseous galaxy cores and persist because of the inefficient dissipation implied by the effective equation of state of the nuclear gas subject to the starburst.

Figure 2 shows the MBH relative distance as a function of time (lower panel), and the corresponding mass growth of the MBH (upper panel). The black (red) solid line describes the mass of the primary (secondary) MBH as a function of time, obtained from the low-resolution simulations using the thermal sound speed in the Bondi formula. Both MBH masses grow initially not because of gas in-
flows stemming from the merger process, but simply because gas can be accreted inside a minimum volume fixed by the numerical resolution (200 pc). This volume is too large to allow for a realistic estimate of the gaseous mass effectively available, thus implying a fast MBH growth even for isolated galaxies. This resolution-dependent effect is usually avoided by tuning parameters in the prescriptions for accretion and feedback, in order to reproduce scaling relations such as the M-sigma (e.g. Springel et al. 2005). The long-dashed lines refer instead to accretion obtained by considering, for the same simulation, a typical turbulent sound speed. We see that mass growth occurs only after the tidal disturbances, excited during the merger, drive gas inflows towards the central regions of the interacting galaxies. In both cases the MBHs become with time equally massive, and the memory of the initial ratio $q_{BH}$ is lost. This is due to the relative importance of tides, that perturb more strongly the lighter galaxy.

At the end of the simulation, $\sigma$ can be measured to infer the mass that the final relic MBH would have in order to lie on the $M_{BH}$ versus $\sigma$ relation (shown as the shaded area in Figure 3). The two MBH histories discussed above imply masses that are in excess of the expected value by an order of magnitude, even well before the merger is completed. These histories imply that the BHs may self-regulate their mass through AGN feedback on the surroundings “during the course of the merger”, sliding along the $M_{BH}$ versus $\sigma$ relation.

The limitations related to the force resolution appear clear when considering the third history (dot-dashed lines of Figure 3). In this last case, we extracted information about the density profile and the effective sound speed of the gas at a much closer distance (comparable to the MBH gravitational sphere of influence) by rescaling results from a refined high resolution simulation of equal mass galaxies. These curves suggest that the MBHs do not undergo any major growth during the first two orbits, and only when the galaxies merge the MBH masses grow by more than an order of magnitude. The MBH mass ratio varies with time but does not seem to increase toward unity. The primary MBH is approaching and exceeding “from below” the $M_{BH}$ versus $\sigma$ expectation value, but in this case the final discrepancy is consistent with the scatter of the relation. The last episode of rapid mass growth may become self-regulated once feedback is included.

Given all these uncertainties on the MBH mass growth, we can depict three possible (speculative) paths in the $M_{BH}$ versus $\sigma$ relation. (1) **Adjustment**: Adjustment occurs if the mass growth time $\tau_{growth}$ of the individual MBHs is long compared to the timescale of the merger $\tau_{merger}$ (either because of turbulence or excess of angular momentum content in the gas near the MBHs). The AGN phase would occur when the MBHs are already in place at the center of the remnant galaxy, and self-regulation would result from balance between (thermal and/or mechanical) energy injection and gravitational energy of the surrounding gas. This can happen either before or after the final coalescence of the MBH binary. In the first case, i.e., before coalescence, a double nuclear pointlike source could be detected in a star-bursting environment at separations closer than $\approx 10$ pc. (2) **Symbiosis**: In this case, both accretion and the buildup of the final galaxy occur almost synchronously ($\tau_{growth} \approx \tau_{merger}$) with step-like episodes of AGN activity and starburst, leading to a “sliding” of $M_{BH}$ and $\sigma$ along the observed relation. In this case we may observe single or double AGN activity. If double, this would occur at separations from several kpc down to subparsec scales. (3) **MBH Dominance**: MBH growth is triggered by tidal torques during the early phase of the merger ($\tau_{growth} < \tau_{merger}$), resulting in a mass exceeding the relation at intermediate stages; in this case the assembly of the galaxy would be dictated by the MBH that may regulate through its (their) luminosity the last starburst and the shape of the final potential well determining the $\sigma$ of the remnant, though it is unclear how this could produce the higher velocity dispersion needed to satisfy the relation. In this scenario, luminous double AGN activity would be observed well before the two galaxies lose their identity, on very large scales.

Most theoretical efforts have been focused on the first scenario (e.g. Silk & Rees 1998,
Granato et al. 2004), and observations seem to find a low fraction of widely separated, interacting galaxies hosting nuclear activity, thus disfavoring – albeit with some uncertainties – MBH dominance as a general trend. On the other hand, numerical experiments on the scale of mergers with force resolution in the commonly employed range (at least an order of magnitude larger than the sphere of influence of a BH) indicate the third case as possible, and avoid/tune excessive MBH mass growth by invoking a certain amount of local AGN feedback acting on the surrounding gas Springel et al. 2005. However, assessing with enough confidence which one of the three paths is actually followed would require convergence of results when going to higher resolution. Instead, the striking differences in the MBH mass accretions shown in Figure 1 suggest that there is a strong degeneracy between numerical resolution and physical assumptions. In the absence of any feedback from the MBH itself, higher numerical resolution produces the same delay in the growth of MBHs as unresolved turbulence. The high resolution simulations indicate that most of the MBH growth occurs after the merger, which favors the "adjustment" scenario.

4. Accretion in circumnuclear discs

Collisions of gas rich spirals may trigger starbursts as those observed in luminous infrared galaxies. A large number of these galaxies host a central massive (up to $10^8 M_\odot$) gaseous disc extending on scales of $\sim 100$ pc (Downes & Solomon 1998). These discs may be the end-product of gas–dynamical, gravitational torques excited during the merger, when large amounts of gas is driven into the core of the remnant (Kazantzidis et al. 2005; Mayer 2005). This phase would follow naturally from the initial stages of galaxy mergers described in the previous section. Inside a massive self-gravitating disc, MBHs can continue their dynamical and accretion evolution.

Let us focus on two high-resolution simulations, that further refine the output of a merger simulation as those described in the previous section. Here the profile and dynamical state of the merger remnant of the simulations described above are re-simulated at much higher resolution, so that the inner sphere of influence of the MBHs is resolved. In these models, two MBHs are placed in the plane of a gaseous disc, embedded in a larger scale stellar spheroid. The gaseous disc is modeled with $2 \times 10^6$ particles, has as total mass $10^9 M_\odot$. The two MBHs are equal mass ($M_{BH} = 4 \times 10^6 M_\odot$). $M_1$ is placed at rest at the centre of the circumnuclear disc, while $M_2$ is initially orbiting in the plane of the disc on an orbit whose eccentricity is $\approx 0.7$, at a separation of 50 pc from $M_1$. $M_2$ can either be co- or counter-rotating with respect to the circumnuclear disc (runs A and B, respectively).

Gas particles are accreted onto the MBHs if the two following criteria are fulfilled:

- the total energy (kinetic + internal + gravitational) of the gas particle is lower than 7/10 of its gravitational energy (all the energies are computed with respect to each MBH).
- the total mass accreted onto a MBH every timestep is lower than the $M$ corresponding to an Eddington luminosity ($L_{Edd}$) assuming a radiative efficiency of 10%.

The spatial resolution of the hydrodynamical force in the highest density regions is $\approx 0.1$ pc. The gravitational softening for the gaseous particles and the MBHs is set at the same value to prevent numerical errors due to the different resolution. With this spatial resolution the influence radius of the holes ($\approx 1$ pc) is resolved, a condition necessary to assess gas accretion.

Let us compare the evolution of the two MBHs in runs A and B. In both runs, $M_1$ accretes at $f_{Edd} = M_1 / M_{Edd} \approx 0.5$, with a slight decrease with time. The accretion rate on $M_2$, instead, differs significantly. In run A, the accretion history of $M_2$ can be divided in two phases. First, for $t \leq 2.5$ Myrs $f_{Edd} \approx 0.3$ on average, showing strong variability while the orbit circularizes and the relative velocity between hole and gas is significant. After circularization, $M_2$ the relative velocity between $M_2$ and the gas is reduced. In this phase, $f_{Edd} \approx 0.45$ on average.

In the counter–rotating case, B, $f_{Edd}$ varies more significantly. We can still distinguish two phases: for $t \lesssim 3$ Myr, $M_2$ is still counter–
rotating, and $f_{\text{Edd}} \approx 0.1$. When the hole becomes co-rotating (angular momentum flip, see middle panel in Figure 3), $M_2$ accretes at $f_{\text{Edd}} \approx 0.45$. These simulations clearly show that the accretion rate is strongly influenced by the details of the dynamical evolution.

5. Gravitational recoil and ejections

Somehow counter-intuitively fast MBH mergers at very high redshift can bring an overall damage to the growth of the MBH population, rather than contribute to the build-up of more massive holes. This is due to the so-called “gravitational recoil”. When the members of a black hole 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. This recoil could be so violent that the merged hole breaks loose from shallow potential wells, especially in small mass pregalactic building blocks (Figure 4).

Comparing the recoil velocity to the escape velocity from their hosts, Volonteri 2007 find that the fraction of “lost” binaries decreases with increasing cosmic time due to a combination of (i) the mass ratio distribution becoming shallower, and, (ii) the hierarchical growth of the hosts. Schnittman 2007 shows in a very elegant way that, even for large recoils, the very hierarchical nature of structure evolution ensures that a substantial fraction of galaxies retain their MBHs, if evolution proceeds over a long series of mergers (see also Menou et al. 2001). As, especially at high redshift, binaries represent the exception, rather than the rule, the possible ejection of most binaries before $z \approx 5$ is not a threat to the evolution of the MBH population that has been detected in nearby galaxies.

Although the gravitational recoil does not damage to the evolution of the MBH population that we observe locally, it can be dangerous in very special cases. Haiman 2004 pointed out that the recoil can be indeed threatening the growth of the MBHs that are believed
to be powering the luminous quasars at $z \approx 6$ detected in the Sloan survey (e.g., Fan et al. 2001). In fact, in such a biased volume, the density of halos where MBH formation can be efficient (either by direct collapse, or via PopIII stars) is highly enhanced. The net result is an higher merger rate, and binarity is especially common for the central galaxy of the main halo. While the “average” MBH experiences at most one merger in its lifetime, a MBH hosted in a rare exceptionally massive halo can experience up to a few tens mergers, and the probability of ejecting the central MBH, halting its growth, is 50-80% at $z > 6$ (Volonteri & Rees 2006). This implies that MBHs at high redshift did not mainly grow via mergers.

References

Kormendy, J., & Gebhardt, K. 2001, ArXiv Astrophysics e-prints