

Chapter 1

THE FORMATION AND EVOLUTION OF THE FIRST MASSIVE BLACK HOLES

Zoltán Haiman

*Department of Astronomy, 1328 Pupin Laboratories
Columbia University, New York, NY 10027, USA **

zoltan@astro.columbia.edu

Eliot Quataert

*Department of Astronomy, 601 Campbell Hall
UC Berkeley, Berkeley, CA 94720, USA †*

eliot@astron.berkeley.edu

Abstract The first massive astrophysical black holes likely formed at high redshifts ($z \gtrsim 10$) at the centers of low-mass ($\sim 10^6 M_\odot$) dark matter concentrations. These black holes grow by mergers and gas accretion, evolve into the population of bright quasars observed at lower redshift, and eventually leave the supermassive black hole remnants that are ubiquitous at the centers of galaxies in the nearby universe. The astrophysical processes responsible for the formation of the earliest seed black holes are poorly understood. The purpose of this review is three-fold: (1) to describe theoretical expectations for the formation and growth of the earliest black holes within the general paradigm of hierarchical cold dark matter cosmologies, (2) to summarize several relevant recent observations that have implications for the formation of the earliest black holes, and (3) to look into the future, and assess the power of forthcoming observations to probe the physics of the first AGNs.

Keywords: Black holes, AGN

*Partial funding provided by NSF grants AST-0307200 and AST-0307291.

†Partial funding provided by NASA grant NAG5-12043, NSF grant AST-0206006, and an Alfred P. Sloan Fellowship.

1. Introduction

It seems established beyond reasonable doubt that some massive black holes (SMBHs) were fully assembled early in the history of the universe. The handful of bright quasars at $z \gtrsim 6$ are likely powered by holes as massive as $\sim 10^9 M_\odot$, and the spectra and metallicity of these objects appear remarkably similar to their counterparts at moderate redshifts (Fan et al. 2003). Indeed, if one selects individual quasars with the same luminosity, their properties show little evolution with cosmic epoch.¹

This implies that the behavior of individual quasars is probably determined by local physics near the SMBH, and is not directly coupled to the cosmological context in which the SMBH is embedded. However, it is clear that the quasar population as a whole does evolve over cosmic timescales. Observations from $0 \lesssim z \lesssim 6$ in optical (e.g. 2dF and SDSS) and radio bands (Shaver et al. 1994) show a pronounced peak in the abundance of bright quasars at $z \approx 2.5$. Recent X-ray observations confirm the rapid rise from $z = 0$ towards $z \approx 2$ (but have not shown evidence for a decline at still higher redshifts; Miyaji et al. 2000).

The cosmic evolution of quasar black holes between $0 \lesssim z \lesssim 6$ is likely driven by a mechanism other than local physics near the hole. This is reinforced by the fact that the timescale of activity of individual quasars is significantly shorter than cosmic timescales at $z \lesssim 6$ – both on theoretical grounds ($\sim 4 \times 10^7$ yr, the e-folding time for mass of a hole whose accretion is limited by its own (Eddington) luminosity, and converts mass to radiation with an efficiency of $\epsilon = \dot{M}c^2/L_{\text{Edd}} \sim 10\%$) and using the duty cycle of quasar activity inferred from various observations ($\sim 10^7$ yr, e.g. Martini 2003 and references therein; see also Ciotti, Haiman & Ostriker 2003).

In the cosmological context, it is tempting to link the evolution of quasars with that of dark matter halos condensing in a cold dark matter (CDM) dominated universe, as the halo population naturally evolves on cosmic time-scales (Efstathiou & Rees 1988). Indeed, this connection has proven enormously fruitful, and has resulted in the following broad picture: the first massive astrophysical black holes appear at high redshifts ($z \gtrsim 10$) in the shallow potential wells of low-mass ($\lesssim 10^8 M_\odot$) dark matter concentrations. These black holes grow by mergers and gas accretion, evolve into the population of bright quasars observed at lower redshift, and eventually leave the supermassive black hole remnants that are ubiquitous at the centers of galaxies in the nearby universe.

Nevertheless, many uncertainties about this scenario remain. Most importantly, the astrophysical process(es) responsible for the formation of the earliest seed black holes (and indeed for the presence of quasar

black holes at all redshift) remain poorly understood. In this review, we focus on the emergence of the first generation of black holes, though many of the important questions are quite general, and apply equally to subsequent generations of black holes. This review is organized as follows. In § 1.2, we summarize several relevant recent observations that have implications for early BHs. In § 1.3, we describe theoretical expectations for the formation and growth of these BHs within the paradigm of hierarchical cold dark matter cosmologies, and also discuss early black holes in the context of cosmological reionization. In § 1.4, we “zoom in” and consider the local physics issues for BH formation. In § 1.5, we look into the future, with the goal to assess the power of forthcoming observations to probe the physics of the first AGNs. We offer our conclusions in § 1.6.

2. Observational Constraints

In this section, we review several recent observations and their implications for the formation of black holes at high redshift.

2.1 High Redshift Quasars in the Sloan Survey

The most distant quasars to date have been discovered in the Sloan Digital Sky Survey (SDSS). This is perhaps somewhat surprising, since the SDSS is a relatively shallow survey ($i \sim 22$), capable of detecting only the rarest bright quasars at a redshift as high as $z \sim 6$. Nevertheless, the large solid-angle searched for high- z quasars to date (~ 2800 sq. degrees) has yielded a handful of such objects (Fan et al. 2000, 2001, 2003). The most important properties (for our purposes) of these sources are that they are probably powered by supermassive black holes as large as (few \times) $10^9 M_\odot$, and they appear to be indistinguishable from bright quasars at moderate ($z \sim 2 - 3$) redshifts, with similar spectra and inferred metallicity. In addition, a large reservoir of molecular gas is already present even in the most distant ($z = 6.41$) source (Walter et al. 2003).

In short, these SMBHs and their surroundings appear as “fully developed” as their lower z counterparts, despite the young age ($\lesssim 10^9$ yrs) of the universe at $z \gtrsim 6$. These rare quasars are likely harbored by massive ($\sim 10^{13} M_\odot$) dark matter halos that form out of $4 - 5\sigma$ peaks of the fluctuating primordial density field. The large halo mass follows directly from the space density of these sources (see below and in Haiman & Loeb 2001; another method to confirm the large halo masses is to study the expected signatures of cosmological gas infall onto such massive halos, as proposed by Barkana & Loeb 2003a). Indeed, the environment and

dynamical history of an individual massive DM halo at $z \sim 6$ and $z \sim 3$ can be similar; it is their abundance that evolves strongly with cosmic epoch. This is broadly consistent with the observations: the bright $z \sim 6$ quasars look similar to their $z \sim 3$ counterparts, but their abundance is much reduced (by a factor of ~ 20).

The fact that these quasars are so rare has important implications. First, they are likely to be the “tip of the iceberg”, and accompanied by a much more numerous populations of fainter quasars at $z \gtrsim 6$. Pushing the magnitude limits of future surveys should prove rewarding. The slope of the luminosity function is expected to be very steep at $i \sim 22$, and only weak constraints are available to date from the source counts (Fan et al. 2003), and from gravitational lensing (Comerford, Haiman & Schaye 2002; Wyithe & Loeb 2002a,b; Richards et al. 2003). Combining these two yields the strongest limit of $-d \log \Phi / d \log L \lesssim 3$ (Wyithe 2003). Second, the steep slope of the dark halo mass function implies that the masses of the host halos can be “measured” from the abundance quite accurately (see discussion in § 1.2.4 below). Conversely, since small changes in the assumed host halo mass results in large changes in the predicted abundance, large uncertainties will remain in other model parameters. In this sense, fainter, but more numerous quasars (or lack thereof) can have more constraining power for models that relate quasars to dark halos (see § 1.2.2).

The most striking feature, however, is the large BH mass already present at $z \sim 6$. This presents interesting constraints on the physics of the growth of these holes, and ways of fueling them (see § 1.4.2 below). In the rest of this section, we critically assess whether the inferred large BH masses are robust.

The masses of the black holes powering the SDSS quasars is inferred by assuming that (1) they shine at the Eddington luminosity with a bolometric correction identical to that of lower redshift quasars (this is justified by their similar spectra), and (2) they are neither beamed nor gravitationally lensed (both of these effects would make the quasars appear brighter). These assumptions lead to masses of $M_{\text{BH}} \approx (2 - 5) \times 10^9 M_{\odot}$ for the four $z > 6$ quasars known to date. These are reasonable assumptions, which have some empirical justification.

The hypothesis that the quasars are strongly beamed can be ruled out based on their line/continuum ratio. If the quasar’s emission was beamed into a solid angle covering a fraction f of 4π , it would only excite lines within this cone, reducing the apparent line/continuum ratio by a factor f . However, the SDSS quasars have strong lines. Haiman & Cen (2002) found that the line/continuum ratio of the $z = 6.28$ quasar SDSS 1030+0524 is about twice that of the median value in SDSS sample at

$z > 2.25$ (Vanden Berk et al. 2001). Willott et al. (2003) apply this argument to the MgII line of the $z = 6.41$ quasar SDSS J1148+5251, and reach a similar conclusion.

Another important uncertainty regarding the inferred BH masses is whether the SDSS quasars may be strongly magnified by gravitational lensing. The optical depth to strong lensing along a random line of sight to $z \sim 6$ is small ($\sim 10^{-3}$; e.g. Kochanek 1998, Barkana & Loeb 2000). Nevertheless, magnification bias can significantly boost the probability of strong lensing. If the intrinsic (unlensed) luminosity function at $z \sim 6$ is steep, and/or it extends to faint magnitudes, the probability of strong lensing for the SDSS quasars could be of order unity (Comerford, Haiman & Schaye 2002; Wyithe & Loeb 2002a,b). The overwhelming majority (more than 90%) of strong lensing events would be expected to show up as multiple images with separations at least as large as $0.3''$ (it is difficult to produce strong magnification without such multiple images even in non-standard lensing models; Kuhlen, Keeton & Haiman 2004). However, HST observations of the highest redshift quasars show no sign of multiple images for any of the $z \gtrsim 6$ sources down to an angle of $0.3''$ (Richards et al. 2003). Another argument against strong lensing comes from the large apparent size of the HII regions around the SDSS quasars (Haiman & Cen 2002). For example, the spectrum of the $z = 6.28$ quasar SDSS 1030+0524 shows transmitted flux over a $\sim 100\text{\AA}$ stretch of wavelength blueward of Lyman α , corresponding to a ~ 30 (comoving) Mpc ionized region around the source. Provided this source is embedded in a neutral IGM (the key assumption for this constraint which has some justification, see below), it is impossible for an intrinsically faint quasar to produce such a large HII region, even for a long source lifetime (Haiman & Cen 2002). White et al. (2003) derive a similar conclusion for the quasar J1148+5251, although this source could be magnified by a factor of \sim few by lensing (subject to the uncertainty of its actual redshift $z = 6.37 - 6.41$).

Finally, whether or not the SDSS quasars are shining at the Eddington limit is difficult to decide empirically. Vestergaard (2004) estimated Eddington ratios in a sample of high redshift quasars using an observed correlation between the size of the broad-line region and the luminosity of the quasar (the correlation is calibrated using reverberation mapping of lower redshift objects; e.g. Kaspi et al. 2000; Vestergaard 2002). She finds values ranging from ≈ 0.1 to $\gtrsim 1$, with the $z \gtrsim 3.5$ quasars having somewhat higher L/L_{EDD} than the lower redshift population. In particular, Vestergaard estimates $L/L_{EDD} \approx 0.3$ for two of the $z \gtrsim 6$ SDSS quasars. Given the uncertainties in these results, this is quite consistent with the assumption of near-Eddington accretion. Note further that in

an extended lower-redshift $0 < z < 1$ sample, Woo & Urry (2002) also find higher Eddington ratios towards $z = 1$, but this may represent a trend towards higher ratios at higher luminosity. Whether the trend is primarily with redshift or luminosity is an important question, but large scatter and selection effects presently preclude a firm answer.

Inferences about Eddington ratios at high redshifts can also be made by utilizing models of the quasar population as a whole. Such models typically assume Eddington luminosity at higher redshift, where fuel is thought to be readily available (Small & Blandford 1992; Haehnelt & Rees 1993). Several semi-analytical models for the quasar population (Haiman & Loeb 1998b; Haehnelt, Natarajan & Rees 1998; Valageas & Silk 1999; Haiman & Menou 2000; Kauffmann & Haehnelt 2000; Wyithe & Loeb 2003b; Volonteri et al. 2003) have found that Eddington ratios of order unity during most of the growth of the BH mass also yield a total remnant SMBH space density at $z = 0$ that is consistent with observations (Magorrian et al. 1998; Ferrarese & Merritt 2000; Gebhardt et al. 2000; Graham et al. 2001). Ciotti & Ostriker (2001) have modeled the behavior of an individual quasar and found that (provided fuel is available), the luminosity is near the Eddington value during the phases when the quasar is on. Despite these arguments, one cannot directly rule out the possibility that the SDSS quasars shine at a super-Eddington luminosity (theoretically, this is possible in the photon bubble models of Begelman 2002; see §4.1). We emphasize that if this were true, then the SDSS quasars would have to be caught at a special time: maintaining the observed luminosities for $\gtrsim 10^7$ years with a radiative efficiency of $\epsilon \equiv L/\dot{m}c^2 = 0.1$ would bring the BH masses up to values of $10^9 M_\odot$ anyway.

2.2 Chandra and Hubble Deep Fields

As discussed above, a relatively shallow but large survey, such as the SDSS, can discover only the rare AGNs at high redshift. To constrain populations models, deeper surveys that reveal the “typical” sources, are more advantageous. When completed, the SDSS will have delivered perhaps 10–20 $z > 6$ quasars, but not more – this is due to the paucity of quasars as bright as $i \sim 22$ at $z \sim 6$. Furthermore, the reddest SDSS filter (the z' band) extends only to $\lambda \approx 9500\text{\AA}$ making the survey insensitive to sources beyond $z \sim 6.5$. In comparison, deep X-ray observations with *Chandra* and *XMM* (and also near-IR observations with *HST*) could directly detect SMBHs to redshifts well beyond the horizon of SDSS, provided such SMBHs exist.

A SMBH at redshift $z = 10$, with a mass of $M_{\text{bh}} = 10^8 M_{\odot}$ (30 times lower than the masses of the $z \sim 6$ SMBHs in the SDSS), would have an observed flux of $\sim 2 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$ in the soft X-ray band (Haiman & Loeb 1999b), under the reasonable assumption that it shines at the Eddington luminosity, and that its emission has a spectral shape similar to a typical quasar near redshift $z \sim 2 - 3$ (with $\sim 3\%$ of the bolometric flux emitted in the range $0.5(1+z)\text{keV} < E < 2(1+z)\text{keV}$ for redshifts $5 < z < 10$; e.g. Elvis et al. 1994).

Semi-analytical models can be utilized to derive the number and redshift distribution of quasars at $z > 5$, by associating quasar activity with the dark matter halos that are present at these redshifts (Haiman & Loeb 1998b; Kauffmann & Haehnelt 2000; Wyithe & Loeb 2003b). The predictions of the simplest version of these semi-analytic models (Haiman & Loeb 1999b) now appear to be significantly higher (by about a factor of ~ 10) than the number of possible $z > 5$ quasar *candidates* (Barger et al. 2003; discrepancies had also been noted earlier by Mushotzky et al. 2000; Alexander et al. 2001 and Hasinger 2002). In the most up-to-date version of such a semi-analytical model, Wyithe & Loeb (2003b) find predictions consistent with (at most) only a few $z > 5$ sources in the CDF. The main difference in this updated model (which is what reduces the expected number of high- z quasars) is that the assumed quasar lifetime is increased from $\sim 10^6$ to $\sim 10^7$ years, and the scaling $M_{\text{bh}} \propto M_{\text{halo}}$ is modified to $M_{\text{bh}} \propto M_{\text{halo}}^{5/3}(1+z)^{5/2}$. These scalings arise due to the radiative feedback assumed to limit the BH growth in these models (see discussion in § 1.2.4 below).

More generally, the number (or, currently, the upper limit) of high-redshift ($z > 6$) sources detected in the *CDFs* will place the best constraints to-date on quasar evolution models at these high redshifts (although no systematic assessment of these constraints yet exists in a suite of models). The current best constraint comes from comparing model predictions to the dearth of high- z quasars in the optical *Hubble Deep Fields* (Haiman, Madau & Loeb 1999). The results of Haiman & Loeb (1999) demonstrate that the *CDF* constraints are superior to these. Models satisfying the upper limit from the *HDF* (e.g., by postulating that SMBHs do not exist, or are not fueled, in halos with circular velocities $v \lesssim 75 \text{ km/s}$), still result in significant over-predictions for the *CDFs*.

2.3 Shedding (Quasar) Light on the Accretion History

It has long been proposed that quasar activity is powered by accretion onto supermassive BHs (Salpeter 1964; Zel'dovich 1964; Lynden-Bell 1967). It has also been realized that the cumulative radiation output of all quasars translates into a significant amount of remnant BH mass, presumably to be found at the centers of local galaxies (Lynden-Bell 1967; Soltan 1982). Over the past few years, there has been renewed interest in this connection because we now have a good estimate for the total present day BH mass density (in addition to an estimate of the cumulative quasar light output). This estimate is allowed by the recent detection of SMBHs at the centers of several dozen nearby galaxies (Magorrian et al. 1998), and the tight correlation of their masses with the masses (Magorrian et al. 1998), velocity dispersions (Ferrarese & Merritt 2000; Gebhardt et al. 2000) and light profiles (Graham et al. 2001) of the spheroids of their host galaxies. The connection between quasar light output and remnant mass provides constraints on the accretion history of the SMBHs, and on the presence of a (yet undetected) population of very high-redshift AGNs.

Many studies have used this correlation to estimate the total present day BH mass density (e.g. Salucci et al. 1999; Haehnelt & Kauffmann 2000; Yu & Tremaine 2002, Haiman, Ciotti & Ostriker 2003). The simplest way to quickly estimate this quantity is to multiply the local spheroid mass density, $\Omega_{\text{sph}} = (0.0018^{+0.00121}_{-0.00085})h^{-1}$ (Fukugita et al. 1998) by the mean ratio $M_{\text{bh}}/M_{\text{spheroid}} = 10^{-2.9}$ (Merritt & Ferrarese 2001). The latter ratio is a factor of ~ 4 times smaller than its value of $10^{-2.28}$ in the original paper (Magorrian et al. 1998); the correction is due mostly to improved models that include velocity anisotropies (see also van der Marel 1997). This results (for $h = 0.72$) in $\rho_{\text{bh}} = 5 \times 10^5 \text{ M}_{\odot} \text{ Mpc}^{-3}$. The most sophisticated analysis to date is by Yu & Tremaine (2002), who utilize the tight $M_{\text{bh}} - \sigma$ relation, and the velocity function of early type galaxies measured in SDSS, to find $\rho_{\text{bh}} = (3 \pm 0.5) \times 10^5 \text{ M}_{\odot} \text{ Mpc}^{-3}$. An additional correlation is observed between BH mass and host circular velocity beyond the optical radius (Ferrarese 2002). This directly ties the BH to its dark matter halo, and, although the results depend strongly on the assumed halo profile, this allows a more refined modeling of the evolution of the quasar BH population (e.g. Wyithe & Loeb 2003b).

The remnant BH mass density has important implications for understanding the first AGN. Several authors have compared the cumulative light output of known quasars (which must be summed over all luminosities and redshifts) with the remnant BH mass density. In general,

this yields the average radiative efficiency of SMBH accretion, with the result $\epsilon \sim 10\%$. Note the paradigm shift: originally, the same analysis was used to argue that (then unobserved) SMBHs must be ubiquitous in local galaxies (Lynden-Bell 1969; Sołtan 1982). Immediately after the discovery of the local SMBHs, it appeared that there was *too much* local BH mass, which was taken to imply that most of the quasar accretion must occur in an optically faint phase (e.g. Haehnelt, Natarajan & Rees 1998). The revised, more accurate estimates of the local SMBH space density (decreased by a factor of four from the original value), appear consistent with the hypothesis that the optical quasar population has a mean radiative efficiency of $\epsilon \sim 10\%$. It then follows that most of the mass of SMBHs was accreted during the luminous quasar phase at $z \sim 2 - 3$, and only a fraction of the total ρ_{bh} could have been built during the formation of the earliest AGNs at $z > 6$. However, there are caveats to this argument: a large (and even dominant) contribution to the total mass from high- z is allowed if the radiative efficiency of the $z \sim 3$ population is as high as 20% (and if the high- z quasars remain undetectable either because they are intrinsically faint or obscured).

The above discussion has focused on optically luminous quasars. A significant fraction of BH growth may, however, occur via obscured objects, which show up in hard X-ray, IR or sub-mm observations, but not in the optical. This possibility is strongly suggested by models of the X-ray background, which require a factor of few more obscured AGN than unobscured AGN (see Fabian 2003 for a review). The precise fraction of BH growth that occurs in an optically obscured phase is uncertain by a factor of few. For example, the local black hole mass density has a factor of \sim two uncertainty, depending on which correlation is adopted with the black hole mass, and on how the correlation is extrapolated to the entire galaxy sample in the universe (Yu & Tremaine 2002; Ferrarese 2002b). This immediately allows for a comparable amount of obscured and unobscured accretion with a typical efficiency of 10%. If, however, most BHs are rapidly rotating, or magnetic torques are important at the last stable orbit (e.g., Gammie 1999) the mean efficiency could be significantly larger $\sim 40\%$. In this case most accretion may occur in an optically obscured phase. Hard X-ray (e.g., the NuSTAR telescope recently selected by NASA for Phase A study as a SMEX mission) and IR (e.g. *SIRTF*) observations are required to provide an unbiased view of the growth of SMBHs.

2.4 Local Black Holes as Fossils

As mentioned above, SMBHs appear ubiquitous in local galaxies, with their masses correlating with the global properties of their host spheroids. Several groups have noted the broad natural implication that the formation of the SMBHs and their host spheroids must be tightly linked (see, e.g. Monaco et al. 2000; Kauffmann & Haehnelt 2001; Granato et al. 2001; Ciotti & van Albada 2001; Cattaneo et al. 2003; Haiman, Ciotti & Ostriker 2003). Various independent evidence suggests that spheroids are assembled at high redshift ($z \sim 2$, see Cattaneo & Bernardi 2003 for the recent age determinations from the Sloan sample, and references to older work), which would be consistent with most of the SMBH mass being accreted around this redshift (coinciding with the peak of the activity of luminous quasars, as discussed in § 1.2.3 above). This then has the unwelcome (but unsurprising) implication that the local SMBHs may contain little direct evidence of the formation of their seeds at $z > 6$. Indeed, it seems most plausible that the observed tight correlations, such as between M_{bh} and σ , are established by a feedback process which operates when most of the BH mass is assembled. However, the significance of this hypothesis is that – with the identification of a specific feedback mechanism – physically motivated extrapolations can be made towards high redshift.

Whether the local $M_{\text{bh}} - \sigma$ relation holds at higher redshift is also an interesting *observational* question; both in normalization and slope (as discussed by several authors), and also in its range (which has received less attention, but see Netzer 2003 and discussion below). The highest redshift SDSS quasars do appear to approximately satisfy the $M_{\text{bh}} - \sigma$ relation of the local SMBHs. If M_{bh} is estimated assuming the Eddington luminosity, and σ is estimated from the circular velocity of the host dark matter halos with the right space density (e.g. Haiman & Loeb 2001), then the SDSS quasars are within the scatter of the $M_{\text{bh}} - \sigma$ relations of Gebhardt et al. (2000) and also of Ferrarese (2002). As explained in § 1.2.1, the mass inference is reasonable. The determination of the halo mass and circular velocity from the observed abundance of quasars is also more robust than it may first appear. This is because despite the dependence on the poorly known duty cycle, the halo mass function is exponentially steep for the massive $M \sim 10^{13} M_{\odot}$ halos at $z \sim 6$; therefore the dependence of the inferred halo mass on the duty cycle (and other uncertainties in the estimated halo abundance) is only logarithmic. The weakest link in the argument is to associate the spheroid velocity dispersion with the circular velocity of the dark matter halo. Ferrarese (2002) shows evidence of a correlation between M_{bh} and σ with the

velocity dispersion measured in the dark matter dominated region of SMBH–host galaxies; this establishes a direct link to the dark halo and puts the above argument on somewhat firmer ground (although there is still large errors in the inferred correlation, depending on the halo profile one adopts to convert the measured circular velocity to total halo mass).

The (tentative) evidence that high redshift AGN also satisfy the $M_{\text{bh}} - \sigma$ relation further supports the idea that the formation of SMBHs and their host galaxy must be tightly coupled by cosmology–independent physical processes (since the SDSS quasars are the rare peaks that have already formed at $z \sim 6$ instead of at $z \sim 2$). Netzer (2003) raises the point that besides the slope and normalization of the $M_{\text{bh}} - \sigma$ relation, the *range* (of masses and velocity dispersions) over which observed galaxies satisfy this relation, has to match between low and high redshift. In particular, the largest $\gtrsim 10^{10} M_{\odot}$ BHs observed at high redshift should also exist at low- z , but have not yet been discovered.

There have been several suggestions in the literature for the nature of the dynamical coupling between the formation of the BH and its spheroid host. The most promising is probably radiative or mechanical feedback from the SMBH on the gas supply in the bulge (Silk & Rees 1998; Haehnelt, Natarajan & Rees 1998; Blandford 1999; King 2003; Wyithe & Loeb 2003b). The essential idea is that when the BH in the center of the galaxy grows too large its outflows and radiation unbind the gas in the bulge or in the disk, quenching further BH growth via accretion and further star formation. Competition with star–formation for the gas supply may also play a role (Di Matteo et al. 2003). Note that these mechanisms can readily work at any redshift.

Alternative possibilities for the origin of the $M_{\text{bh}} - \sigma$ relation include: (1) filling the dark matter loss cone (Ostriker 2001). This model runs into difficulties with the Sołtan argument discussed in §2.2; since the SMBHs are fed mostly dark matter rather than gas there is no associated radiation, (2) direct capture of stars on high–eccentricity orbits by the SMBH (Zhao, Haehnelt & Rees 2002, Merritt & Poon 2003); this model has a similar problem because black holes more massive than $\gtrsim 10^8 M_{\odot}$ don’t tidally disrupt stars so there is again no radiative output associated with the BH growth, (3) stellar captures by the accretion disk feeding the hole (Miralda–Escudé & Kollmeier 2003).

3. First Structure Formation

In this section, we sketch some basic theoretical arguments relevant to the formation of structure in the universe; we then discuss formation mechanisms for SMBHs.

3.1 Cosmological Perturbations as the Sites of the First Black Holes

Recent measurements of the cosmic microwave background (CMB) temperature anisotropies by the Wilkinson Microwave Anisotropy Probe (*WMAP*), determinations of the luminosity distance to distant type Ia Supernovae (SNe), and other observations have led to the emergence of a robust “best-fit” cosmological model with energy densities in cold dark matter (CDM) and “dark energy” of $(\Omega_M, \Omega_\Lambda) \approx (0.3, 0.7)$ (e.g. Spergel et al. 2003).

The growth of density fluctuations, and their evolution into non-linear dark matter structures can be followed in this cosmological model in detail from first principles by semi-analytic methods (Press & Schechter 1974; Sheth et al. 2001). More recently, it has become possible to derive accurate dark matter halo mass functions directly in large cosmological N -body simulations (Jenkins et al. 2001). Structure formation in a CDM dominated universe is “bottom-up”, with low-mass halos condensing first. Dark matter halos with the masses of globular clusters, 10^5 – $10^6 M_\odot$, are predicted to have condensed from $\sim 3\sigma$ peaks of the initial primordial density field as early as $\sim 1\%$ of the current age of the universe, or at redshift $z \sim 25$.

It is natural to identify these condensations as the sites where the first astrophysical objects, including the first AGNs, were born. The nature of the objects that form in these early dark matter halos is currently one of the most rapidly evolving research topics in cosmology.

3.2 Chemistry and Gas Cooling at High Redshift

Baryonic gas that falls into the earliest nonlinear dark matter halos is shock-heated to the characteristic virial temperatures of a few hundred Kelvin. It has long been pointed out (Rees & Ostriker 1977; White & Rees 1978) that such gas needs to lose its thermal energy efficiently (within a \sim dynamical time) in order to continue contracting, or in order to fragment – in the absence of any dissipation, it would simply reach hydrostatic equilibrium, and would eventually be incorporated into a more massive halo further down the halo merger hierarchy. While the formation of non-linear dark matter halos can be followed from first-principles, the cooling and contraction of the baryons, and the ultimate formation of stars or black holes (BHs) in these halos, is much more difficult to model ab-initio.

The gas content of a cosmological perturbation can contract together with the dark matter only in dark halos above the cosmological Jeans

mass, $M_J \approx 10^4 M_\odot [(1+z)/11]^{3/2}$, in which the gravity of dark matter can overwhelm thermal gas pressure. In these early, chemically pristine clouds, radiative cooling is dominated by H_2 molecules. As a result, gas phase H_2 “astro-chemistry” is likely to determine the epoch when the first AGNs appear (the role of H_2 molecules for early structure formation was reviewed by Abel & Haiman 2001). Several papers constructed complete gas-phase reaction networks, and identified the two possible ways of gas-phase formation of H_2 via the H^- or H_2^+ channels. These were applied to derive the H_2 abundance under densities and temperatures expected in collapsing high-redshift objects (Hirasawa 1969; Matsuda et al 1969; Palla et al. 1983; Lepp & Shull 1984; Shapiro & Kang 1987; Kang et al. 1990; Kang and Shapiro 1992; Shapiro, Giroux & Babul 1994). Studies that incorporate H_2 chemistry into cosmological models and address issues such as non-equilibrium chemistry, dynamics, or radiative transfer, have appeared only relatively more recently. Haiman, Thoul & Loeb (1996) used spherically symmetric simulations to study the masses and redshifts of the earliest objects that can collapse and cool via H_2 ; their findings were confirmed by semi-analytic treatment in Tegmark et al. (1997). The first three-dimensional cosmological simulations that incorporate H_2 cooling date back to Gnedin & Ostriker (1996, 1997) and Abel et al. (1997).

The basic picture that emerged from these papers is as follows. The H_2 fraction after recombination in the smooth ‘protogalactic’ gas is small ($x_{H_2} = n_{H_2}/n_H \sim 10^{-6}$). At high redshifts ($z \gtrsim 100$), H_2 formation is inhibited even in overdense regions because the required intermediaries H_2^+ and H^- are dissociated by cosmic “microwave” background (CMB, but with the typical wavelength in the infrared) photons. However, at lower redshifts, when the CMB photons redshift their energies, the intermediaries survive, and a sufficiently large H_2 abundance builds up inside collapsed clouds ($x_{H_2} \sim 10^{-3}$) at redshifts $z \lesssim 100$ to cause cooling on a timescale shorter than the dynamical time. Sufficient H_2 formation and cooling is possible only if the gas reaches temperatures in excess of ~ 200 Kelvin; or masses of few $\sim 10^5 M_\odot [(1+z)/11]^{-3/2}$. The efficient gas cooling in these halos suggests that the first nonlinear objects in the universe were born inside $\sim 10^5 M_\odot$ dark matter halos at redshift $z \sim 20$ (corresponding to a $\sim 3\sigma$ peak of the primordial density peak).

The behavior of metal-free gas in such a cosmological “minihalo” is a well posed problem that has recently been addressed in three dimensional numerical simulations (Abel, Bryan & Norman 2000, 2002; Bromm, Coppi & Larson 1999, 2002). These works have been able to follow the contraction of gas to much higher densities than previous studies. They have shown convergence toward a temperature/density

regime of $T \sim 200$ K, $n \sim 10^4 \text{ cm}^{-3}$, dictated by the critical density at which the excited states of H_2 reach equilibrium, and cooling becomes less efficient (Galli & Palla 1998). The 3D simulations suggest that the mass of the gas does not fragment further into clumps below sizes of $10^2 - 10^3 M_\odot$, but rather it forms unusually massive stars. Such stars would naturally leave behind black hole seeds, which can subsequently grow into by mergers and accretion into the SMBHs. Interestingly, massive stars have an “either/or” behavior. Nonrotating stars with masses between $\sim 40 - 140 M_\odot$ and above $\sim 260 M_\odot$ collapse directly into a BH without an explosion, and hence without ejecting their metal yields into the surrounding medium, whereas stars in the range $\sim 140 - 260 M_\odot$ explode without leaving a remnant (Heger et al. 2003). This dichotomy is especially interesting because early massive stars are attractive candidates for polluting the IGM with metals at high redshift (Madau et al. 2001). It is likely that the first stars had a range of masses, in which case they could contribute to both metal enrichment and to the seed BH population, with a relative fraction that depends sensitively on their initial mass function (IMF).

3.3 Cosmological Reionization: Do the First Black Holes Contribute?

Perhaps the most conspicuous effect of the first generation of light sources, once they collectively reach a critical emissivity of ionizing radiation, is the reionization of the intergalactic medium (IGM). As has long been known, the absence of strong HI Lyman α absorption (i.e., a so-called Gunn–Peterson trough, Gunn & Peterson 1965, hereafter GP) in the spectra of distant sources implies that the IGM is highly ionized (with volume averaged neutral fractions $\lesssim 10^{-4}$) at all redshifts $z \lesssim 6$. There have been two observational break-throughs recently. On the one hand, there is good evidence, from the strong absorption in the spectra of the highest redshift SDSS quasars, that the transition from a neutral to a highly ionized state of the IGM is occurring close to $z \sim 6$ (Becker et al. 2001; Fan et al. 2003; White et al. 2003).² On the other hand, the recent detection of a large electron scattering optical depth by the *WMAP* satellite implies that significant ionization had taken place at much higher redshifts ($z \sim 15$, Spergel et al. 2003). There is currently a flurry of activity trying to interpret these results in the context of reionization models (see Haiman 2003 for a recent review); and the electron scattering optical depth measured by *WMAP* still has a significant uncertainty, $\tau = 0.17 \pm 0.04$ (Kogut et al. 2003; Spergel et al. 2003). Nevertheless, these developments bring into sharp focus an interesting

“old” question: could AGNs have contributed to the reionization of the IGM? A natural follow-up question would then be: can we use reionization as a probe of the earliest AGNs? In this section, we highlight the need for an early population of ionizing sources, and assess whether these could be early AGNs. We start with a critical review of the recent observations and their implications.

Reionization and the Gunn–Peterson Troughs.

At the time of this writing, there are four known quasars at $z > 6$ (Fan et al. 2003). This redshift appears to coincide with the tail end of the epoch of reionization. In the spectra of a \sim dozen bright quasars at $z > 5$, the amount of neutral hydrogen absorption increases significantly (by about an order of magnitude) from $z \sim 5.5$ to $z \sim 6$ (McDonald & Miralda–Escudé 2001; Cen & McDonald 2002; Fan et al. 2002; Lidz et al. 2002), with the highest redshift quasars showing full Gunn–Peterson troughs consistent with no transmitted flux.³ Assuming photo-ionization equilibrium, this corresponds to a sharp increase, by at least an order of magnitude, in the H-ionizing background radiation from $z \sim 6$ to $z \sim 5.5$.

The actual numerical limit on the mean mass (volume) weighted neutral fraction at $z \sim 6$ is only $x_{\text{HI}} > 10^{-2}$ ($x_{\text{HI}} > 10^{-3}$), which does not directly establish that we are probing the neutral epoch of the IGM with $x_{\text{HI}} \sim 1$. However, the observed steep evolution of the ionizing background suggests that this is the case. Note that the ionizing background scales as $J \propto \epsilon \lambda$, where ϵ is the emissivity (per unit volume), and λ is the mean free path of ionizing photons. Specifically, the increase from $z \sim 6$ to $z \sim 5.5$ is much steeper than the evolution of the emissivity of the known galaxy population (e.g. Madau & Pozzetti 2000). The emissivity of the bright quasar population evolves faster, but they cannot account for the background needed to cause the observed high level of ionization (Shapiro, Giroux & Babul 1994; Haiman, Abel & Madau 2001). Alternatively, the observed steep evolution of J could come from a rapid evolution of the mean free path. At lower redshifts, the mean free path is dominated by the poorly known abundance of Lyman limit systems (Haardt & Madau 1996), and evolves rapidly. Nevertheless, this evolution, scaling approximately as $\lambda_{\text{mfp}} \propto (1+z)^{-6}$ (Cen & McDonald 2002), still accounts only for a factor of \sim two change in J from $z \sim 5.5$ to $z \sim 6$.

It therefore appears that the steep evolution of J seen in the SDSS quasar spectra is difficult to understand without invoking some additional physical effect(s). On the other hand, the steep evolution would be naturally expected if we were detecting the end of the reionization epoch. Before the discrete ionized bubbles around isolated sources first

overlap, each hydrogen atom is exposed at most only to a single (or a few, if the sources are clustered) ionizing sources. However, after the bubbles percolate, the mean free path to ionizing photons undergoes a sudden drop, and the background seen by a typical atom will sharply increase to a sum over the entire “Olbers’ integral”, dominated by numerous distant sources (e.g. Haiman & Loeb 1999; Gnedin 2000). The time-scale for the buildup of the background, the light-crossing time across individual HII bubbles, is a small fraction of the Hubble time. This would be the most economical explanation of the observed steep evolution of J .

Note that Songaila & Cowie (2002) reach a different conclusion, and find that the spectra are consistent with a constant ionizing background over this redshift interval. The strongest conclusions about the ionizing background can be drawn from the amount of Lyman γ and β absorption. Because of the smaller oscillator strengths of these higher transitions, the IGM is less opaque in these lines than in the corresponding Lyman α line. Hence, these higher lines yield a stronger lower limit on the neutral fraction, and a stronger upper limit on the ionizing background. It appears that the differences between Songaila & Cowie’s results and that of the other groups are in the treatment of these higher lines. Given the significance of whether or not we are detecting the reionization epoch, this will be important to clarify in future work.

Reionization and the CMB.

The free electrons produced by reionization scatter a fraction of the background CMB photons, offering an alternative way of probing the reionization history (in particular at redshifts much higher than one can currently take a spectrum of an individual source). The scattering of the photons damps the temperature fluctuations on small angular scales (i.e. on scales below the apparent size of the horizon at reionization, corresponding to spherical index $\ell \gtrsim 10$; this is a purely geometrical effect), but boosts the “primordial” polarization signal at large angles ($\ell \lesssim 10$; see Zaldarriaga 1997). The effect on the temperature anisotropies is essentially an overall suppression of power on the scales where it is measurable, and it is therefore unfortunately nearly degenerate with the intrinsic amplitude of the fluctuation power spectrum. However, Kaplinghat et al. (2003) showed that one can break this degeneracy, and detect the reionization signature, in the polarization power spectrum. It is indeed in the “TE” (temperature–polarization angular cross–correlation) map that the *WMAP* team discovered evidence of early reionization, and measured the value $\tau = 0.17 \pm 0.04$ for the electron scattering optical depth (Spergel et al. 2003; Kogut et al. 2003). Assuming a single-step

reionization (sudden transition from a neutral to a fully ionized IGM at redshift z_r), this translates to a reionization redshift of $z_r = 17 \pm 4$.

Reionization and Early Black Holes.

The above *WMAP* result is inconsistent at the 3σ level with a sudden percolation of HII bubbles occurring at $z \sim 6$, which would correspond to the low optical depth of $\tau = 0.04$. This discrepancy is reduced (to the $\sim 2\sigma$ level) even in the simplest models for reionization, in which the ionizing emissivity traces the collapse of DM structures. With a reasonable choice of efficiency parameters in such a model, percolation indeed occurs around $z \sim 6$, satisfying the GP trough detections. In such models, there is a natural “tail” of partial ionization, extending to redshifts beyond the percolation epoch, which predicts the total $\tau \sim 0.08$ (Haiman & Holder 2003; Ciardi et al. 2003). However, if the high value of $\tau = 0.17$ is confirmed in future CMB polarization data (e.g., by several additional years of *WMAP* data), the implication will remain: there are additional sources of ionizing radiation at $z \sim 15$. Most importantly, with further improved CMB polarization measurements by *Planck*, the reionization history at high redshifts can be mapped out to high precision (Kaplighat et al. 2003; Holder et al. 2003).

The emissivity of the bright optical quasar population drops steeply at high redshift ($z \gtrsim 3$; e.g. Fan et al. 2002). There is a hint that the evolution towards high redshift is flatter in X-rays (Miyaji et al. 2000). While this could be explained if optical quasars were selectively more dust-obscured at high redshift, this interpretation would fail to explain the sharp decline towards high z that is also seen in the radio (Shaver et al. 1996). If the sharp decline is real, it is easy to show that quasars don’t contribute significantly to the ionizing background at $z \gtrsim 6$ (Madau, Haardt & Rees 1999; Haiman, Abel & Madau 2001; Fan et al. 2001), and thus cannot account for the GP troughs detected in the SDSS quasars at this redshift.

However, there is, at least in principle, still room for AGNs to contribute to reionization. First, the above ignores the possible presence of faint “miniquasars” (a terminology introduced by Haiman, Madau & Loeb 1999) below the current detection thresholds. It has been pointed out (Haiman & Loeb 1998b; Haehnelt, Natarajan & Rees 1998) that there could be a significant population of such faint quasars, and that their expected abundance depends crucially on the duty cycle of quasar activity. If the quasar lifetime is short ($\lesssim 10^7$ years), then quasars must reside in intrinsically abundant, low-mass halos, in order to match their observed surface density on the sky. Conversely, if quasars are long-lived, they must be harbored by the rarer, more massive halos (for the same apparent abundance). The abundance of low-mass halos de-

clines less rapidly (and can even increase for $M_{\text{halo}} < M_*$) towards high redshifts, and therefore if the quasar duty cycle is short, one expects a larger number of yet-to-be detected “miniquasars”. Quasar lifetimes are currently uncertain, but are constrained to lie in the range $10^6 - 10^8$ years (see the review by Martini 2003). A particularly relevant method to obtain the lifetime (and thus host halo mass) for the typical quasar at a fixed luminosity is to study the spatial clustering of quasars in large surveys such as 2dF or SDSS (Haiman & Hui 2001; Martini & Weinberg 2001). Current results from 2dF favor $t \lesssim 10^7$ years (see Croom et al. 2003).

In the simple models of Haiman & Loeb (1998) that assume a short quasar lifetime, quasars can reionize the IGM by $z \gtrsim 10$. That model was “calibrated” to reproduce the original observed relation between SMBH mass and bulge mass at $z \sim 0$ by Magorrian et al. (1998). However, this model runs into difficulties with more recent observations: (1) it overproduces the expected counts of faint X-ray sources in the *Chandra Deep Fields*, and (2) it is no longer consistent with the more recent local SMBH mass estimates (which are reduced by a factor of ~ 4) and their steeper dependence on the velocity dispersion $M_{\text{bh}} \propto \sigma^{4-5}$ rather than $\propto \sigma^3$. Wyithe & Loeb (2003b) recently presented an updated model satisfying these constraints. In their model, the abundance of faint quasars at high redshift falls short of reionizing the universe at $z \sim 6$.

Despite the above conclusions, it is natural to ask whether the abundance of fainter *miniquasars* could be higher, and whether they could then significantly contribute to the reionization history. Ricotti & Ostriker (2003) show that such quasar black holes can significantly ionize the universe if they contain a fraction $\gtrsim 10^{-5}$ of all baryons. Another example is a large population of intermediate $\sim 100 M_{\odot}$ black halos, which have a harder spectrum, and are more efficient ionizers (Madau et al. 2003). A general constraint on the quasar contribution to reionization comes from the delay between HI and HeII reionization epochs. Quasars have a hard spectrum, with the ratio of the number of photons above 4 and 1 Rydbergs about $\sim 10\%$, roughly the ratio of He vs. H atoms (the spectra of IMBHs are harder and they produce a factor of \sim two more He-ionizing photons). One may naively expect that hydrogen and helium would be reionized simultaneously, in contrast with the observed redshift $z_r(\text{H}) \gtrsim 6$ and $z_r(\text{HeII}) \sim 3$. Including the fact that HeII recombines ~ 5 times faster than HI would translate to a delay that is consistent with the values above (Miralda-Escudé & Rees 1993). However, a delay from $z_r(\text{H}) \sim 15$ to $z_r(\text{HeII}) \sim 3$ would be inconsistent with a pure miniquasar reionization scenario. This means that if hydrogen

reionization was caused by miniquasars at $z \sim 15$, then HeII was likely reionized around the same redshift; it then subsequently recombined (as stellar sources overtook miniquasars as the dominant ionizing sources), and was reionized again by the bright quasar population at $z \sim 3$. This non-trivial evolution implies a complex thermal history of the IGM, which may leave detectable imprints on its temperature distribution at lower redshifts (Hui & Haiman 2003).

The hard spectra of quasars produces several other distinguishing features for reionization (Oh 2001; Venkatesan et al. 2001). Because the mean free path is longer than the Hubble length for photons with energies $\gtrsim [(1+z)/10]^{1/2}$ keV, there is no sharp 'edge' for the discrete HII regions surrounding the ionizing sources. As a result, the neutral fraction should decrease gradually throughout most of the entire IGM. This is in sharp contrast with the swiss-cheese topology of reionization by softer photons. Furthermore, X-ray photons deposit a significant fraction ($\sim 1/3$) of their energy into ionizations only while the IGM is close to neutral. Once the ionized fraction reaches $\sim 30\%$, most of their energy is thermalized with the electrons (e.g. Shull & van Steenberg 1985). As a result, reionization by quasars would be quite different from the stellar case: the IGM would be gradually ionized to the ionized fraction of $\sim 30\%$ (as opposed to suddenly fully ionized). These features make it unlikely the quasars contributed significantly to the sudden elimination of the GP troughs at $z \sim 6$. However, the same features would be attractive in producing partial reionization at high redshift, and thus helping to explain the large optical depth measured by *WMAP* (Madau et al. 2003; Ostriker et al. 2003). Note that in this scenario, normal stars would "take over" and dominate the ionizing background at $z \sim 6$, to cause the overlap of highly ionized regions. The stars would then concurrently heat the IGM to $\sim 2 \times 10^4$ K. Hui & Haiman (2003) have argued (see also Theuns et al. 2000) that the IGM could not be kept *fully* ionized continuously from $z = 15$ to $z = 4$, because it would then cool adiabatically to a temperature that is below the observed value at $z \sim 4$. The above scenario could naturally avoid this constraint.

We have therefore seen that the first AGNs at $z > 6$ could, in principle, still be important contributors to reionization at high redshift. To conclude this section, we will point out yet another potential constraint. At energies above ~ 1 keV (rest frame at $z = 0$), there is little absorption, and whatever radiation was produced by the high- z quasar population would add cumulatively to the present-day background. Most of the soft X-ray background has already been resolved into low-redshift sources (Mushotzky et al. 2000; see also Wu & Xue 2001 and references therein), which raises the question whether these putative sources, if they are to

reionize the IGM, would necessarily overproduce the soft X-ray background (Dijkstra, Haiman & Loeb 2003).

4. Massive Black Hole Formation

Having reviewed the general problem of structure formation at high redshift, we now focus on the poorly understood question of how SMBHs were assembled in the first place. This is an outstanding problem, and it is not even clear whether the first nonlinear objects in the universe were stars or BHs, and whether galaxies or their central BHs formed first (see below). The leading ideas related to the formation of SMBHs at high redshift can be broadly divided into three areas: (1) formation of seed BHs from “normal” stellar evolution, and subsequent accretion to form SMHBs, (2) direct collapse of gas to a SMBH, usually via a supermassive star/disk (3) formation of a SMBH (or an intermediate mass black hole seed) by stellar dynamical processes in dense stellar systems such as star clusters or galactic nuclei. It is, of course, likely that all of these processes could be relevant (e.g., Begelman & Rees 1978; Rees 1984).

4.1 Seed BHs and Accretion

In view of the evidence described in §2.1, it is quite convincing that the SDSS quasars at $z \sim 6$ have masses of several $10^9 M_\odot$. Having BHs as massive as this at such an early stage in the evolution of the universe requires explanation. The simplest possibility is that they grow by gas accretion out of a stellar-mass seed BH, left behind by an early massive star. The earliest stars, forming out of metal-free gas, are thought to be massive, with $M =$ several $100 M_\odot$ (Abel, Bryan & Norman 2000, 2002; Bromm, Coppi & Larson 1999, 2002). Such stars can leave behind a substantial fraction of their original mass as a BH (Heger et al. 2003; Carr et al. 1984). As emphasized by Haiman & Loeb (2001), if the subsequent gas accretion obeys the Eddington limit, and the quasar shines with a radiative efficiency of 10%, then the time it takes for a SMBH to grow to the size of $3 \times 10^9 M_\odot$ from a stellar seed of $\sim 100 M_\odot$ is $3 \times 10^7 \ln(3 \times 10^9/100) \text{ yr} \sim 7 \times 10^8 \text{ yr}$. This is comparable to the age of the universe at $z = 6$ ($\sim 9 \times 10^8 \text{ yr}$ for a flat Λ CDM universe with $H_0 = 70 \text{ km/s/Mpc}$ and $\Omega_m = 0.3$). Therefore, the presence of these BHs is consistent with the simplest model for BH growth, *provided the seeds are present early on*, at $z \gtrsim 15$.

In this context, the crucial question is whether gas can accrete at a highly super-Eddington rate onto a BH, i.e., with $\dot{M} \gg \dot{M}_{EDD}$, where $\dot{M}_{EDD} = 10L_{Edd}/c^2 \approx 1.7M_8M_\odot \text{ yr}^{-1}$ is the accretion rate that would produce an Eddington luminosity if accretion onto a BH of mass

$10^8 M_8 M_\odot$ proceeded with 10% radiative efficiency. If so, this could lead to rapid BH growth at high redshift. Constraints on BH seeds, and their formation redshift, would therefore be much less stringent. If mass is supplied to a BH at $\dot{m} \equiv \dot{M}/\dot{M}_{EDD} \gg 1$, the photons are trapped in the inflowing gas because the photon diffusion time out of the flow becomes longer than the time it takes the gas to accrete into the BH (e.g., Begelman 1978; Begelman & Meier 1982). The resulting accretion is thus not via the usual thin disk (Shakura & Sunyaev 1973), but via a radiatively inefficient flow (RIAF); the luminosity is still set by the Eddington limit, but most of the gravitational binding energy released by the accretion process is not radiated away (being trapped in the flow).

The growth of SMBHs at high redshift probably proceeds via an optically thick photon trapped accretion flow with $\dot{m} \gg 1$. Indeed, it would be a remarkable coincidence if the mass supply rate were precisely $\sim \dot{M}_{EDD}$ (required for a thin accretion disk) during the entire growth of massive BHs. It is more likely that the mass supply rate is initially much larger in the dense environments of high redshift galaxies ($\dot{m} \gg 1$) and then slowly decreases with time as the galaxy is assembled (e.g., Small & Blandford 1992; Cavaliere et al. 2000). Recent theoretical calculations imply that if $\dot{m} \gg 1$, very little of the mass supplied to the black hole actually reaches the horizon; most of it is driven away in an outflow (see, e.g., simulations of RIAFs by Stone, Pringle, & Begelman 1999; Igumenshchev & Abramowicz 1999; Stone & Pringle 2001; Hawley & Balbus 2002; Igumenshchev et al. 2003; Proga & Begelman 2003; and the analytical models of Blandford & Begelman 1999, 2003; Quataert & Gruzinov 2000). The accretion rate onto a BH thus probably cannot exceed $\sim \dot{M}_{EDD}$ by a very large factor even if the mass supply rate is large (see Shakura & Sunyaev 1973 for an early discussion of this point).⁴

The above discussion focuses on whether highly super-Eddington accretion is possible. The question of whether the Eddington-limit can be exceeded by a modest factor of ~ 10 is a bit more subtle. Magnetized radiation dominated accretion disks are subject to a “photon bubble” instability that nonlinearly appears to lead to strong density inhomogeneities (see, in particular, Begelman 2001, 2002; and Arons 1992; Gammie 1998; Blaes & Socrates 2001). Density inhomogeneities allow super-Eddington fluxes from the accretion flow because radiation leaks out of the low density regions while most of the matter is contained in high density regions. Begelman (2002) estimates that the Eddington limit can potentially be exceeded by a factor of $\sim 10 - 100$. This would allow much more rapid growth of BHs at high redshift, circumventing the above arguments that seed BHs at $z \sim 15$ are required. Magnetohydrodynamic (MHD) simulations of radiation dominated accretion flows

are in progress that should help assess the nonlinear saturation of the photon bubble instability (Neal Turner and collaborators).

4.2 Accretion vs. Mergers

Mergers between halos can help build up the mass of individual BHs (without significantly changing the total mass of the population), provided the central BHs in the halos coalesce rapidly. The mean accretion efficiency of $\sim 10\%$ inferred from comparing the local black hole mass density with the integrated quasar light suggests that accretion dominates at least the last e-folding of the BH mass (Yu & Tremaine 2002). Mergers may, however, be significant earlier on (Haiman, Ciotti & Ostriker 2003). In addition, uncertainties in the *expected* radiative efficiency of BH accretion limit how accurately one can constrain the growth of BH mass by mergers. For example, if the typical efficiency was $\approx 40\%$ as for a maximally rotating Kerr BH, mergers could clearly dominate BH growth (on the other hand, note that multiple mergers would have a tendency to cancel the BH spin; Hughes & Blandford 2003). In order for mergers to contribute significantly to the growth of individual BH masses, stellar seeds must be present in most of the numerous minihalos that form at $z \gtrsim 15$, down to small halo masses.

For concreteness, consider possible merger histories for the $z = 5.82$ SDSS quasar (Haiman & Loeb 2001; the following arguments would be stronger for the $z = 6.41$ quasar). One can estimate the mass of the dark matter halo harboring the quasar by its abundance. SDSS searched a comoving volume of $\sim 1 \text{ Gpc}^3$ to find each quasar. Assuming a duty cycle of a few times 10^7 yrs, one estimates that the dark matter halos corresponding to this space density have masses of $10^{13} M_\odot$ (using the halo mass function in Jenkins et al. 2001; the original Press & Schechter (1974) formula would give a similar answer). A $10^{13} M_\odot$ halo at $z = 6$ has typically only ~ 10 progenitors with circular velocities of $v > 50 \text{ km/s}$ (the other progenitors being smaller). This implies that mergers can only help build up the BH mass if seed BHs are present in progenitor halos with much smaller masses. Haiman & Loeb (2001) argued for a cutoff in the BH mass function around halos with $v = 50 \text{ km/s}$ because the cosmic UV background can suppress gas infall into smaller halos (e.g., Efstathiou 1992; Thoul & Weinberg 1996; Navarro & Steinmetz 1997; Kitayama & Ikeuchi 2000). However, Dijkstra et al. (2003) have recently shown that this suppression is ineffective at redshifts as high as $z \gtrsim 6$. Thus there is no known obstacle to forming seed BHs in halos down to $v \sim 10 \text{ km/s}$ (below this threshold, atomic H cooling becomes inefficient). It therefore needs to be reassessed whether

a large fraction of the mass growth can be accounted for by mergers among halos. Clearly, placing a seed BH in each arbitrarily low-mass progenitor halo, with the same BH-halo mass ratio as inferred for the SDSS quasars ($M_{\text{bh}}/M_{\text{halo}} \sim 10^{-4}$) could account for the observed BH masses in quasars by redshift $z = 6$ even without any gas accretion (Haiman, Ciotti & Ostriker 2003).

A promising way of assessing the role of mergers in BH growth and evolution is via their gravity wave signature (see, e.g., Menou 2003 or Haehnelt 2003 for reviews). In particular, mergers occur frequently between the dark matter halos that host high-redshift black holes. If each such merger results in the coalescence of two massive black holes, the expected event rates by *LISA* are significant (see below).

The question of whether halo mergers necessarily lead to BH mergers is, however, still not resolved (see Milosavljevic & Merritt 2003 for a review). During a galaxy merger, the BHs sink via dynamical friction to the center of the galaxy and form a BH binary on scales of \sim pc. The BH binary can continue to shrink by ejecting low angular momentum stars that pass close to the binary (those in the “loss cone”). In spherical galaxies, this process is inefficient because the loss cone must be replenished by two-body relaxation. The BH binary thus appears to stall and cannot coalesce even during a Hubble time (e.g., Begelman, Blandford, & Rees 1980). Several ideas for circumventing this difficulty have been proposed. Gas accretion may drag the binary together in a manner similar to Type II migration in planetary systems (Gould & Rix 2000; Armitage & Natarajan 2002). In addition, in non-axisymmetric triaxial galaxies, low angular momentum orbits are populated much more efficiently because the stellar orbits can be chaotic; the resulting binary decay times are in many cases significantly less than a Hubble time, even if only a few percent of the stellar mass is on chaotic orbits (e.g., Yu 2002; Merritt & Poon 2003). Finally, if SMBHs are brought together by successive halo mergers at a rate higher than the rate at which they can coalesce, the lowest mass SMBH is likely to be ejected out of the nucleus of the merger remnant by the slingshot mechanism (Saslaw, Valtonen & Aarseth 1974), with implications both for gravity wave event rates, and for SMBH mass buildup.

Essentially all of the work on the gravity wave signal from BH-BH inspiral has assumed efficient (nearly instantaneous) mergers. Because *LISA* has spectacular sensitivity, it can detect such mergers at any redshift (if BHs are present). A more important constraint is on the masses of the merging BHs – the sum of the two coalescing holes needs to be $10^3 M_{\odot} \lesssim M \lesssim 10^6 M_{\odot}$ in order for the resulting gravity waves to be within *LISA*’s frequency range (e.g. Menou 2003). Several authors

(Haehnelt 1994; Menou, Haiman & Narayanan 2001; Islam, Taylor & Silk 2003b; Wyithe & Loeb 2003a) have made predictions for *LISA* event rates. If every galaxy hosts a massive BH, *LISA* should detect several hundred mergers per year, with most events at high redshifts (Menou, Haiman & Narayanan 2001; see also Wyithe & Loeb 2003a). On the other hand, Kauffmann & Haehnelt (2000) argue that only galaxies with deep potential wells ($v_c \gtrsim 100$ km/s) will form SMBHs; in this case the event rate is much less, $\sim 1 \text{ yr}^{-1}$ and is dominated by $z \lesssim 5$ (Haehnelt 2003). On a related note, Menou et al. (2001) showed that the *LISA* event rate is very sensitive to the fraction of dark matter halos that host massive BHs. This can be $\ll 1$ at high z (implying a low *LISA* event rate) because mergers ensure that every galaxy will end up with a BH at its center by $z = 0$ anyway. Note that in this case, the predicted redshift distribution of *LISA* events would be very different, peaking at low z ($z \sim 2$ in Figure 2 of Menou et al. 2001). These examples highlight the fact that gravity waves will provide a powerful probe of the formation and growth of SMBHs. We also note that predictions for the gravity wave “lightcurve” have been published to date only for equal mass black holes (Hughes 1998); since the typical merger will take place with large mass ratios, it is necessary to work out predictions for the general case.

Finally, we note that the observed morphologies of quasar hosts can, in principle, provide constraints on the prevalence of mergers. In numerous existing models, quasar activity is exclusively triggered by mergers; one then expects the images of quasar hosts to appear disturbed. Direct interpretation is difficult because galaxies may relax and display an undisturbed morphology on a timescale shorter than the lifetime of the activated quasar, especially after minor mergers (with large mass ratios). However, it is interesting to note that hosts appear clumpy at high redshift, and smoother and relaxed at lower redshifts (e.g. Kukula et al. 2001), broadly consistent with the merger rates of dark halos peaking at high redshift (Haiman & Menou 2000; Kauffmann & Haehnelt 2000).

4.3 Stars vs. Black Holes

Instead of growing by accretion/mergers from solar mass progenitors, SMBHs may form directly in the collapse of gas clouds at high redshift, via a supermassive star or disk. This depends critically on whether fragmentation of the gas cloud into stars can be avoided, particularly in view of the large angular momentum barrier that must be overcome to reach small scales in a galactic nucleus.

A number of papers have sketched how this may occur (e.g., Haehnelt & Rees 1993; Loeb & Rasio 1994). The essential idea is that when

contracting gas in a proto-galactic nucleus becomes optically thick and radiation pressure supported, it becomes less susceptible to fragmentation and star formation. It is, however, unlikely that radiation pressure becomes important before angular momentum does, implying that the gas forms a viscous accretion disk in the galactic nucleus (fragmentation before the disk forms can be avoided because the forming fragments would collide and 'coalesce' before they can separate out to discrete dense clumps; Kashlinsky & Rees 1983). On the other hand, if self-gravitating, the resulting disk is strongly gravitationally unstable and becomes prone to fragmentation and star formation (e.g. Shlosman & Begelman 1989; Goodman 2003). Whether this fragmentation can be avoided is unclear. One possibility is to stabilize the disk by keeping its temperature high, which may be possible in a virtually metal-free high redshift halo. In particular, H_2 molecules are fragile, and can be easily dissociated by an early soft UV background (Haiman, Rees & Loeb 1997). If molecular hydrogen cooling can be suppressed, the gas will lack coolants and collapse isothermally at a temperature of ~ 8000 K set by atomic line cooling. If molecules are prevented from forming, it may then be unable to fragment into stars, and form a $\sim 10^6 M_\odot$ SMBH instead (this seems difficult to arrange, but options to achieve this are discussed in Oh & Haiman 2003). Bromm & Loeb (2003) have carried out numerical simulations of this scenario and indeed find that if the temperature is kept at 10^4K , $\gtrsim 10^6 M_\odot$ can condense to scales of $\lesssim 1$ pc. At the end of their simulations the gas is still inflowing with no indication of fragmentation.

Another possibility is that, even in the presence of significant cooling, angular momentum transport by gravitational instabilities, spiral waves, bars, etc., can drive a fraction of the gas to yet smaller scales in the galactic nucleus. Eisenstein & Loeb (1995) argued that this was particularly likely to occur in rare low angular momentum dark matter halos because the disk could viscously evolve before star formation commenced. In addition, even if most of the gas is initially converted into stars, stellar winds and supernovae will eject a significant amount of this gas back into the nucleus, some of which will eventually collapse to smaller scales (Begelman & Rees 1978).

Although the detailed evolutionary pathways are still not understood, a possible outcome of the above scenarios is the continued collapse of some gas to smaller scales in the galactic nucleus. As the gas flows in it becomes optically thick and the photon diffusion time eventually exceeds the inflow time. Radiation pressure dominates for sufficiently massive objects so that the adiabatic index is $\Gamma \approx 4/3$. Radiation pressure may temporarily balance gravity, forming a supermassive

star or disk (SMS; e.g., Hoyle & Fowler 1963; Wagoner 1969; see, e.g., Shapiro & Teukolsky 1983 for a review and additional references to earlier work). The SMS will radiate at the Eddington limit (but see § 1.4.1 above) and continue contracting. When the SMS is sufficiently compact ($GM/Rc^2 \approx 10^{-4}M_8^{-1/2}$ for nonrotating stars), general relativistic corrections to the gravitational potential become important and the star becomes dynamically unstable because its effective polytropic index is $\lesssim 4/3$. For masses $\lesssim 10^5 M_\odot$, thermonuclear reactions halt the collapse and generate an explosion (e.g., Fuller Woosley, & Weaver 1986), but more massive objects appear to collapse directly to a SMBH (see Shapiro 2003 for a review; and, e.g., Shibata & Shapiro 2002; Saijo et al. 2002 for recent simulations).

4.4 The Formation of BHs in Stellar Clusters

The negative heat capacity of self-gravitating stellar systems makes them vulnerable to gravitational collapse in which the core of the cluster collapses on a timescale t_{cc} comparable to the two-body relaxation time of the cluster (Binney & Tremaine 1987). If core-collapse proceeds unimpeded, the resulting high stellar densities can lead naturally to the runaway collisional growth of a single massive object which may evolve to form a BH (as in the discussion of SMSs above). This process provides an additional route for the direct formation of SMBHs at high redshift (or, more likely, intermediate mass seeds).

Early work suggested that the fate of stellar clusters depends sensitively on the number of stars in the cluster. Lee (1987) and Quinlan & Shapiro (1990) found that very dense massive stars clusters ($N \gtrsim 10^6 - 10^7$ stars) were required to have successful core-collapse and runaway growth of a single massive object. In less massive clusters, core collapse was halted by binary heating, in which the cluster gains energy at the expense of binaries via three-body interactions (Heggie 1975; Hut et al. 1992). Successful core-collapse also requires that t_{cc} is shorter than the timescale for the most massive stars to evolve off the main sequence (Rasio et al. 2003; this requirement implies compact clusters \lesssim pc in size). Otherwise mass loss from evolved stars and supernovae prevents the core from collapsing (in much the same way as binary stars systems can become unbound by supernovae).

Recent work has revived earlier ideas that stellar clusters are subject to a “mass-segregation instability” which makes even less massive clusters prone to forming BHs (Spitzer 1969; Vishniac 1978; Begelman & Rees 1978). Because massive stars in a cluster sink by dynamical friction towards the center (mass segregation), they invariably dominate

the dynamics of the cluster core and can undergo core-collapse on a timescale much shorter than that of the cluster as a whole (and on a timescale shorter than their main sequence lifetime). Portegies-Zwart & McMillan (2002) showed with N-body simulations that the resulting core-collapse likely leads to runaway merger and formation of a single BH. Gürkan et al. (2003) reached a similar conclusion for much larger $N \sim 10^7$ using Monte-Carlo simulations.

The above processes provide a promising channel for the formation of intermediate mass black hole (IMBH) seeds which can grow via mergers and/or accretion to form SMHBs. For example, Volonteri et al. (2003) and Islam et al. (2003a) have recently incorporated such early BH seeds into Monte Carlo simulations of the black hole merger histories. With reasonable prescriptions for the merging and accretion of BHs inside dark halos, these models can account for the observed evolution of the quasar luminosity functions at $z < 5$, and can serve for physically motivated extrapolations to high redshifts to describe the first AGNs.

It should be noted that IMBHs may have been directly detected using stellar dynamics in the globular clusters G1 (Gebhardt et al. 2002) and M15 (van der Marel et al. 2002; although this object can be modeled without an IMBH; van der Marel 2003) and or as ultra-luminous X-ray sources in nearby galaxies (e.g., Colbert & Mushotzky 1999; Kaaret et al. 2001). There are, however, viable non-IMBH interpretations of both the globular cluster (e.g., Baumgardt et al. 2003) and ULX observations (e.g., King et al. 2001; Begelman 2002).

5. The Future

In this section, we briefly summarize the possibility of probing the continuum and line emission from AGN beyond the current redshift horizon of $z \sim 6$. This discussion is necessarily based on models for how the AGN population evolves at $z > 6$. Such models can be constructed by assuming that SMBHs populate dark matter halos, e.g., in accordance with the locally-measured $M_{\text{bh}} - \sigma$ relation (or an extrapolation of the relation to higher redshifts). While there is no direct measurement of this relation at high redshift, this assumption is at least plausible. There is, e.g., tentative evidence that the relation holds for $z \sim 3$ quasars (this is based on using the $\text{H}\beta/\text{OIII}$ lines as proxies for BH mass and σ , respectively; e.g. Shields et al. 2003), and also at $z \sim 6$ (based on the argument in section 2.1 above). No doubt the observational constraints will improve as both BH masses and velocity dispersions are measured in larger samples of distant quasars (e.g. from the SDSS). Correspondingly, extrapolations to high- z will be more reliable as the feedback processes

that regulate BH growth are better understood. Here we summarize predictions from the simplest models.

5.1 Broad Band Detections

Predictions for the number counts of high z AGN have been made using simple semi-analytic models for the near-IR (Haiman & Loeb 1998b) and in soft X-rays (Haiman & Loeb 1999b). In these early models, the quasar BH was assumed to have a fixed fraction $\sim 10^{-4}$ of the halo mass, shine at the Eddington luminosity, and have a duty cycle of bright activity of $\sim 10^6$ years.

In such models, the surface density of sources is very high in optical/near-IR bands, even at $z \sim 10$. For example, in the $1-5\mu\text{m}$ band, the $\sim 1\text{nJy}$ sensitivity of the *James Webb Space Telescope* allows the detection of a $\sim 10^5 M_\odot$ black hole at $z = 10$ (provided that the BH shines at the Eddington limit, with the Elvis et al. 1994 spectrum). Several sources per sq. amin are predicted at this threshold from $z \gtrsim 5$, with most of these sources at $z \gtrsim 10$ (Haiman & Loeb 1999a). It would also be interesting to detect the host galaxies of ultra-high redshift AGNs, which should be feasible with *JWST*'s sensitivity. If the galaxies occupy a fair fraction of the virial radius of their host halos, many of them can potentially be resolved with *JWST*'s planned angular resolution of $\sim 0.05''$. The Large Synoptic Survey Telescope (LSST; Tyson 2002⁵), with a planned capability of going ~ 5 magnitudes deeper than SDSS in a ~ 3 times larger solid angle, would be an ideal instrument for studying high- z quasars in the optical/near-IR, provided it is equipped with a sufficiently red filter.

In the soft X-rays, the $1-2\text{ keV}$ flux of $10^{-16}\text{ erg s}^{-1}\text{ cm}^{-2}$ reached by the *CDFs* corresponds to a much larger $\sim 10^8 M_\odot$ black hole at $z = 10$, but nevertheless, thousands of sources are predicted at $z \gtrsim 5$ per sq. degree, and tens per sq. degree at $z > 10$. This would imply that tens of $z > 5$ sources should have been detectable already in the *CDFs*, whereas only a handful of potential candidates, and no confirmed sources, have been found (as discussed in § 1.2.2 above). In revised models, with longer quasar lifetimes and thus a steeper scaling of M_{bh} with M_{halo} , these numbers can be sharply decreased (Haiman & Loeb 1998b; Haehnelt, Natarajan & Rees 1998), which can bring the expected counts into agreement with current non-detections (Wyithe & Loeb 2003b).

Large numbers of dusty $z > 6$ AGN could be detected at mid-IR wavelengths. Although we are not aware of predictions at these wavelengths for AGNs, strong dust-enshrouded starbursts that turn most of the gas into stars would result in large source counts at longer wavelengths. Hundreds of galaxies per sq. amin could be detectable at the

$\sim 100\text{nJy}$ threshold (in the semi-analytical models of Haiman, Spergel & Turner 2003). This flux level can be reached in a $\sim 10^6$ s exposure with the *Space Infrared Telescope Facility (SIRTF)*, and is well within the sensitivity of the *JWST* and of the proposed *Terrestrial Planet Finder (TPF)*.

The radio sensitivity of the extended VLA and other forthcoming instruments (e.g., ATA, SKA) is also promising for probing AGN beyond $z \sim 6$. Using the updated scaling of BH mass with halo mass and redshift from Wyithe & Loeb (2003b), and assuming the same radio loud fraction ($\sim 10\%$) as at lower redshifts, we find that a few $10\mu\text{Jy}$ sources per sq. amin should be detectable at 1–10 GHz (Haiman & Quataert 2004).

In addition to direct detection of AGN at very high z it may also be possible to detect lower mass seed BHs at comparable redshifts (or higher). In particular, a plausible model for gamma-ray bursts (GRBs) invokes accretion onto a newly formed $\sim 10M_{\odot}$ BH (the collapsar model; e.g., Woosley 1993). GRB afterglow emission may be directly detectable from $z \sim 10 - 20$ (e.g., Lamb & Reichart 2000; Ciardi & Loeb 2000). Such afterglows would show up as, e.g., fading I-band dropouts in IR surveys (which are under development; Josh Bloom, private communication). Their detection would open up a new probe of BH formation and evolution at high redshift (as well as a new probe of the IGM along the line of sight; e.g. Barkana & Loeb 2003b).

In summary: model predictions for the continuum emission of $z > 6$ AGN are very sensitive to how one extrapolates the $M_{\text{bh}} - M_{\text{halo}}$ relation to $z \gtrsim 6$. However, this should be viewed as “good news”: (1) large numbers of detectable AGNs at these redshifts are certainly possible, and (2) their detection will put strong constraints on models for the origin and evolution of the BH population.

5.2 Emission Line Measurements

The strongest recombinations lines of H and He from $5 < z < 20$ AGN will fall in the near-IR bands of *JWST*, and could be bright enough to be detectable. Specific predictions have been made for the source counts in the $\text{H}\alpha$ emission line (Oh 2001), and for the three strongest HeII lines (Oh, Haiman & Rees 2001; Tumlinson, Giroux & Shull 2001). The key assumption is that most of the ionizing radiation produced by the miniquasars is processed into such recombination lines (rather than escaping into the IGM). Under this assumption, the lines are detectable for a fiducial $10^5 M_{\odot}$ miniquasar at $z = 10$. The Lyman α line is more susceptible to absorption by neutral hydrogen in the IGM near the source, but should be detectable for bright sources that are surrounded by a

large enough HII region so that Ly α photons shift out of resonance before hitting the neutral IGM (Cen & Haiman 2000).

The simultaneous detection of H and He lines would be especially significant. As already argued above, the hardness of the ionizing continuum from the first sources of UV radiation plays a crucial role in the reionization of the intergalactic medium (IGM). It would therefore be very interesting to directly measure the ionizing continuum of any $z > 6$ source. While this may be feasible at X-ray energies for exceptionally bright sources, the absorption by neutral gas within the source and in the intervening IGM will render the ionizing continuum of high-redshift sources inaccessible to direct observation in IR bands. This is a problem if the ionizing sources are black holes with $M < 10^8 M_\odot$ at $z \sim 10$ (easily detectable at wavelengths redward of redshifted Lyman α in the near IR by *JWST*, but too faint to see in X-rays). The comparison of H α and HeII line strengths can be used to infer the ratio of HeII to HI ionizing photons, $Q = \dot{N}_{\text{ion}}^{\text{HeII}} / \dot{N}_{\text{ion}}^{\text{HI}}$. A measurement of this ratio would shed light on the nature of the first luminous sources, and, in particular, it could reveal if the source has a soft (stellar), or hard (AGN-like) spectrum. Note that this technique has already been successfully applied to constrain the spectra of sources in several nearby extragalactic HII regions (Garnett et al. 1991).

Provided the gas in the high- z AGNs is enriched to solar levels, several molecular lines may be visible. In fact, CO has already been detected in the most distant $z = 6.41$ quasar (Walter et al. 2003). The detectability of CO for high- z sources in general has been considered by Silk & Spaans (1997) and by Gnedin, Silk & Spaans (2001). If AGN activity is accompanied by a star formation rate of $\gtrsim 30 M_\odot/\text{yr}$, the CO lines are detectable at all redshifts $z = 5 - 30$ by the Millimeter Array (the redshift-independent sensitivity is due to the increasing CMB temperature with redshift); while ALMA could reveal fainter CO emission. The detection of these molecular lines will provide valuable information on the stellar content and gas kinematics near the AGN.

6. Conclusions

In this review, we have summarized theoretical ideas and observational constraints on how massive black holes form at the centers of galaxies, and how such BHs grow via accretion and mergers to give rise to the observed population of BHs in the local and moderate-redshift universe. This remains a poorly understood but important problem. In addition to being of intrinsic interest for understanding the AGN phenomena, sources of gravity waves, etc., there is strong evidence that the

formation and evolution of black holes is coupled to the formation and evolution of the host galaxy in which the black hole resides (e.g., the $M_{\text{bh}} - \sigma$ relation), and thus to the cosmological formation of non-linear dark matter structures (i.e., the dark halos surrounding these galaxies). We anticipate that this will remain a growth area of research in the coming years, with continued rapid progress on both the observational and theoretical fronts.

Acknowledgments

The authors wish to thank the editor of this volume, Amy Barger, for her patience during the preparation of this article.

Notes

1. A possibly important exception is tentative evidence for increasing Eddington ratios towards higher redshift, discussed in § 1.2.1
2. The above statement refers to a transition from neutral to ionized hydrogen. A transition from HeII to HeIII appears to be occurring at a lower redshift, $z \sim 3$ (e.g. Heap et al. 2000).
3. White et al. (2003) detect flux blueward of Lyman α in the spectra of the $z = 6.41$ quasar, but they attribute this flux to an intervening $z = 4.9$ galaxy. Furthermore, even if the feeble flux originated from the quasar, the strong lower limits on the mean neutral fraction would remain.
4. Note that this is fully consistent with the mean accretion efficiency of $\sim 10\%$ inferred from comparing the local black hole mass density with the integrated quasar light, even though accretion is not via a thin disk.
5. www.lsst.org

References

- Abel, T., Anninos, P., Zhang, Y., Norman, M. L. 1997, *NewA*, 2, 181
 Abel, T., Bryan, G. L., & Norman, M. L. 2000, *ApJ*, 540, 39
 Abel, T., Bryan, G. L., & Norman, M. L. 2002, *Science*, 295, 93
 Abel, T., & Haiman, Z. 2001, in “Molecular hydrogen in space”, Cambridge contemporary astrophysics, Eds. F. Combes, and G. Pineau des Forrets, Cambridge University Press: Cambridge, UK, p. 237
 Alexander, D. M., et al. 2001, *AJ*, 122, 2156
 Armitage, P. J. & Natarajan, P. 2002, *ApJ*, 267, L9
 Arons, J., 1992, *ApJ*, 388, 561
 Barger, A., J., et al. 2003, *ApJ*, 584, 61
 Barkana, R., & Loeb, A. 2000, *ApJ*, 531, 613
 Barkana, R., & Loeb, A. 2003a, *Nature*, 421, 341
 Barkana, R., & Loeb, A. 2003b, *ApJ*, in press, astro-ph/0305470
 Baumgardt. H., Makino, J., Hut, P., McMillan, S., & Portegies-Zwart, S. 2003, *ApJ* submitted (astro-ph/0301469)
 Becker, R. H., et al. 2001, *AJ*, 122, 2850

- Begelman, M. C. 1978, MNRAS, 184, 53
Begelman, M. C. 2001, ApJ, 551, 897
Begelman, M. C. 2002, ApJ, 568, L97
Begelman, M. C., Blandford, R. D. & Rees, M. J. 1980, Nature, 287, 307
Begelman, M. C. & Meier, D. L. 1982, ApJ, 253, 87
Begelman, M. C. & Rees, M. J. 1978, MNRAS, 185, 847
Bennett, C. L., et al. 2003, ApJS, 148, 1
Binney, J. & Tremaine, S. 1987, Galactic Dynamics (Princeton: Princeton Univ. Press)
Blaes, O. & Socrates, A. 2001, ApJ, 553, 987
Blandford, R.D. 1999, in Galaxy Dynamics, ASP Conf. Ser. vol. 182, Eds.D.R. Merritt, M. Valluri and J.A. Sellwood, (San Francisco: ASP), p.87
Blandford, R. D., & Begelman, M. C. 1999, MNRAS, 303, L1
Blandford, R. D., & Begelman, M. C. 2003, MNRAS, in press (astro-ph/0306184)
Bromm, V., Coppi, P. S., & Larson, R. B. 1999, ApJ, 527, 5
Bromm, V., Coppi, P. S., & Larson, R. B. 2002, ApJ, 564, 23
Bromm, V., & Loeb, A. 2003, ApJ, 596, 34
Carr, B. J., Bond, J. R., Arnett, W. D. 1984, ApJ, 277, 445
Cattaneo, A., & Bernardi, M. 2003, MNRAS, in press, astro-ph/0305298
Cavaliere et al. 2000
Cen, R., & Haiman, Z. 2000, ApJL, 542, 75
Cen, R., & McDonald, P. 2002, ApJ, 570, 457
Ciardi, B., Ferrara, A., & White, S. D. M. 2003, MNRAS, 344, 7
Ciardi, B. & Loeb, A. 2000, ApJ, 540, 687
Ciotti, L., & Ostriker, J. P. 2001, ApJ, 551, 131
Ciotti, L., & van Albada, T. S. 2001, ApJ, 552, L13
Colbert, E. J. M. & Mushotzky, R. F. 1999, ApJ, 519, 89
Comerford, J., Haiman, Z., & Schaye, J. 2002, ApJ, 580, 63
Croom, S., et al. 2003, in "AGN Physics with the Sloan Digital Sky Survey", Eds. G.T. Richards and P.B. Hall, in press, astro-ph/0310533
Di Matteo, T., Croft, R. A. C., Springel, V., & Hernquist, L. 2003, ApJ, 593, 56
Dijkstra, M., Haiman, Z., & Loeb, A. 2003, in preparation
Dijkstra, M., Haiman, Z., Rees, M. J., & Weinberg, D. H. 2003, ApJ, in press, astro-ph/0308042
Eisenstein, D. & Loeb, A. 1995, ApJ, 443, 11
Efstathiou, G. 1992, MNRAS, 256, 43p
Efstathiou, G., & Rees, M. J. 1988, MNRAS, 230, 5p

- Elvis, M., Wilkes, B. J., McDowell, J. C., Green, R. F., Bechtold, J., Willner, S. P., Oey, M. S., Polomski, E., & Cutri, R. 1994, *ApJS*, 95, 1
- Fabian, A. C. 2003, to appear in *Carnegie Observatories Astrophysics Series, Vol. 1: Coevolution of Black Holes and Galaxies*, ed. L. C. Ho (Cambridge: Cambridge Univ. Press)
- Fan, X., Narayanan, V. K., Strauss, M. A., White, R. L., Becker, R. H., Pentericci, L., and Rix, H. 2002, *AJ*, 123, 1247–1257.
- Fan, X., et al. 2000, *AJ*, 120, 1167
- Fan, X., et al. 2001, *AJ*, 122, 2833
- Fan, X., et al. 2002, *AJ*, 123, 1247
- Fan, X., et al. 2003, *AJ*, 125, 1649
- Ferrarese, L., Merritt, D. 2000, *ApJ*, 539, L9
- Ferrarese, L. 2002, *ApJ*, 578, 90
- Ferrarese, L. 2002b, in *Current High-Energy Emission around Black Holes*, ed. C-H. Lee & H.-Y. Chang (Singapore: World Scientific), 3
- Fukugita, M., Hogan, C. J., & Peebles, P. J. E. 1998, *ApJ*, 503, 518
- Fuller, G. M., Woosley, S. E., & Weaver, T. A. 1986, *ApJ*, 307
- Galli, D., & Palla, F. 1998, *A&A*, 335, 403
- Gammie, C. F., 1998, *MNRAS*, 297, 929
- Gammie, C. F., 1999, *ApJ*, 522, L57
- Garnett, D. R., Kennicutt, R. C., Chu, Y.-H., & Skillman, E. D. 1991, *ApJ*, 373, 458
- Gebhardt, K., et al. 2000, *ApJ*, 539, L13
- Gebhardt, K., Rich, R. M., & Ho, L. 2002, *ApJ*, 578, L41
- Gnedin, N.Y. 2000, *ApJ*, 535, 530
- Gnedin, N. Y., & Ostriker, J. P. 1996, *ApJ*, 472, 63
- Gnedin, N. Y., & Ostriker, J. P. 1997, *ApJ*, 486, 581
- Gnedin, N. Y., Silk, J., & Spaans, M. 2001, *MNRAS*,
- Goodman, J., 2003, *MNRAS*, 339, 937
- Gould, A. & Rix, H.-W. 2000, *ApJ*, 532, L29
- Graham, A.W., Erwin, P., Caon, N., & Trujillo, I. 2001, *ApJ*, 563, L13
- Granato, G. L., Silva, L., Monaco, P., Panuzzo, P., Salucci, P., De Zotti, G., & Danese, L. 2001, *MNRAS*, 324, 757
- Gunn, J. E., & Peterson, B. A. 1965, *ApJ*, 142, 1633
- Gürkan, M. A., Freitag, M., & Rasio, F. A. 2004, *ApJ* submitted (astro-ph/0308449)
- Haardt, F. & Madau, P. 1996, *ApJ*, 461, 20
- Haehnelt, M. G. 1994, *MNRAS*, 269, 199
- Haehnelt, M. G. 2003, *Classical and Quantum Gravity*, 20, S31 (astro-ph/0307379)

- Haehnelt, M. G., & Rees, M.J. 1993, MNRAS
- Haehnelt, M. G., Natarajan, P., Rees, M.J. 1998, MNRAS, 300, 827
- Haiman, Z. 2003, to appear in Carnegie Observatories Astrophysics Series, Vol. 1: Coevolution of Black Holes and Galaxies, ed. L. C. Ho (Cambridge: Cambridge Univ. Press), astro-ph/0304131
- Haiman, Z., Abel, T., & Madau, P. 2001, ApJ, 551, 599
- Haiman, Z., & Cen, R. 2002, ApJ, 578, 702
- Haiman, Z., Ciotti, L. & Ostriker, J. P. 2003, ApJ, submitted, astro-ph/0304129
- Haiman, Z., & Holder, G. P. 2003, ApJ, 595, 1
- Haiman, Z., & Hui, L. 2001, ApJ, 547, 27
- Haiman, Z., & Loeb, A. 1998a, in “Science with the NGST”, Eds. E.P. Smith and A. Koratkar, ASP Conf. Ser. 133, p. 251
- Haiman, Z., & Loeb, A. 1998b, ApJ, 503, 505
- Haiman, Z., & Loeb, A. 1999a, in “After the Dark Ages: When Galaxies were Young”, Eds. S. Holt and E. Smith, AIP Press, p. 34
- Haiman, Z., & Loeb, A. 1999b, ApJ, 519, 479
- Haiman, Z., & Loeb, A. 2001, ApJ, 552, 459
- Haiman, Z., Madau, P., & Loeb, A. 1999, ApJ, 514, 535
- Haiman, Z., & Menou, K. 2000, ApJ, 531, 42
- Haiman, Z., Thoul, A. A., & Loeb, A. 1996, ApJ, 464, 523
- Haiman, Z., Rees, M. J., & Loeb, A. 1997, ApJ, 476, 458 [erratum: 1997, ApJ, 484, 985]
- Haiman, Z., Spergel, D. N., & Turner, E. 2003, ApJ, 585, 630
- Haiman, Z., & Quataert, E. 2004, in preparation
- Hasinger, G., in “New Visions of the X-ray Universe in the XMM-Newton and Chandra Era”, ed. F. Jansen, (ESA: Noordwijk), in press, astro-ph/0202430
- Heap, S. R., Williger, G. M., Smette, A., Hubeny, I., Sahu, M., Jenkins, E. B., Tripp, T. M., & Winkler, J. N. 2000, ApJ, 534, 69
- Heger, A., et al. 2003, ApJ, 591, 288
- Heggie, D. C. 1975, MNRAS, 173, 729
- Hirasawa, T. 1969 Prog. Theor. Phys. 42(3), 523
- Holder, G. P., Haiman, Z., Kaplinghat, M., & Knox, L. 2003, ApJ, 595, 13
- Hoyle, F. & Fowler, W. A. 1963, MNRAS, 125, 169
- Hughes, S. A., & Blandford, R. D. 2003, ApJ, 585, 101
- Hui, L., & Haiman, Z. 2003, ApJ, 596, 9
- Hut, P. et al. 1992, PASP, 104, 981
- Igumenshchev, I., Narayan, R., & Abramowicz, M. 2003, ApJ, 592, 1042
- Igumenshchev, I. & Abramowicz, M. 1999, MNRAS, 303

- Islam, R. R., Taylor, J. E., & Silk, J. 2003a, MNRAS, submitted, astro-ph/0307171
- Islam, R. R., Taylor, J. E., & Silk, J. 2003b, MNRAS, submitted, astro-ph/0309559
- Jenkins, A. et al. 2001, MNRAS, 321, 372
- Kaaret, P. et al. 2001, MNRAS, 321, L29
- Kang, H., & Shapiro, P. R. 1992, ApJ, 386, 432
- Kang, H., Shapiro, P. R., Fall, S. M., & Rees, M. J. 1990, ApJ, 363, 488
- Kaplinghat, M., Chu, M., Haiman, Z., Holder, G., Knox, L., & Skordis, C. 2003, ApJ, 583, 24
- Kashlinsky, A., & Rees, M. J. 1983, MNRAS, 205, 955
- Kaspi, S., Smith, P. S., Netzer, H., Maoz, D., Jannuzi, B. T., & Giveon, U. 2000, ApJ, 533, 631
- Kauffmann, G. & Haehnelt, M. 2000, MNRAS, 311, 576
- King, A. R. et al. 2001, ApJ, 552, L109
- King, A. R. 2003, ApJ, 596, L27
- Kitayama, T., & Ikeuchi, S. 2000, ApJ, 529, 615
- Kochanek, C. S. 1998, in Science with the NGST, ed. E. P. Smith & A. Koratkar, ASP Conference Series vol. 133, p. 96
- Kogut, A., et al. 2003, ApJS, 148, 161
- Kuhlen, M., Keeton, C. R., & Haiman, Z. 2004, in preparation
- Kukula, M. J., et al. 2001, MNRAS, 326, 1533
- Lamb, D. Q. & Reichart, D. E. 2000, ApJ, 536, L1
- Lee, H. M. 1987, ApJ, 319, 801
- Lepp, S., & Shull, J. M. 1984, ApJ, 280, 465
- Lidz, A., Hui, L., Zaldarriaga, M., & Scoccimarro, R. 2002, ApJ, 579, 491
- Loeb, A., & Rasio, F. 1994, 432, 52
- Lynden-Bell, D. 1967, MNRAS, 136, 101
- Lynden-Bell, D. 1969, Nature, 223, 690
- Madau, P., Haardt, F., & Rees, M. J. 1999, ApJ, 514, 648
- Madau, P., Ferrara, A., & Rees, M. J. 2001, ApJ, 555, 92
- Madau, P., & Pozzetti, L. 2000, MNRAS, 312, 9
- Madau, P., & Rees, M. J. 2001, ApJ, 551, L27
- Madau, P., Rees, M. J., Volonteri, M., Haardt, F., & Oh, S.-P. 2003, ApJ, submitted, astro-ph/0310223
- Magorrian, J., et al. 1998, AJ, 115, 2285
- Martini, P. 2003, to appear in Carnegie Observatories Astrophysics Series, Vol. 1: Coevolution of Black Holes and Galaxies, ed. L. C. Ho (Cambridge: Cambridge Univ. Press), astro-ph/0304009
- Martini, P., & Weinberg, D. H. 2001, ApJ, 547, 12

- Matsuda, T., Sato, H., & Takeda, H. 1969, *Prog. Theor. Phys.*, 42(2), 219
- McDonald, P., & Miralda-Escudé, J. 2001, *ApJ*, 549, 11L
- Merritt, D., & Ferrarese, L. 2001, *MNRAS*, 320, L30
- Merritt, D., & Poon, M. Y. 2003, *ApJ*, submitted, astro-ph/0302296
- Menou, K. 2003, *Classical and Quantum Gravity*, 20, S37 (astro-ph/0301397)
- Menou, K., Haiman, Z., & Narayanan, V. K. 2001, *ApJ*, 558, 535
- Milosavljevic, M. & Merritt, D. 2003, in *The Astrophysics of Gravitational Wave Sources*, ed. J. Centrella, AIP, in press (astro-ph/0211270)
- Miralda-Escudé, J., & Kollmeier, J. A. 2003, *ApJ*, submitted, astro-ph/0310717
- Miralda-Escudé, J., & Rees, M. J. 1993, *MNRAS*, 260, 617
- Miyaji, T., Hasinger, G., & Schmidt, M. 2000, *A&A*, 353, 25
- Monaco, P., Salucci, P., & Danese, L. 2000, *MNRAS*, 311, 279
- Mushotzky, R. F., Cowie, L. L., Barger, A. J., & Arnaud, K. A. 2000, *Nature* 404, 459
- Navarro, J. F., & Steinmetz, M. 1997, *ApJ*, 478, 13
- Netzer, H. 2003, *ApJ*, 583, L5
- Oh, S. P. 2001, *ApJ*, 553, 499
- Oh, S. P., & Haiman, Z. 2002, *ApJ*, 569, 558
- Oh, S. P., & Haiman, Z. 2003, *MNRAS*, in press, astro-ph/0307135
- Oh, S. P., Haiman, Z., & Rees, M. J. 2001, *ApJ*, 553, 73
- Palla, F., Salpeter, E. E., & Stahler, S. W. 1983, *ApJ*, 271, 632
- Portegies-Zwart, S. F., McMillan, S. L. W. 2002, *ApJ*, 576, 899
- Press, W. H., & Schechter, P. L. 1974, *ApJ*, 181, 425
- Proga, D. & Begelman, M. C. 2003, *ApJ*, 592, 767
- Quataert, E. & Gruzinov 2000, *ApJ*, 545, 842
- Quinlan, G. D. & Shapiro, S. L. 1990, *ApJ*, 356, 483
- Rasio, F. A., Freitag, M., & Gürkan, M. A. 2003, to appear in *Carnegie Observatories Astrophysics Series, Vol. 1: Coevolution of Black Holes and Galaxies*, ed. L. C. Ho (Cambridge: Cambridge Univ. Press), astro-ph/0304038
- Rees, M. J. 1984, *ARA&A*, 22, 471
- Rees, M. J., & Ostriker, J. P. 1977, *ApJ*, 179, 541
- Richards, G. T., Strauss, M. A., Pindor, B., Haiman, Z., Eisenstein, D., Fan, X. & Schneider, D. P. 2003, *AJ*, in press, astro-ph/0309274
- Ricotti, M., & Ostriker, J. P. 2003, *MNRAS*, submitted, astro-ph/0311003
- Saijo, M., Shibata, M., Baumgarte, T. W., & Shapiro, S. L. 2002, *ApJ*, 569, 349
- Salpeter, E. E. 1964, *ApJ*, 140, 796
- Salucci, P., et al. 1999, *MNRAS*, 307, 637
- Saslaw, W. C., Valtonen, M. J. & Aarseth, S. J. 1974, *ApJ*, 190, 253

- Shakura, N. I., & Sunyaev, R. A. 1973, *A&A*, 24, 337
- Shapiro, P. R., Giroux, M. L., & Babul, A. 1994, *ApJ*, 427, 25
- Shapiro, P. R., & Kang, H. 1987, *ApJ*, 318, 32
- Shapiro, S. L. 2003, to appear in *Carnegie Observatories Astrophysics Series, Vol. 1: Coevolution of Black Holes and Galaxies*, ed. L. C. Ho (Cambridge: Cambridge Univ. Press), astro-ph/0304202
- Shapiro, S. L. & Teukolsky, S. A. 1983, *Black Holes, White Dwarfs, and Neutron Stars* (New York: Wiley)
- Shaver, P. A., et al. 1996, *Nature*, 384, 439
- Shibata, M. & Shapiro, S. L. 2002, *ApJ*, 527, L39
- Shields, G. A., Gebhardt, K., Salviander, S., Wills, B. J., Xie, B., Brotherton, M. S., Yuan, J., & Dietrich, M. 2003, *ApJ*, 583, 124
- Small, T. A., & Blandford, R. D. 1992, *MNRAS*, 259, 725
- Sheth, R. K., Mo, H. J., & Tormen, G. 2001, *MNRAS*, 323, 1
- Shlosman, I., & Begelman, M. C. 1989, *ApJ*, 341, 685
- Shull, J. M., & van Steenberg, M. E. 1985, *ApJ*, 298, 268
- Silk, J., & Rees, M. J., *A&A*, 331, L1
- Silk, J., & Spaans, M. 1997, *ApJ*, 488, L79
- Sołtan, A. 1982, *MNRAS*, 200, 115
- Songaila, A., & Cowie, L. L. 2002, *AJ*, 123, 2183
- Spergel, D. N. et al. 2003, *ApJS*, 148, 175
- Spitzer, L, L, Jr. 1969, *ApJ*, 158, L139
- Stone, J. M. & Pringle, J. E. 2001, *MNRAS*, 322, 461
- Stone, J., Pringle, J., & Begelman, M. 1999, *MNRAS*, 310, 1002
- Susa, H., Uehara, H., Nishi, R., & Yamada, M. 1998, *Prog. Theor. Phys.* 100, 63
- Tegmark, M., Silk, J., Rees, M. J., Blanchard, A., Abel, T., & Palla, F. 1997, *ApJ*, 474, 1
- Thoul, A. A., & Weinberg, D. H. 1996, *ApJ*, 465, 608
- Tumlinson, J., Giroux, M. L., & Shull, M. J. 2001, *ApJ*, 550, L1
- Tyson, J. A. 2002, *Proc. SPIE Int. Soc. Opt. Eng.* 4836, 10 (astro-ph/0302102)
- Yu, Q. 2002, *MNRAS*, 331, 935
- Yu, Q. & Tremaine, S. 2002, *MNRAS*, 335, 965
- Valageas, P., & Silk, J. 1999, *A & A*, 347, 1
- Van der Marel, R. P. 1997, in “Galaxy Interactions at Low and High Redshift”, *Proc. IAU Symp.* 186., Eds. J.E. Barnes and D.B. Sanders, Kluwer, Dordrecht, p. 102
- Van der Marel, R. P. 2003, to appear in *Carnegie Observatories Astrophysics Series, Vol. 1: Coevolution of Black Holes and Galaxies*, ed. L. C. Ho (Cambridge: Cambridge Univ. Press), astro-ph/0302101
- van der Marel, R. P. et al. 2002, *AJ*, 124, 3255

- Vanden Berk, D. E., et al. 2001, *AJ*, 122, 549
Venkatesan, A., Giroux, M. L., & Shull, J. M. 2001, *ApJ*, 563, 1
Vestergaard, M. 2002, *ApJ*, 571, 733
Vestergaard, M. 2004, *ApJ*, in press, astro-ph/0309521
Vishniac, E. T. 1978, *ApJ*, 223, 986
Volonteri, M., Haardt, F., & Madau, P. 2003, *ApJ*, 582, 559
Wagoner, R. V. 1969, *ARA&A*, 7, 553
Walter, F., Bertoldi, F., Carilli, C., et al. 2003, *Nature*, 424, 406
Wandel, A., Peterson, B. M., & Malkan, M. A. 1999, *ApJ*, 526, 579
White, R. L., Becker, R. H., Fan, X., & Strauss, M. A. 2003, *AJ*, 126, 1
White, S. D. M., & Rees, M. J. 1978, *MNRAS*, 183, 341
Willott, C. J., McLure, R. J., & Jarvis, M. J. 2003, *ApJ*, 587, L15
Woo, J.-H., & Urry, C. M. 2002, 579, 530
Woosley, S. E. 1993, *ApJ*, 405, 273
Wu, X., & Xue, Y. 2001, *ApJ*, 560, 544
Wyithe, S. 2003, *ApJ*, submitted, astro-ph/0308290
Wyithe, S., & Loeb, A. 2002a, *Nature*, 417, 923
Wyithe, S., & Loeb, A. 2002b, *ApJ*, 577, 57
Wyithe, S., & Loeb, A. 2003a, *ApJ*, 590, 691
Wyithe, S., & Loeb, A. 2003b, *ApJ*, 595, 614
Zaldarriaga, M. 1997, *Phys. Rev. D.*, 55, 1822
Zel'dovich, Y. B. 1964, *Dok. Akad. Nauk SSSR*, 155, 67
Zhao, H., Haehnelt, M. G., & Rees, M. J. 2002, *NewA*, 7, 385