Piero Madau

Department of Astronomy and Astrophysics, University of California, Santa Cruz, Ca 95060, USA e-mail: pmadau@ucolick.org

Abstract. I review scenarios for the assembly of supermassive black holes (MBHs) at the center of galaxies that trace their hierarchical build-up far up in the dark matter halo “merger tree”. Monte Carlo realizations of the merger hierarchy in a ΛCDM cosmology, coupled to semi-analytical recipes, are a powerful tool to follow the merger history of halos and the dynamics and growth of the MBHs they host. High-resolution hydrodynamics simulations of early structure formation in ΛCDM cosmologies are needed to track in detail the gas-phase H$_2$ “astrochemistry”, the thermal and ionization history of a clumpy intergalactic medium, and to guide studies of early reheating. X-ray photons from miniquasars powered by intermediate-mass “seed” holes may permeate the universe more uniformly than EUV radiation, make the low-density diffuse intergalactic medium warm and weakly ionized prior to the epoch of reionization breakthrough, enhance molecular cooling and increase the amount of dense material available for star formation. The spin distribution of MBHs is determined by gas accretion, and is predicted to be heavily skewed towards fast-rotating Kerr holes, already in place at early epochs, and not to change significantly below redshift 5. Decaying MBH binaries may shape the innermost central regions of galaxies and should be detected in significant numbers by LISA.

Key words. black hole physics – cosmology: theory – galaxies: active – methods: numerical – X-rays: general

1. Introduction

The strong link observed between the masses of supermassive black holes (MBHs) at the center of most galaxies and the gravitational potential wells that host them suggests a fundamental mechanism for assembling black holes and forming spheroids in galaxy halos. The $m_{\text{BH}}-\sigma$ relation (2000) implies a rough proportionality between MBH mass and the mass of the baryonic component of the bulge. It is not yet understood whether this relation was set in primordial structures, and consequently how it is maintained throughout cosmic time with such a small dispersion, or indeed which physical processes established such a correlation in the first place (2000) (2000) (2000).

In cosmologies dominated by cold dark matter (CDM) galaxy halos experience multiple mergers during their lifetime, with those between comparable-mass systems (“major mergers”) expected to result in the formation of elliptical galaxies (2000). Simple models in which MBHs are also assumed to grow during major mergers and to be present in every galaxy at any redshift – while only a fraction of them is “active” at any given time – have been shown to explain many aspects of the observed
evolution of quasars (2000) (2000) (2000). The coevolution of MBHs and their host galaxies in hierarchical structure formation scenarios gives origin to a number of of important questions, most notably:

- Did the first MBHs form in subgalactic units far up in the merger hierarchy, well before the bulk of the stars observed today? The seeds of the $z \sim 6$ quasars discovered in the Sloan Digital SkySurvey had to appear at very high redshift, $z \gtrsim 10$, if they are accreting no faster than the Eddington rate. In hierarchical cosmologies, the ubiquity of MBHs in nearby luminous galaxies can arise even if only a small fraction of halos harbor MBHs at very high redshift (2000).

- How massive were the initial seeds, and is there a population of relic pregalactic MBHs lurking in present-day galaxy halos? A clue to these questions may lie in the numerous population of ultraluminous off-nuclear (“non-AGN”) X-ray sources that have been detected in nearby galaxies (2000). Assuming isotropic emission, the inferred masses of these “ULXs” may suggest intermediate-mass black holes with masses $\gtrsim$ a few hundred $M_\odot$ (2000).

- Can coalescing MBH binaries at very high redshift be detected in significant numbers by the planned Laser Interferometer Space Antenna (LISA)? If MBHs were common in the past (as implied by the notion that many distant galaxies harbor active nuclei for a short period of their life), and if their host galaxies undergo multiple mergers, then MBH binaries will inevitably form in large numbers during cosmic history. MBH pairs that are able to coalesce in less than a Hubble time will give origin to the loudest gravitational wave events in the universe.

- Active galactic nuclei powered by supermassive holes keep the universe ionized at $z \lesssim 4$, structure the intergalactic medium (IGM), and probably regulate star formation in their host galaxies. Intermediate-mass holes accreting gas from the surrounding medium may shine as “miniquasars” at redshifts as high as $z \sim 20$. What is the thermodynamic effect of miniquasars on the IGM at early times?

- Besides their masses, astrophysical black holes are completely characterized by their spins, $S = aGm_{\text{BH}}/c$, $0 \leq a/m_{\text{BH}} \leq 1$. The spin of a MBH is expected to have a significant effect on its observational manifestation, such as the efficiency of converting accreted mass into radiation and the existence and direction of jets in active nuclei. What is the expected distribution of MBH spins and how does this evolve with cosmic time?

In this talk I will review some recent developments in our understanding of the assembly, growth, emission history, and environmental impact of MBHs from early epochs to the present. Unless otherwise stated, all the results shown below refer to the currently favoured (by a variety of observations) $\Lambda$CDM world model with $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, $h = 0.7$, $\Omega_\Lambda = 0.045$, $\sigma_8 = 0.93$, and $n = 1$.

2. MBHs as Population III remnants

The first stars in the universe must have formed out of metal-free gas, in dark matter “minihalos” of total mass $\gtrsim 5 \times 10^3 M_\odot$ (2000) condensing from the high-$\sigma$ peaks of the primordial density field at redshift $z = 20 - 30$. Numerical simulations of the fragmentation of primordial clouds in standard CDM theories all show the formation of Jeans unstable clumps with masses exceeding a few hundred solar masses; because of the slow subsonic contraction – a regime set up by the main gas coolant, molecular hydrogen – further fragmentation into sub-components is not seen, and a single very massive star forms from the inside out (2000) (2000) (2000).

At zero metallicity mass loss through radiatively-driven stellar winds or nuclear-powered stellar pulsations is expected to be negligible, and Population III stars will likely die losing only a small fraction of their mass (except for $100 < m_* < 140 M_\odot$). Nonrotating very massive stars in the mass window $140 \lesssim m_* \lesssim 260 M_\odot$ will disappear as pair-instability supernovae (2000), leaving no compact remnants and polluting the universe with the first heavy elements (2000) (2000). Stars with $40 < m_* < 140 M_\odot$ and $m_* > 260 M_\odot$ are predicted instead to collapse to black holes with
masses exceeding half of the initial stellar mass (2000). Barring any fine tuning of the initial mass function of Pop III stars, intermediate-mass black holes (IMBHs) – with masses above the 4–18 \( M_\odot \) range of known “stellar-mass” holes – may then be the inevitable end-product of the first episodes of pregalactic star formation (2000). Since they form in high-\( \sigma \) rare density peaks, relic IMBHs are expected to cluster in the bulges of present-day galaxies as they become incorporated through a series of mergers into larger and larger systems (see Fig. 1). The presence of a small cluster of IMBHs in galaxy nuclei may have several interesting consequences associated with tidal captures of ordinary stars (likely followed by disruption), capture by the central supermassive hole, and gravitational wave radiation from such coalescences. Accreting pregalactic IMBHs may be detectable as ultra-luminous, off-nuclear X-ray sources (2000).

3. The first miniquasars

Physical conditions in the central potential wells of young and gas-rich protogalaxies may have been propitious for black hole gas accretion. Perhaps seed black holes grew efficiently in small minihalos just above the cosmological Jeans mass (with shallow potential wells), or maybe gas accretion had to await the buildup of more massive galaxies (with virial temperatures above the threshold for atomic cooling). This issue is important for the detectability of high-\( \zeta \) miniquasars: it also determines whether the radiation background at very high redshifts had an X-ray component component able to preheat and partially ionize the IGM.

As mentioned above, gas condensation in the first baryonic objects is possible through the formation of \( \text{H}_2 \) molecules, which cool via roto-vibrational transitions down to temperatures of a few hundred kelvins. In the absence of a UV photodissociating flux and of ionizing X-ray radiation, three-dimensional simulations of early structure formation show that the fraction of cold, dense gas available for accretion onto seed holes or star formation exceeds 20\% for halos more massive than \( 10^6 \ M_\odot \) (2000). On the other hand, a zero-metallicity progenitor star in the range \( 40 < m_\star < 500 \ M_\odot \) emits about 70,000 photons above 1 ryd per stellar baryon (2000). The ensuing ionization front will completely overrun the host halo, photoevaporating most of the surrounding gas (2000). Black hole remnants of the first stars that created \( \text{H}_\text{II} \) regions are then unlikely to accrete significant mass until new cold material will be made available through the hierarchical merging of many gaseous subunits.

Accretion onto IMBHs may be an attractive way to (partially) reionize the low-density IGM (2000) (2000). A large fraction of the UV radiation from massive stars may not escape the dense sites of star formation, or may be deposited locally in halo gas that recombines almost immediately. The harder radiation emitted from miniquasars is instead more likely to escape from the hosts into intergalactic space, and may then produce more ‘durable’ (albeit partial) ionization in the diffuse IGM. High-resolution hydrodynamics simulations of...
Fig. 2. A 3D volume rendering of the IGM in the inner 0.5 Mpc simulated box at $z = 21$ and $z = 15.5$. Only gas with overdensity $4 < \delta < 10$ is shown: the locations of dark matter minihaloes are marked by spheres. The size and grey scale of the spheres indicate halo mass. At these epochs, the halo finder algorithm identifies 55 ($z = 21$) and 262 ($z = 15.5$) minihaloes within the simulated volume.

Early structure formation in ΛCDM cosmologies are a powerful tool to track in detail the thermal and ionization history of a clumpy IGM and guide studies of early reheating. We (2000) have used ENZO, an adaptive mesh refinement (AMR), grid-based hybrid (hydro+N-body) code developed by Bryan & Norman (see http://cosmos.ucsd.edu/enzo/) to solve the cosmological hydrodynamics equations and simulate the effect of a miniquasar turning on at very high redshift in a volume 1 Mpc on a side (comoving). We first identify in a low-resolution pure N-body simulation the Lagrangian volume of a resolved protogalactic halo with a total mass $7 \times 10^5 \, M_\odot$ at $z = 25$, above the cosmological Jeans mass. We then generate new initial conditions with an $128^3$ initial static grid that covers a 0.5 Mpc volume centered around the identified high-σ peak. During the evolution, refined grids (for a maximum of 5 additional levels) are introduced with twice the spatial resolution of the parent (coarser) grid to home in, with progressively finer resolution, on the densest parts of the “cosmic web”. The simulation follows the non-equilibrium chemistry of the dominant nine species (H, H+, H++, e, He, He++, He+++, H$_2$, and H$_2^+$) in primordial gas, and includes radiative losses from atomic and molecular line cooling.

At $z = 21$, a miniquasar powered by a 150 $M_\odot$ black hole accreting at the Eddington rate is turned on in the protogalactic halo. The miniquasar shines for a Salpeter time (i.e. down to $z = 17.5$) and is a copious source of soft X-ray photons, which permeate the IGM more uniformly than possible with extreme ultraviolet (EUV, $\geq 13.6 \, eV$) radiation (2000) and make it warm and weakly ionized prior to the epoch of reionization breakthrough (2000). A spectrum with $\nu L_\nu = \text{const}$ (like the nonthermal component observed in ULXs) was assumed for photons with energies in the range 0.2-10 keV, to which the simulation box is transparent. X-rays alone do not produce a fully ionized medium, but partially photoionize the gas by repeated secondary ionizations. A primary nonthermal photoelectron of energy $E = 1 \, \text{keV}$ in a medium with residual ionization (from the recombination epoch) $x = 2 \times 10^{-4}$ will create over two dozens secondary electrons, depositing a fraction $f_{\text{ion}} \approx 37\%$ of its initial energy as secondary ionizations of hydrogen, and only $f_{\text{heat}} \approx 13\%$ as heat (2000). Once the IGM ionized fraction increases to $x \approx 0.1$, the number of secondary ionizations per ionizing photon drops to a few, and the bulk of the primary’s energy goes into heat ($f_{\text{heat}} \approx 0.6$) via elastic Coulomb collisions with thermal electrons.
We find that, after one Salpeter timescale, the miniquasar has heated up the simulation box to a volume-averaged temperature of 2800 K. The mean electron and H$_2$ fractions are now 0.03 and $4 \times 10^{-5}$; the latter is 20 times larger than the primordial value, and will delay the buildup of a uniform UV photodissociating background. The net effect of the X-rays is to reduce gas clumping in the IGM by as much as a factor of 3. While the suppression of baryonic infall and the photoevaporation of some halo gas lower the gas mass fraction at overdensities $\delta$ in the range 20-2000, enhanced molecular cooling increases the amount of dense material at $\delta > 2000$. In many haloes within the proximity of our miniquasar the H$_2$-boosting effect of X-rays is too weak to overcome heating, and the cold and dense gas mass actually decreases. There is little evidence for an entropy floor preventing gas contraction and H$_2$ formation; instead, molecular cooling can affect the dynamics of baryonic material before it has fallen into the potential well of dark matter haloes and virialized. Overall, the radiative feedback from X-rays enhances gas cooling in lower-$\sigma$ peaks that are far away from the initial site of star formation, thus decreasing the clustering bias of the early pregalactic population, but does not appear to dramatically reverse or promote the collapse of pregalactic clouds as a whole.

4. Gravitational radiation from inspiraling MBH binaries

Frequent galaxy mergers will inevitably lead to the formation of MBH binaries. As dark matter halos assemble, MBHs get incorporated into larger and larger halos, sink to the center owing to dynamical friction, accrete a fraction of the gas in the merger remnant to become supermassive, form a binary system, and eventually coalesce (2000) (2000). In a stellar background a “hard” binary shrinks by capturing the stars that pass close to the holes and ejecting them at much higher velocities, a suprelastic scattering process that depletes the nuclear region and turns a stellar “cusp” into a low-density core (2000). Rapid coalescence eventually ensues due to the emission of gravitational radiation.

MBH binaries, with masses in the range $10^3 - 10^7$ M$_\odot$, are one of the primary targets for LISA (2000). Interferometers operate as all-sky monitors, and the data streams collect the contributions from a large number of sources belonging to different cosmic populations. To optimize the subtraction of resolved sources from the data stream, it is important to have a detailed description of the expected rate, duration, amplitude, and waveforms of events. Figure [3] shows the number of MBH binary coalescences per unit redshift per unit observed year predicted by (2000), using a detailed model of MBH binaries dynamics. The observed event rate is obtained by dividing the rate per unit proper time by the $(1 + z)$ cosmological time dilation factor. Each panel shows the rate for different $m_{BH} = M_1 + M_2$ mass intervals, and lists the integrated event rate, $dN/dt$, across the entire sky. The number of events per observed year per unit redshift peaks at $z = 2$ for $10^7 < m_{BH} < 10^9$ M$_\odot$, at $z = 3 - 4$ for $10^5 < m_{BH} < 10^7$ M$_\odot$, and at $z = 10$ for $m_{BH} < 10^5$ M$_\odot$, i.e. the lower the black hole mass, the higher the peak redshift. Beyond the peak, the event rate decreases steeply with cosmic time.

As recently shown by (2000), the GW signal from MBH binaries will be resolved (assuming a 3-year LISA observation) into 100 discrete events, 40 of which will be observed above threshold until coalescence. These “merging events” involve relatively massive binaries, $M \sim 10^5$ M$_\odot$, in the redshift range $2 \lesssim z \lesssim 6$. The remaining 60 events come from higher redshift, less massive binaries ($M \sim 5 \times 10^3$ M$_\odot$ at $z \gtrsim 6$) and, although their $S/N$ integrated over the duration of the observation can be substantial, the final coalescence phase is at too high frequency to be directly observable by LISA. The total number of detected events accounts for a fraction $\gtrsim 90\%$ of all coalescences of massive black hole binaries at $z \lesssim 5$. The residual confusion noise from unresolved massive black hole binaries is expected to be at least an order of magnitude below the estimated stochastic LISA noise.
Fig. 3. Number of MBH binary coalescences observed per year at $z = 0$, per unit redshift, in different $m_{BH} = M_1 + M_2$ mass intervals. Each panel also lists the integrated event rate, $dN/dt$, predicted by (2000). The rates (solid lines) are compared to a case in which triple black hole interactions are switched off (dotted lines). Triple hole interactions increase the coalescence rate at very high redshifts, while, for $10 < z < 15$, the rate is decreased because of the reduced number of surviving binaries. Dashed lines: rates computed assuming binary hardening is instantaneous, i.e. MBHs coalesce after a dynamical friction timescale.

5. MBH spins

The spin of a MBH is determined by the competition between a number of physical processes. Black holes forming from the gravitational collapse of very massive stars endowed with rotation will in general be born with non-zero spin (2000). An initially non-rotating hole that increases its mass by (say) 50% by swallowing material from an accretion disk may be spun up to $a/m_{BH} = 0.84$ (2000). While the coalescence of two non-spinning black holes of comparable mass will immediately drive the spin parameter of the merged hole to $a/m_{BH} \geq 0.8$ (2000), the capture of smaller companions in randomly-oriented orbits may spin down a Kerr hole instead (2000).

In (2000) we have made a first attempt at estimating the distribution of MBH spins and its evolution with cosmic time in the context of hierarchical structure formation theories, following the combined effects of black hole-black hole coalescences and accretion from a gaseous disk on the magnitude and orientation of MBH spins. Here I will briefly summarize our findings. Binary coalescences appear to cause no significant systematic spin-up or spin-down of MBHs: because of the relatively flat distribution of MBH binary mass ratios in hierarchical models, the holes random-walk around the spin parameter they are endowed with at birth, and the spin distribution retains significant memory of the initial rotation of “seed” holes.

It is accretion, not binary coalescences, that dominates the spin evolution of MBHs (Fig. 4). Accretion can lead to efficient spin-up of MBHs even if the angular momentum of the inflowing material varies in time. This is because, for a thin accretion disk, the hole is aligned with the outer disk on a timescale that is much shorter than the Salpeter time (2000), leading to accretion via prograde equatorial orbits. As a result, most of the mass accreted by the hole acts to spin it up, even if the orientation of the spin axis changes in time. For a geometrically thick disk, alignment of the hole with the outer disk is much less efficient, occurring on a timescale comparable to the Salpeter time. Even in this case most holes will be rotating rapidly. This is because, in any model in which MBH growth is triggered by major mergers, every accretion episode must typically increase a hole’s mass by about one e-folding to account for the local MBH mass density and the $m_{BH} - \sigma^*$ relation. Most individual accretion episodes thus produce rapidly-rotating holes independent of the initial spin.

Under the combined effects of accretion and binary coalescences, we find that the spin distribution is heavily skewed towards fast-rotating Kerr holes, is already in place at early epochs, and does not change significantly below redshift 5. About 70% of all MBHs are maximally rotating and have mass-to-energy conversion efficiencies approaching 30%. Note that if the equilibrium spin attained by accreting MBHs is lower than the value of $\hat{a} \equiv a/m_{BH} = 0.998$ used here, as in the thick disk MHD simulations of (2000), where $\hat{a} \approx 0.93$,
then the accretion efficiency will be lower as well, \( \approx 17\% \).

**Fig. 4.** Distribution of MBH spins in different redshift intervals. **Left panel:** effect of black hole binary coalescences only. **Solid histogram:** seed holes are born with \( \hat{a} \equiv a/m_{BH} = 0.6 \). **Dashed histogram:** seed holes are born non-spinning. **Right panel:** spin distribution from binary coalescences and gas accretion. Seed holes are born with \( \hat{a} = 0.6 \), and are efficiently spun up by accretion via a thin disk.

Even in the conservative case where accretion is via a geometrically thick disk (and hence the spin/disk alignment is relatively inefficient) and the initial orientation between the hole’s spin and the disk rotation axis is assumed to be random, we find that most MBHs rotate rapidly with spin parameters \( \hat{a} > 0.8 \) and accretion efficiencies \( \epsilon > 12\% \). As recently shown by (2000), a direct comparison between the local MBH mass density and the mass density accreted by luminous quasars shows that quasars have a mass-to-energy conversion efficiency \( \epsilon \geq 0.1 \) (a simple and elegant argument originally provided by (2000)). This high average accretion efficiency may suggest rapidly rotating Kerr holes, in agreement with our findings. Since most holes rotate rapidly at all epochs, our results suggest that spin is not a necessary and sufficient condition for producing a radio-loud quasar.

**Acknowledgements.** I would like to thank all my collaborators for their contributions to the ideas presented here. Support for this work was provided by NASA grant NNG04GK85G and by NSF grants AST-0205738.

**References**