HOW MUCH MASS DO SUPERMASSIVE BLACK HOLES EAT IN THEIR OLD AGE?

Philip F. Hopkins,[1] Ramesh Narayan,[1] and Lars Hernquist[1]
Received 2005 October 9; accepted 2006 February 2

ABSTRACT

We consider the distribution of local supermassive black hole Eddington ratios and accretion rates, accounting for the dependence of radiative efficiency and bolometric corrections on the accretion rate. We find that black hole mass growth, both of the integrated mass density and the masses of most individual objects, must be dominated by an earlier, radiatively efficient, high accretion rate stage, and not by the radiatively inefficient low accretion rate phase in which most local supermassive black holes are currently observed. This conclusion is particularly true of supermassive black holes in elliptical host galaxies, as expected if they have undergone merger activity in the past that would fuel quasar activity and rapid growth. We discuss models of the time evolution of accretion rates and show that they all predict significant mass growth in a prior radiatively efficient state. The only way to avoid this conclusion is through careful fine-tuning of the accretion/quasar timescale to a value that is inconsistent with observations. Our results agree with a wide range of observational inferences drawn from the quasar luminosity function and X-ray background synthesis models, but our approach has the virtue of being independent of the modeling of source populations. Models in which black holes spend the great majority of their time in low accretion rate phases are thus completely consistent both with observations implying mass gain in relatively short, high accretion rate phases and with the local distribution of accretion rates.

Subject headings: cosmology: theory — galaxies: active — galaxies: evolution — quasars: general

1. INTRODUCTION

The growth of quasars and supermassive black holes is a topic of fundamental interest in cosmology, critically informing theories of black hole formation and accretion models, and the role of black hole energetics in the X-ray, UV, and infrared backgrounds and their consequences for reionization. It is now believed that supermassive black holes reside at the centers of most if not all galaxies (e.g., Kormendy & Richstone 1995; Richstone et al. 1998; Kormendy & Gebhardt 2001). Furthermore, recent discoveries of correlations between masses of black holes in nearby galaxies and either the mass (Magorrian et al. 1998; McLure & Dunlop 2002; Marconi & Hunt 2003) or velocity dispersion (i.e., the M-σ relation; Ferrarese & Merritt 2000; Gebhardt et al. 2000) of spheroids demonstrate a fundamental link between the growth of supermassive black holes and galaxy formation.

Observations of the nearby universe suggest that rapid black hole growth may be related to gas inflows in the centers of galaxies, powering intense infrared emission (e.g., Sanders et al. 1988), with mergers or interactions of galaxies providing the gravitational torques necessary to drive large quantities of gas to galaxy centers (e.g., Barnes & Hernquist 1991, 1996) to fuel both starbursts (Mihos & Hernquist 1994, 1996) and rapid, self-regulating black hole growth (Di Matteo et al. 2005). This picture of quasar activity, whether represented in analytical or numerical models of the M-σ relation (Silk & Rees 1998; Fabian 1999; Ciotti & Ostriker 2001; Di Matteo et al. 2005), semianalytical models or computer simulations of galaxy and quasar formation (Kauffmann & Haehnelt 2000; Haiman & Menou 2000; Di Matteo et al. 2003, 2004; Wyithe & Loeb 2003; Granato et al. 2004), or modeling of quasar light curves and the quasar and elliptical galaxy luminosity functions (Hopkins et al. 2005a, 2005b, 2005c, 2005d, 2006a,

2006b, 2006c), implies that black hole growth should be dominated by radiatively efficient, relatively short phases at high accretion rates ($\gtrsim$10% of Eddington).

This picture is supported by inferred black hole accretion histories from the quasar luminosity function (e.g., Soltan 1982; Salucci et al. 1999; Yu & Tremaine 2002; Marconi et al. 2004; Shankar et al. 2004; Merloni 2004; Merloni et al. 2004) and synthesis models of the X-ray background (e.g., Comastri et al. 1995; Gilli et al. 1999; Elvis et al. 2002; Ueda et al. 2003; Cao 2005; Barger et al. 2005), as well as direct observations of quasar accretion rates at different redshifts (Vestergaard 2004; Heckman et al. 2004; McLure & Dunlop 2004) or models that combine all of these based on quasar evolution and the observed luminosity function (Hopkins et al. 2005b, 2006a). However, these inferences are generally model dependent, and more importantly are based on quasar observations that, while in some cases probing X-ray luminosities well below L^* and much fainter than bright optically selected quasars, do not include the (at least locally) very large low accretion rate, radiatively inefficient population. It is not clear how strongly constrained is the possibility of black holes gaining a significant amount of mass in much lower accretion rate phases after periods of early, bright activity; in these latter phases the radiative efficiency may drop, and thus the black holes will not usually be observed.

Various authors have attempted to use observational estimates of the distribution of quasar accretion rates as a function of redshift to answer the question of whether or not all black holes pass through and gain their mass in a high Eddington ratio bright quasar phase (e.g., Heckman et al. 2004; McLure & Dunlop 2004; Kollmeier et al. 2005). Although the determination of the distribution and evolution of quasar accretion rates remains a key outstanding problem, selection effects limit the questions these high-redshift observations can answer. These surveys are generally complete only down to accretion rates ~0.1 of Eddington, and at any given redshift only a small fraction of black holes are "active" with Eddington ratios above this fraction. Surveys that

[1] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138.

can probe lower accretion rates have yielded very different results for the number and distribution of objects with accretion rates in the range ~ 0.01–0.1 Eddington (compare, e.g., the estimates of Vanden Berk et al. [2005] from the Sloan Digital Sky Survey [SDSS] and Kollmeier et al. [2005] from the AGN and Galaxy Evolution Survey). Consequently, although the observations described above may suggest otherwise, these high-redshift measurements of the accretion rate distribution at large Eddington ratios cannot determine whether or not a significant population of black holes exist at accretion rates $\sim 1\%$ of Eddington, sufficient to dominate their mass growth and total mass density by $z = 0$.

Furthermore, estimating the mass growth of individual black holes (as opposed to the integrated growth of the population as a whole) in the observed high accretion rate phases is not directly possible, as the duration of this phase is degenerate with the rate at which quasars are "triggered" or "activated" at these rates. Even indirectly, neither quantity is well constrained, as, e.g., the duration of this phase is currently restricted observationally only to the range $\sim 10^6$–10^8 yr (Martini 2004).

At low accretion rates (relative to Eddington), black holes transition to radiatively inefficient/advection-dominated accretion flows (RIAF/ADAF; e.g., Narayan & Yi 1995b; Quataert 2001; Narayan 2006; Yuan & Narayan 2004), making these objects especially difficult to observe, and implying that observational surveys may be an order of magnitude less deep in terms of intrinsic accretion rate than they are in terms of a radiative ($L/L_{\rm Edd}$) Eddington ratio. It is therefore especially important to consider the distribution of accretion rates at low redshifts, where objects with low Eddington ratios can be reliably detected, as a constraint on possible black hole growth in low Eddington ratio states. The fact that the population of low Eddington ratio sources, which must exist at all moderate and low redshifts (owing to limits on the integrated black hole mass and X-ray luminosity density as discussed above), is observed at $z = 0$ to extend to accretion rates as low as 10^{-6} of Eddington strongly emphasizes this point, as these accretion rates are unlikely to be observable in high-redshift surveys in the near future.

In this paper, we consider the $z = 0$ observed distribution of accretion rates and determine whether or not it allows for significant mass growth in radiatively inefficient, low accretion rate states. There are three basic questions one could ask:

1. In which state—radiatively inefficient, low accretion rate, or radiatively efficient, high accretion rate—do black holes spend most of their *time*?
2. In which state do black holes gain most of their *mass*?
3. In which state do black holes radiate most of their *luminosity*?

According to current wisdom, the answer to question (1) is the radiatively *inefficient* state, and that to question (3) is the radiatively *efficient* state. However, question (2) is rather delicate; it is by no means straightforward to deduce its answer from the other answers. The purpose of the present paper is to investigate question (2) in depth using the $z = 0$ distribution of black hole masses and accretion rates.

We describe the observed sample from Ho (2002), compare with that of Marchesini et al. (2004), and discuss possible selection effects in § 2.1. In § 2.2 we discuss the dependence of bolometric corrections and radiative efficiency on accretion rate and our accounting for this in determining the distribution of accretion rates. In § 2.3 we consider the distribution of accretion rates as a function of host galaxy morphology. We discuss several models for the mass growth of black holes and their implications for the possibility of mass gain in low accretion rate states in § 3 and finally summarize our conclusions in § 4.

2. THE OBSERVED DISTRIBUTION OF LOW-EFFICIENCY ACCRETION RATES

2.1. *The Sample and Possible Selection Effects*

We consider the observed sample of Ho (2002), itself an update of the compilation of black hole masses in Ho (1999a), which includes black hole masses, B-band luminosities, redshifts, and host galaxy morphological information for 80 nearby supermassive black holes. The measurement methods for black hole masses and luminosities, and possible selection effects, are discussed in detail in Ho (2002), but we briefly review them here.

The majority of the black hole masses are derived from spatially resolved observations of gas and/or stellar kinematics (for a review of the methods, see, e.g., Kormendy & Richstone [1995]). There are a few exceptions, namely, Sgr A* in the Milky Way, whose mass is estimated based on proper motions of individual stars (e.g., Schödel et al. 2002; Eckart et al. 2002; Ghez et al. 2005), four galaxies that have strong water maser emission (NGC 1068, NGC 4945, Circinus, and NGC 4258; Greenhill et al. 1996, 1997, 2001; and Miyoshi et al. 1995, respectively), and Arp 102B, for which the orbital period of the accretion disk has been determined (Newman et al. 1997). In Ho (2002), the differences between masses estimated by different means of dynamical modeling of the integrated spectroscopy (primarily *Hubble Space Telescope* spectroscopy of stellar dynamics in the central regions of each galaxy) are discussed. Three-integral dynamical modeling of the galaxy surface brightness profiles and stellar absorption lines is used to estimate the black hole masses, which tends to give slightly lower black hole masses than the simplified axisymmetric two-integral modeling of, e.g., Magorrian et al. (1998) (van der Marel 1999; Ferrarese & Merritt 2000; Gebhardt et al. 2000).

For more luminous or distant active galactic nuclei (AGNs), direct dynamical measurements are no longer feasible, as the AGN luminosity will overpower the central stellar emission features. Thus, a large fraction of the sample masses are determined via reverberation mapping (Blandford & McKee 1982) of the central emission lines. The uncertainties of this method are discussed in Ho (1999a), Wandel et al. (1999), Kaspi et al. (2000), McLure & Dunlop (2001), and Krolik (2001), but it is generally true that black hole masses estimated via reverberation mapping agree with those determined from dynamical estimates to within a factor of ~ 2–3. The reverberation-mapped objects of Ho (2002) are compiled from Kaspi et al. (2000) and Ho (1999a), in which the individual objects are described further.

The optical (B-band) luminosities of the sources in the sample are determined using the correlation between Hβ luminosity and B-band absolute magnitude (Yee 1980; Shuder 1981), as calibrated by Ho & Peng (2001), which is in principle a more isotropic quantity than optical continuum luminosity and is measurable in obscured (type II) AGNs. Where the B-band magnitude is directly available (including all of the reverberation-mapped sources), it agrees reasonably well with this determination, with no significant systematic bias (Ho & Peng 2001; Ho 2002). We do not consider the 17 Palomar-Green quasars in the Ho (2002) sample, as these objects are chosen from yet another parent sample, at nonnegligible redshifts $z = 0.1$–0.2, introducing stronger selection effects. Moreover, they are all at reasonably large Eddington ratios (Ho 1999a, 2002) and thus are members of the class of higher redshift, high accretion rate radiatively efficient objects in which we are explicitly not interested. We have, however, checked that including them in our total sample results in a negligible difference in the cumulative Eddington ratio distribution and our subsequent calculations, after normalizing the relative fraction for the
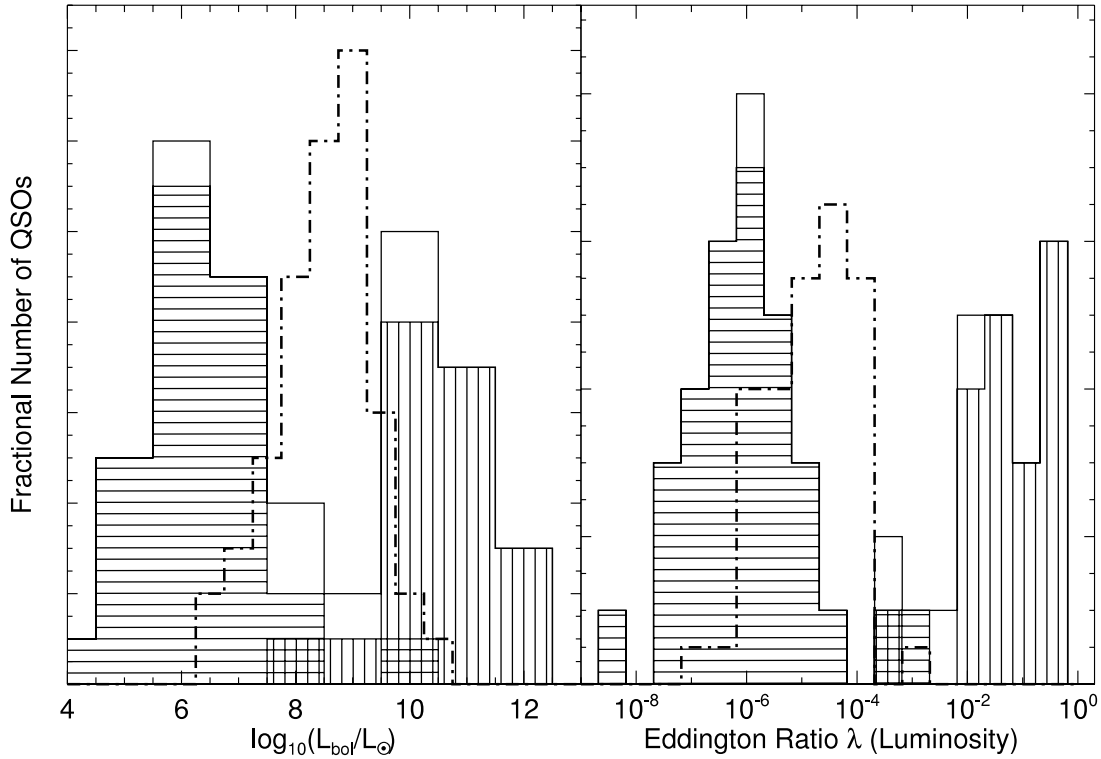
Fig. 1.—Distribution of bolometric luminosities (*left*) and radiative Eddington ratios ($\lambda \equiv L/L_{\mathrm{Edd}}$, *right*) in the Ho (2002) sample (*solid histograms*) and the Marchesini et al. (2004) radio-selected sample (*dot-dashed histograms*). Open histograms show the cumulative sample of Ho (2002), horizontally shaded show the distribution for objects with stellar or gasdynamical mass determinations, and vertically shaded show the distribution for objects with reverberation mapping mass determinations.

effective volume of the subsample (i.e., converting the observed number to a number density) in Figure 2.

We briefly consider possible selection effects of the different methods by which the black hole masses in this sample are measured. As noted above, strong or bright AGNs will overwhelm the stellar emission in the central regions of the observed galaxy and make direct measurements of stellar kinematics difficult. Therefore, the objects with masses measured via stellar kinematics may be systematically biased toward lower nuclear luminosities, and correspondingly lower Eddington ratios and accretion rates. However, the reverberation-mapped objects are expected to have the opposite bias, as this method of mass measurement requires resolving the AGN emission lines and thus may bias the observed sample to higher Eddington ratios and accretion rates.

Figure 1 shows these effects explicitly. Here, the distribution of bolometric luminosities of our entire adopted Ho (2002) sample is shown (*open histograms, left*). Since we simply wish to show the luminosity range sampled by the observations we use a constant bolometric correction ($L = 17L_B$), but we adopt more accurate bolometric corrections in § 2.2. The right panel shows the corresponding distribution of radiative Eddington ratios. Horizontally shaded histograms show the distribution for objects with masses determined using gas or stellar dynamics, whereas vertically shaded histograms show the distribution for objects with masses determined using reverberation mapping. There is a considerable gap in the luminosities probed around $10^9 \, L_{\odot}$ and a similar gap in Eddington-scaled luminosities around 10^{-4} to 10^{-2}. Owing to the minimal overlap in the bolometric luminosities covered by either method, it is difficult to reliably determine whether the observed gap in the Eddington ratio distribution is real or represents the combination of the selection effects of the two methods.

To test this, we consider the sample of black hole masses and bolometric luminosities from Marchesini et al. (2004), shown in both panels as dot-dashed histograms. We consider the Fanaroff-Riley I sample studied by these authors, based on the 298 radio galaxies and 53 radio-loud quasars of the Third Cambridge Revised Catalog of Radio Sources catalog (Spinrad et al. 1985). The black hole masses are estimated from the stellar luminosities of the host galaxy bulges (except for the bright radio-loud quasars, which we do not consider here for the same reason we neglect the high accretion rate quasars in the Ho [2002] sample, although again we have checked and found that the relative number density is not large enough to change our results). The bolometric luminosities are based on fitting the nuclear optical luminosities to the light profiles of the observed galaxies.

The advantage of considering the Marchesini et al. (2004) sample is clear in the left panel of Figure 1, as it covers the range of bolometric luminosities in which the Ho (2002) samples show a deficit, yielding a combined sample with a smooth (nonbimodal) luminosity distribution. The Eddington ratio distribution for these objects is shown in the right panel, and although it peaks at a higher Eddington ratio ($\sim 10^{-4}$) than the stellar/gas dynamics–measured sample of Ho (2002), it has a similar qualitative behavior. Furthermore, this peak being at higher Eddington ratio most likely owes to the higher luminosity limit in the Marchesini et al. (2004) sample.

Although including the Marchesini et al. (2004) sample fills the gap in bolometric luminosity, the gap in Eddington ratio at 10^{-3} to 10^{-2} is preserved, a point that is discussed in Marchesini et al. (2004). If we consider the combined distribution from the Ho (2002) and Marchesini et al. (2004) samples, the low Eddington ratio distribution peaks at roughly 10^{-5}, more than low enough so that our subsequent results are all qualitatively unchanged regardless

of whether or not we include the objects in the Marchesini et al. (2004) sample. Based on this, the "gap" at Eddington ratios $\sim 10^{-2}$ does not appear to be a selection effect, and we can rely on the Ho (2002) sample in our following analysis.

In the following, we generally do not include the sample of Marchesini et al. (2004) except where otherwise specified, because it is unclear how to correct for possible bias introduced in considering a radio-loud sample, given the correlation between Eddington ratio and radio loudness observed in Ho (2002). For further discussion of these distinctions between samples and possible selection effects, we refer to Ho (2002), Marchesini et al. (2004), and Jester (2005).

2.2. *Determining the Accretion Rate*

The bolometric luminosity, which we denote by L, of a black hole accreting at a rate $\dot{M}$ is $L = \epsilon_r \dot{M} c^2$, where ϵ_r is the radiative efficiency. Therefore, the Eddington rate of accretion for a black hole of mass M is $\dot{M}_{\rm Edd} = L_{\rm Edd}(M)/\epsilon_r c^2$, where the Eddington luminosity of a black hole is well defined,

$$L_{\rm Edd} \approx 3.3 \times 10^4 \, L_\odot (M/M_\odot). \tag{1}$$

Because the Eddington accretion rate depends on the radiative efficiency ϵ_r, which itself can depend on the accretion rate, these definitions are circular unless we define the Eddington accretion rate with respect to a canonical radiative efficiency ϵ_r^0. We follow standard practice (e.g., Narayan & Yi 1995b; Esin et al. 1997; Jester 2005; but see also Chen et al. 1995; Blandford & Begelman 1999; Marchesini et al. 2004, who all choose $\epsilon_r^0 = 1$) in defining

$$\epsilon_r^0 = 0.1, \tag{2}$$

which corresponds to the typical radiative efficiency expected from accretion through a standard efficient Shakura & Sunyaev (1973) thin disk. We then define the Eddington rate as

$$\dot{M}_{\rm Edd} = \frac{L_{\rm Edd}(M)}{\epsilon_r^0 c^2} = \frac{M}{t_{\rm S}}, \tag{3}$$

where

$$t_{\rm S} = 4.2 \times 10^7 \; {\rm yr} \tag{4}$$

is the Salpeter time, the mass e-folding time for a black hole accreting at the Eddington rate with $\epsilon_r = 0.1$ (Salpeter 1964). We can further define the dimensionless luminosity Eddington ratio λ and mass accretion rate Eddington ratio $\dot{m}$ by

$$\lambda \equiv L/L_{\rm Edd}, \qquad \dot{m} \equiv \dot{M}/\dot{M}_{\rm Edd}. \tag{5}$$

With these definitions, we have

$$\lambda = \frac{L}{\epsilon_r^0 M c^2/t_{\rm S}} = \dot{m} \, \frac{\epsilon_r(\dot{m})}{\epsilon_r^0}, \tag{6}$$

where we allow explicitly for the possibility that the radiative efficiency may depend on $\dot{m}$.

We can also define the black hole growth timescale, which corresponds to the mass e-folding time for growth at a constant Eddington ratio,

$$t_{\rm BH} = \frac{M}{\dot{M}} = \frac{1}{\dot{m}} \, t_{\rm S}. \tag{7}$$

Note that this further lets us define a special value of $\dot{m}$,

$$\dot{m}_{\rm H} \equiv t_{\rm S}/t_{\rm Hubble} \approx 0.003, \tag{8}$$

which is the dimensionless accretion rate for which the growth time is equal to the Hubble time (~ 14 Gyr). For significantly lower accretion rates $\dot{m} \ll \dot{m}_{\rm H}$, the growth timescale is much greater than the age of the universe, and a given black hole cannot possibly gain a significant fraction of its mass at that $\dot{m}$.

For a simple estimate of the accretion luminosity L, we have followed Ho (2002) and converted from B-band to bolometric luminosities adopting a constant bolometric conversion ($L = c_B L_B$, $c_B = 17$). However, the actual bolometric conversion varies, ranging from ~ 11 to 17 for bright AGNs and quasars (e.g., Elvis et al. 1994) to ~ 24 (e.g., Ho 1999b; Ho et al. 2000) for the most low-luminosity RIAF/ADAF systems. For bright AGNs and quasars, the bolometric conversion c_B appears to depend just on luminosity, although this of course may instead reflect a dependence on the accretion rate. For objects with moderate to large accretion rates, we adopt the luminosity-dependent bolometric conversions of Marconi et al. (2004), based on observations of the quasar spectrum over a wide range of wavelengths and as a function of luminosity (e.g., Elvis et al. 1994; George et al. 1998; Vanden Berk et al. 2001; Perola et al. 2002; Telfer et al. 2002; Ueda et al. 2003; Vignali et al. 2003). This gives a B-band luminosity

$$\log (L_B) = 0.80 - 0.067\mathcal{L} + 0.017\mathcal{L}^2 - 0.0023\mathcal{L}^3, \tag{9}$$

where $\mathcal{L} = \log (L/L_\odot) - 12$. The important point is that c_B *decreases* with increasing luminosity; i.e., the brightest objects are the most dominated by the optical-UV portion of the spectrum (see also, e.g., Wilkes et al. 1994; Green et al. 1995; Strateva et al. 2005).

At low luminosities, we consider the subsample of low-luminosity AGNs from Ho (1999b) and Ho et al. (2000), for which bolometric luminosities from radio, optical, UV, and X-ray observations have been determined directly. Calculating the c_B correction for each of these objects, we find no evidence for a direct correlation with L, but we do find marginal evidence for a correlation of c_B with observed Eddington ratio $\lambda = L/L_{\rm Edd}$. We can fit this dependence roughly to a power law and obtain the best fit,

$$c_B \approx 11(\lambda/0.01)^{-0.25}. \tag{10}$$

As essentially all these objects have $\dot{m} \lesssim \dot{m}_{\rm crit}$ (defined below, but roughly $\dot{m}_{\rm crit} \sim 0.01 - 0.1$), we do not need to worry about a "break" in the slope in terms of λ induced by the change in dependence of ϵ_r on λ (i.e., on $\dot{m}$) at $\dot{m}_{\rm crit}$. Again, we find that the ratio of total to B-band luminosity decreases with increasing luminosity (although in this case with dimensionless luminosity). This is also consistent with theoretical spectral models of advection-dominated accretion (e.g., Mahadevan 1997), which generally feature an increasingly hard X-ray spectrum and possibly increased contribution from jets at low-$\dot{m}$.

In order to determine which objects should receive which bolometric correction, we divide our sample in two on the basis of our original calculation of λ. For objects with $\lambda > \dot{m}_{\rm crit}$ (defined below), we apply the bolometric corrections of Marconi et al. (2004), which are primarily derived from bright, high accretion rate objects. For objects with $\lambda < \dot{m}_{\rm crit}$, we apply the corrections derived from Ho (1999b) and Ho et al. (2000), whose samples are essentially all below this accretion rate. Although the exact point at which we determine the divide in bolometric corrections

is somewhat arbitrary, this choice is consistent with theory and observations in both cases, and it is not a significant uncertainty, as the two sets of corrections give similar results for the objects in the transition region.

Observations of X-ray binaries have revealed a number of distinct spectral states, such as the high/soft state, low/hard state, quiescent state, etc. (see McClintock & Remillard [2006] for a review). Moreover, the same black hole can exhibit different states at different times. A primary cause of the various states is changes in the mode of accretion, as suggested by Narayan (1996) and Esin et al. (1997). According to this model, accretion flows with $\dot{m}$ greater than a critical value $\dot{m}_{\mathrm{crit}}$ are radiatively efficient and are well described by a thin-disk model. These systems correspond to the high/soft spectral state. However, for $\dot{m} < \dot{m}_{\mathrm{crit}}$, the accretion flow switches to a two-zone state in which the thin disk is restricted to radii R greater than a transition radius R_{trans}, while the gas inside R_{trans} accretes via a RIAF, also known as an ADAF (Narayan & Yi 1994, 1995b; Narayan 2006). This is the low/hard state, or (at very low $\dot{m}$) the quiescent state. The transition radius R_{trans} is only slightly larger than the radius of the innermost stable circular orbit when $\dot{m}$ is just below $\dot{m}_{\mathrm{crit}}$, but it increases to larger values as $\dot{m}$ decreases further (see Yuan & Narayan 2004). Consequently, the radiative efficiency is fairly close to $\epsilon_r^0 = 0.1$ when $\dot{m}$ is just below $\dot{m}_{\mathrm{crit}}$, but it decreases rapidly (but smoothly) as $\dot{m}$ decreases below $\dot{m}_{\mathrm{crit}}$ (see eq. [11]).

Observations suggest that the different accretion states observed in X-ray binaries exist as well in AGNs (Narayan et al. 1995, 1996; Meier 2001; Maccarone et al. 2003; Yuan & Narayan 2004). It appears that the transition between accretion states occurs at more or less the same critical Eddington ratio $\dot{m} = \dot{m}_{\mathrm{crit}}$, regardless of the mass of the accreting black hole. Although the value of $\dot{m}_{\mathrm{crit}}$ is somewhat uncertain (see the discussion below on hysteresis), observations of black hole binaries (Maccarone 2003) as well as theoretical extensions of accretion models (e.g., Narayan & Yi 1995b; Esin et al. 1997; Meyer et al. 2000) suggest $\dot{m}_{\mathrm{crit}} \sim 0.01$. Recent observations of both a bimodal distribution of Eddington ratios at low redshift (Marchesini et al. 2004) and the distribution of objects in the L-M plane (Jester 2005) also suggest that the transition occurs at $\dot{m} \sim 0.01$ (see Cao & Xu 2005). Note that the similarity between $\dot{m}_{\mathrm{crit}}$ and $\dot{m}_{\mathrm{H}}$ is entirely coincidental.

In accretion models, the radiative efficiency does not depend on the absolute value of M or $\dot{M}$, but only on the dimensionless accretion rate $\dot{m}$ (e.g., Chen et al. 1995; Esin et al. 1997; Abramowicz 2005). Based on observations and theoretical models, we model the radiative efficiency $\epsilon_r(\dot{m})$ as

$$\epsilon_r = \begin{cases} \epsilon_r^0 & \text{if } \dot{m} > \dot{m}_{\mathrm{crit}}, \\ \epsilon_r^0 \left(\dfrac{\dot{m}}{\dot{m}_{\mathrm{crit}}} \right) & \text{if } \dot{m} \leq \dot{m}_{\mathrm{crit}}. \end{cases} \tag{11}$$

This particular choice for the efficiency factor was originally suggested by early ADAF models (Narayan & Yi 1995b), but it is generally consistent with observations and ensures that the radiative efficiency is continuous across the critical Eddington ratio $\dot{m}_{\mathrm{crit}}$. With this model, we can determine the accretion rate $\dot{m}$ for any black hole given the bolometric luminosity L and black hole mass M. The model of $\epsilon_r(\dot{m})$ may actually overestimate the accretion rate for a given luminosity at low $\dot{m} \ll \dot{m}_{\mathrm{crit}}$. This is because ADAFs tend to suffer large mass loss through winds (Narayan & Yi 1994, 1995a; Blandford & Begelman 1999). As a result, $\dot{m}$ decreases with decreasing radius, so that the efficiency depends on whether one considers the accretion rate at the

black hole or at the transition radius. In terms of $\dot{m}$ at the black hole, which is the quantity of interest for this paper, ϵ_r decreases less steeply with decreasing $\dot{m}$, say as $\sim (\dot{m}/\dot{m}_{\mathrm{crit}})^{1/2}$; i.e., the accretion rate for a given luminosity would be less than equation (11) predicts. However, for our analysis we choose the prescription given in equation (11), which is the conservative choice since it errs on the side of overestimating the accretion rate at low λ.

What value of $\dot{m}_{\mathrm{crit}}$ should we use? It has been observed in X-ray binaries that there is a hysteresis phenomenon in the transition between radiatively inefficient and radiatively efficient accretion flows (see, e.g., Maccarone & Coppi 2003 and references therein; Barret & Olive 2002). The transition from the radiatively inefficient state (ADAF) to the radiatively efficient (thin-disk) state often occurs at a fairly large accretion rate $\dot{m} \sim 0.1$, whereas the reverse transition (from radiatively efficient to inefficient accretion) occurs at a much lower $\dot{m} \sim 0.01$. This implies that when $\dot{m}$ increases with time, the critical accretion rate may be as high as $\dot{m}_{\mathrm{crit}} \sim 0.1$, whereas when $\dot{m}$ decreases with time, $\dot{m}_{\mathrm{crit}} \sim 0.01$. Since the peak of quasar activity occurred at high redshift and the activity declined more or less monotonically to the present epoch, one expects that the lower value of $\dot{m}_{\mathrm{crit}} \sim 0.01$ is more relevant for our analysis. However, the possibility of hysteresis does introduce a degree of uncertainty in our model. Therefore, we consider in our subsequent analysis both $\dot{m}_{\mathrm{crit}} = 0.01$ and 0.1, and we ultimately find that the choice does not qualitatively change our conclusions.

Figure 2 (top) shows the distribution of dimensionless accretion rates $\dot{m}$ obtained from the observed sample using equation (11). The top left panel shows the result for $\dot{m}_{\mathrm{crit}} = 0.1$, the top right panel for $\dot{m}_{\mathrm{crit}} = 0.01$, essentially bracketing the range allowed by the hysteresis effect discussed above. In each case, we show the distribution for all objects (open histograms) and for just those objects that are morphologically identified as elliptical galaxies (morphological type index $T \leq -2$, shaded histograms). We also show the value of $\dot{m}_{\mathrm{H}}$ (dashed line), the accretion rate for which the black hole growth timescale is equal to the Hubble time. Although the result is slightly stronger for the elliptical subsample, the critical point is the same in both cases: even after accounting for the fact that the black holes are in a radiatively inefficient state and thus the actual accretion rate is significantly larger than the observed luminosity Eddington ratio, most local objects are accreting at Eddington ratios $\dot{m} \ll \dot{m}_{\mathrm{H}}$, i.e., at very low accretion rates at which significant fractional mass gain in less than the age of the universe is not possible. Although the choice of $\dot{m}_{\mathrm{crit}}$ systematically shifts the inferred $\dot{m}$ at very low $\dot{m} \ll \dot{m}_{\mathrm{crit}}$, it has almost no effect in the range of interest, $\dot{m} \sim 10^{-3}$ to 10^{-1}. The bolometric corrections we adopt also slightly increase the characteristic $\dot{m}$ over that implied by, e.g., a constant $c_B = 17$ (in other words, simpler bolometric corrections only strengthen this point), although the difference is negligible for all but the smallest ($\dot{m} \lesssim 10^{-4}$) and largest ($\dot{m} \gtrsim 0.1$) accretion rates.

We can also attempt to use the observed sample accretion rate distribution to (very roughly) infer that of the population as a whole. We estimate a simple visibility volume correction, using the distances to each object in Ho (2002). Comparing the maximum distances or distances within which $\gtrsim 90\%$ of each (dynamical or reverberation-mapping mass measurements) subsample are located, this gives a rough ratio of the subsample effective volumes. Alternatively, we correct the relative number density of sources in each L bin to the value implied by the $z \sim 0$ low-luminosity SDSS AGN luminosity function of Hao et al. (2005). In either case, the estimate is similar. Whether or not we include the Marchesini et al. (2004) sample in this estimate also makes little difference. In the bottom panels of Figure 2 we use this
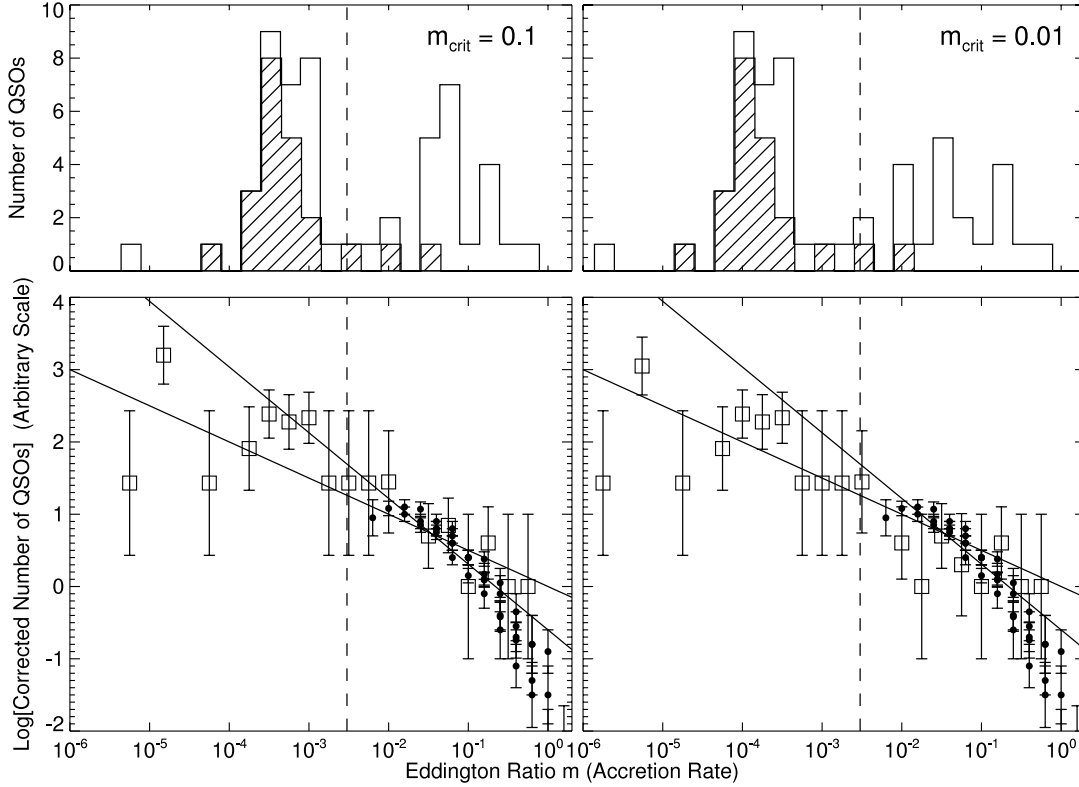
Fig. 2.—*Top:* Distribution of dimensionless accretion rate ($\dot{m} \equiv \dot{M}/\dot{M}_{\rm Edd}$) calculated through eq. (11) and binned by $\Delta \log \dot{m} = 0.25$, from the sample of Ho (2002). Left panels assume $\dot{m}_{\rm crit} = 0.1$, and right panels assume $\dot{m}_{\rm crit} = 0.01$. Shaded histograms show the distribution for AGNs with elliptical host galaxies only. *Bottom:* Estimated relative number density of objects as a function of $\dot{m}$ (*points*), corrected using a simple visibility volume correction (alternatively, corrected to match the local AGN luminosity functions of Hao et al. [2005]) for the combined samples of Ho (2002) and Marchesini et al. (2004, *squares*). Circles show the SDSS Eddington ratio distributions determined in Yu et al. (2005) for systems with velocity dispersions $\sigma = 70-110$ km s^{-1}. The vertical dashed line in all panels corresponds to the accretion rate below which the black hole growth timescale $M/\dot{M}$ is larger than the age of the universe.

correction (technically, the luminosity function correction including the Marchesini et al. [2004] objects) to estimate a *relative* $\dot{m}$ distribution (*squares*), with assumed Poisson errors (logarithmically scaled). We caution that these distributions are rough estimates and should be taken heuristically, as in truth the selection effects involved in measuring individual black hole masses are much more complicated than, e.g., a simple magnitude limit. We also show the observed SDSS Eddington ratio distributions estimated in Yu et al. (2005), in a more robust statistical sense, but for relatively high $\dot{m} \gtrsim 10^{-3}$ to 10^{-2}. Again, the normalization of these is essentially arbitrary, but the trend with $\dot{m}$ should be robust. We show these observations as the circles, for systems with velocity dispersions $\sigma = 70, 80, 90, 100,$ and 110 km s^{-1}.

In all cases, the $\dot{m}$ distributions broadly agree and suggest that the $\dot{m}$ distribution can be approximated as a power law (or series of power laws) over a wide range in $\dot{m}$. The solid lines in the bottom panels of Figure 2 show power laws with slopes of $-1/1.23$ estimated from Yu et al. (2005) and $-\frac{1}{2}$ predicted in Hopkins et al. (2006a), which give reasonable approximations over a wide range of $\dot{m}$ and demonstrate the range in slope allowed. For a power law $\dot{m}$ down to a turnover at $\dot{m}_{\rm min}$, the differential duty cycle takes the form

$$\frac{df}{d \log \dot{m}} = (\beta \ln 10 \dot{m}_{\rm min}^{\beta}) \dot{m}^{-\beta}. \qquad (12)$$

Essentially all of the observational uncertainty regarding duty cycles, the proper effective volume and number density corrections for these samples, and the depth at which the $\dot{m}$ distribution turns over (as compared to, e.g., selection incompleteness)

is neatly contained in the parameter $\dot{m}_{\rm min}$. The estimates from observations in Figure 2 do at least provide a reasonable upper limit $\dot{m}_{\rm min} \lesssim 10^{-5}$, which allows us to calculate the quantities of particular interest (e.g., upper limits to the mass gain in low-efficiency phases), given that β is also reasonably well constrained in the range above. Although these estimates are by no means a rigorous determination of the $\dot{m}$ distribution, they do provide approximate constraints and upper limits, which imply that our subsequent calculations (e.g., the mass-weighted effective $\dot{m}$ and ϵ_r) cannot be qualitatively changed within the allowed uncertainty in Figure 2.

2.3. Accretion Rates as a Function of Host Morphology

Using the intrinsic mass accretion rates estimated above, accounting for both the dependence of radiative efficiency on accretion rate and spectral shape (bolometric corrections) on luminosity and accretion rate, we can consider whether there is a difference in the distribution of accretion rates in systems of different morphological types. In Figure 3, we show the black hole growth timescale $t_{\rm BH} = t_{\rm S}/\dot{m}$ for each object in the sample of Ho (2002) as a function of the host galaxy morphology (morphological type index), where the host galaxy Hubble types are determined in de Vaucouleurs et al. (1991). The dashed line in the figure shows the Hubble time, $t_{\rm H} \approx 14$ Gyr. It is clear that essentially all objects with moderate to large accretion rates (i.e., $t_{\rm BH} < t_{\rm H}$) are spirals, whereas almost every elliptical has a negligible mass accretion rate ($t_{\rm BH} \gg t_{\rm H}$). This is also clear in Figure 2 and is well known from a number of other observations (see, e.g., Heckman et al. 2004).

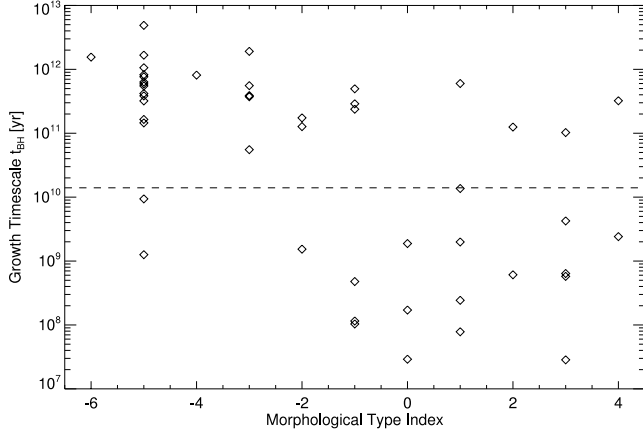FIG. 3.—Black hole growth timescale $t_{BH} = t_S/\dot{m}$ as a function of host galaxy morphology, for the sample from Ho (2002). The horizontal dashed line marks the Hubble time.
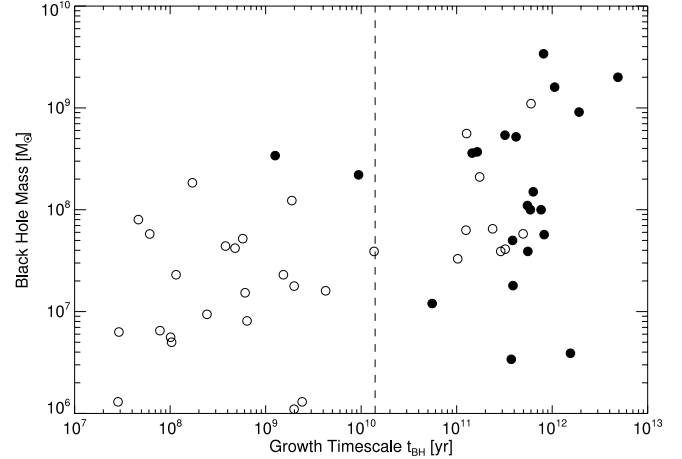

FIG. 4.—Black hole growth timescale as a function of black hole mass; open circles are the entire sample, and filled circles are elliptical galaxies. The vertical dashed line shows the Hubble time.

Observational studies of the $z = 0$ black hole mass function estimated from observations of galaxy luminosity functions and the distribution of spheroid velocity dispersions have established that the total present black hole mass density is dominated by black holes in elliptical hosts, with at most a comparable contribution from black holes in S0 host galaxies (e.g., Aller & Richstone 2002; Marconi et al. 2004; Shankar et al. 2004). If, as is clear in Figure 3, the large majority of present elliptical galaxies have $t_{BH} \gg t_H$, this implies that the majority of the present black hole mass density cannot have been accumulated in the observed low accretion rate phase, but that ellipticals must have experienced a previous phase with high-efficiency accretion.

Furthermore, regardless of whether or not all bright quasar activity is the result of galaxy mergers or interactions, it is difficult to imagine a major merger that does *not* excite bright quasar activity (so long as a reasonable amount of cold, rotationally supported gas is present in either of the merging galaxies). If elliptical galaxies are formed from the merger of gas-rich spiral progenitors, as is expected in hierarchical scenarios of structure formation (e.g., Toomre & Toomre 1972; Toomre 1977; for a review, see Barnes & Hernquist [1992]), then it is difficult to construct a scenario whereby an elliptical galaxy forms without *some* period of rapid, high-$\dot{m}$ accretion, which need only occur for $\gtrsim 10^6$ yr to dominate an entire Hubble time worth of accretion at the present ($z = 0$) rate. Indeed, the gas dissipation that would fuel the growth of supermassive black holes is required if galaxy mergers are to explain the high phase-space densities (e.g., Hernquist et al. 1993), kinematic properties (e.g., Cox et al. 2006), and fundamental scaling relations (e.g., Robertson et al. 2006) of ellipticals.

3. MODELS FOR THE EVOLUTION OF BLACK HOLE MASS

3.1. *If Black Holes Accrete at Constant $\dot{m}$*

Could each black hole accrete at roughly a constant $\dot{m}$ during its entire lifetime? If this is the case, then objects with $\dot{m} \ll \dot{m}_H = 0.003$ have $t_{BH} \gg t_H$ and do not gain mass in a Hubble time, whereas objects with $\dot{m} \gtrsim 10^{-2}$ with $t_{BH} \ll t_H$ experience many e-foldings. Thus, there should be a strong change in the black hole mass distribution at $t_{BH} = t_H$, with much higher black hole masses for $t_{BH} < t_H$. Figure 4 shows the black hole mass as a function of growth time t_{BH} for each object in our sample (*filled circles*, elliptical galaxies; *open circles*, late-type galaxies), with the dashed line marking the Hubble time. Clearly, the predicted trend is not evident; if anything, the most massive objects in our sample

have the *lowest* accretion rates (highest t_{BH}). This trend has been studied in greater statistical detail by Heckman et al. (2004).

3.2. *Models Involving Time Evolution of $\dot{m}$*

The conclusion of § 3.1 is that $\dot{m}$ must vary with time and that the mass growth in many systems must occur in an earlier phase of high-$\dot{m}$ accretion. We now ask whether low-efficiency, low accretion rate growth can give a significant fractional contribution to the absolute black hole mass density or individual black hole masses at $z = 0$. The answer to this depends on the time history of the accretion rate, for which we need to adopt some reasonable model.

Consider a general model in which $\dot{m} = 1$ at early times and then decays as a power law in time, peak phase with $\dot{m} = 1$,

$$\dot{m} = 1/\left[1 + (t/t_Q)^\eta\right]. \tag{13}$$

Any such model will naturally produce a power-law distribution of accretion rates at low-$\dot{m}$ (eq. [12]) with $\beta = 1/\eta$ and $\dot{m}_{min} = (t_H/t_Q)^{-\eta}$ (and in the appropriate limits can also represent exponential or step-function decay). The observed $\dot{m}$ distributions are broadly consistent with the above model for $\eta \sim 1.25$–2.0 (see Fig. 2). Simulations of quasar light curves (Hopkins et al. 2006c) suggest $\eta \approx 2$, as does connecting $z = 0$ accretion rates at $t \sim t_H$ to a bright $\dot{m} \sim 1$ phase of duration $\sim 10^7$–10^8 yr implied by a wide range of observations (see Martini 2004 and references therein). We therefore adopt $\eta = 2$ (this choice also allows analytical solutions below), but our calculations are only changed by factors $\sim \eta/2$ for other values of η. Simulations and theoretical models (e.g., Silk & Rees 1998; Fabian 1999; Ciotti & Ostriker 2001; Wyithe & Loeb 2003; Di Matteo et al. 2003, 2004, 2005), a wide range of observations (e.g., Martini 2004), and our estimated upper limit $\dot{m}_{min} \lesssim 10^{-5}$ (simply from the existence of a significant population of low-$\dot{m}$ observed elliptical galaxies) imply a bright (high-$\dot{m}$) quasar lifetime $\sim 10^6$–10^8 yr, setting a fairly strict upper limit $t_Q \lesssim t_S \sim 4 \times 10^7$ yr.

The above model of $\dot{m}(t)$ with $\eta = 2$ gives, for an initial black hole mass M_0, the following result for the mass growth as a function of time:

$$M(t) = M_0 \exp\left[\frac{t_Q}{t_S} \tan^{-1}\left(\frac{t}{t_Q}\right)\right]; \tag{14}$$

i.e., the object grows exponentially as $M = M_0 \exp(t/t_S)$ for $t \ll t_Q$ and asymptotes to $M_f = M_0 e^{(\pi/2)(t_Q/t_S)}$ as $t \rightarrow \infty$. If $t_Q \gg t_S$, then growth is large, but this is because the black hole spends a great deal of time in bright, high-$\dot{m}$ phases before the accretion rate slowly drops off. In fact, even after a Hubble time, the accretion rate would still be large. To reach $\dot{m} \leq 0.01$ by $t = t_H$ requires $t_Q \lesssim 3t_S$, yielding a small mass gain at low-$\dot{m}$; the maximum fraction of the present black hole mass that can be accreted at $\dot{m} < 0.01$ by the present is $\sim 8\%$ with $t_Q \approx 1.8t_S$.

More complex and host galaxy–dependent quasar light curves have been obtained in hydrodynamic simulations of galaxy mergers incorporating gas cooling, a multiphase interstellar medium, Bondi-Hoyle accretion determined from the surrounding gas properties, and thermal feedback from accretion (Springel & Hernquist 2002, 2003; Springel et al. 2005b). Despite the fact that these light curves show that most objects spend significantly more time at luminosities and accretion rates well below the peak quasar luminosity (Hopkins et al. 2005b, 2005c, 2006a), the authors find that the mass growth is dominated by efficient phases, with $\gtrsim 70\%$–80% of mass accumulated in bright, optically observable $\dot{m} \sim 1$ quasar phases (rising to $\sim 100\%$ of the mass in the most massive objects $M \gtrsim 10^9 \ M_\odot$), and the remaining mass essentially entirely acquired in the short following phase of declining $\dot{m}$ at moderate $\dot{m} \sim 0.1$ (Hopkins et al. 2006a).

3.3. The Mass-Weighted Eddington Ratio and Radiative Efficiency

Using the parameterization of the quasar light curve given in equation (13), we can define an "effective" mass-weighted accretion rate $\langle \dot{m} \rangle$, i.e.,

$$\langle \dot{m} \rangle = \frac{1}{M_f} \int \dot{m}(t) \dot{M} \, dt, \quad (15)$$

which can be evaluated numerically for a given t_Q/t_S. This gives $\langle \dot{m} \rangle \approx 0.50, \ 0.34$ for $t_Q = 10^7$ yr and $t_Q = t_S$, respectively, in good agreement with other estimates of the typical accretion rates at which black holes gain most of their mass (e.g., Kollmeier et al. 2005). It requires a very large $t_Q \gtrsim 10t_S$ to give low $\langle \dot{m} \rangle \lesssim 0.02$, but such long quasar lifetimes are ruled out by a number of arguments as discussed above. Using the fitted power laws of Yu et al. (2005) gives a significantly shallower decay, $\eta \approx 1.23$, but similar $\langle \dot{m} \rangle \approx 0.21, \ 0.14$ for $t_Q = 10^7$ yr and $t_Q = t_S$. If we use the binned $\dot{m}$ distribution of Figure 2 directly, we obtain $\langle \dot{m} \rangle \approx 0.3, \ 0.2$ for $t_Q = 10^7$ yr and $t_Q = t_S$, with a negligible difference if we assume $\dot{m}_{crit} = 0.01$ or 0.1. The difference due to how we estimate the correction to an "intrinsic" $\dot{m}$ distribution is also small compared to the expected Poisson noise from the small sample. Assuming the distribution cuts off at the lowest observed $\dot{m}$ sets an upper limit on the duty cycle (lower limit on $\langle \dot{m} \rangle$) independent of t_Q (see eq. [12]), giving $\langle \dot{m} \rangle \gtrsim 0.18$.

We can similarly define an effective mass or luminosity-weighted radiative efficiency, replacing $\dot{m}$ with ϵ_r in equation (15) above. Our "canonical" radiative efficiency $\epsilon_r^0 = 0.1$ can be factored out of this equation, and we can then numerically calculate the resulting $\langle \epsilon_r \rangle$ for a given t_Q/t_S and $\dot{m}_{crit}$. For a relatively large $t_Q = t_S$, we obtain $\langle \epsilon_r \rangle = 0.81\epsilon_r^0(\dot{m}_{crit} = 0.1); \ 0.93\epsilon_r^0(\dot{m}_{crit} = 0.01)$. For a shorter timescale $t_Q \sim 10^7$ yr these estimates increase to $0.95\epsilon_r^0, \ 0.98\epsilon_r^0$. Likewise, if we adopt just the observed distribution of accretion rates as the probability distribution function for accretion rate over each AGN lifetime as for $\langle \dot{m} \rangle$, we obtain $\langle \epsilon_r \rangle = 0.92\epsilon_r^0(\dot{m}_{crit} = 0.1); \ 0.96\epsilon_r^0(\dot{m}_{crit} = 0.01)$ for $t_Q \sim 10^7$ yr, and $\langle \epsilon_r \rangle = 0.72\epsilon_r^0(\dot{m}_{crit} = 0.1); \ 0.85\epsilon_r^0(\dot{m}_{crit} = 0.01)$ for $t_Q \sim t_S$ (and values closer to $1.0\epsilon_r^0$ for $t_Q < t_S$). If we were to

weight by luminosity instead of mass, we would be even further biased toward $\langle \epsilon_r \rangle \approx \epsilon_r^0$.

Thus, for reasonable values of the quasar lifetime, the effective radiative efficiencies expected are quite similar to the canonical values adopted for bright quasars; essentially, this is a restatement of our previous derivation that most mass/luminosity is accumulated/radiated in radiatively efficient, high accretion rate phases.

3.4. Comparison with Expected Quiescent Accretion Rates

We can also attempt to estimate the accretion rates and growth timescales of the elliptical galaxies observed. If we assume that the gas in these galaxies is virialized and is in spherical hydrostatic equilibrium and adopt a Hernquist (1990) profile for the static potential (set by stars and dark matter), we can determine the central sound speed and density and estimate the accretion rate at the Bondi-Hoyle rate. Using the M-σ and M-M_{sph} relations measured by Tremaine et al. (2002) and Marconi & Hunt (2003), respectively, we obtain

$$t_{BH} = \alpha \, \frac{1}{f_{gas}} \, \frac{M_{sph}}{M_{BH}} \, t_{dyn}$$
$$\sim 10^{12} \text{ yr } \alpha \left(\frac{0.01}{f_{gas}} \right) \left(\frac{M_{M\text{-}\sigma}}{M} \right) \left(\frac{\sigma}{200 \text{ km s}^{-1}} \right), \quad (16)$$

where $\alpha \sim 1$ is a constant that depends on the profile and gas equation of state, ≈ 0.6 for the assumptions above and $\gamma = 5/3$, t_{dyn} is the spheroid dynamical time, f_{gas} is the gas mass fraction, and $M_{M\text{-}\sigma}$ is the expected M-σ black hole mass.

Comparison with Figure 4 suggests that equation (16) provides a good order-of-magnitude estimate for the elliptical galaxies in our observed sample (typical $M \sim 10^8 \ M_\odot$). Of course, there is no particular evidence in the observations for the weak dependence on σ given by our model. But this further suggests that these objects must have gone through a previous bright phase, to heat the gas to virial temperatures and lower the gas fraction to $\sim 1\%$ typical in observed ellipticals. Moreover, note that $t_{BH} \propto M_{M\text{-}\sigma}/M$; i.e., if a black hole is undermassive, it will take longer to grow, meaning it is unlikely that these black holes grew to their observed M-σ masses via slow, quiescent accretion.

Finally, in essentially all cases in which the Bondi accretion rates of low-$\dot{m}$ AGNs have been determined (Fabian & Canizares 1988; Di Matteo et al. 2000, 2001; Loewenstein et al. 2001; Bower et al. 2003; Pellegrini 2005), as well as the same in X-ray binaries (e.g., Perna et al. 2003), it is observed that objects are actually accreting well *below* the Bondi rate. Thus, the fact that our estimate above provides an approximate agreement with the observed growth timescale suggests that we may actually be significantly overestimating the accretion rates of low-$\dot{m}$ AGNs, making our conclusions even stronger. This may be related, for example, to the fact that in the Milky Way case, an external observer would likely associate a significant amount of nearby but (owing to our proximity compared to AGNs in other galaxies) separately resolved emission with Sgr A*, because the AGN luminosity is not large enough to overwhelm these sources (see the discussion in Ho 2002).

4. CONCLUSIONS

We have determined the distribution of accretion rates of $z = 0$ supermassive black holes from the sample of Ho (2002), including the effects of changing radiative efficiency at low accretion rates in radiatively inefficient accretion flows and the dependence of bolometric corrections on both luminosity and accretion rate.

We find that most local black holes have very low accretion rates $\dot{m} \sim 10^{-4}$, well below the minimum accretion rate $\dot{m}_{\rm H} = 0.003$ required to grow significantly over the entire age of the universe. This is especially true in the case of supermassive black holes in elliptical galaxies, implying that the dominant part of the integrated black hole mass density was formed in radiatively efficient high-$\dot{m}$ phases, as is expected if these galaxies have undergone mergers in the past that would fuel quasar activity and eventually heat or expel gas in some form of self-regulated growth (Di Matteo et al. 2005; Springel et al. 2005a).

We have further discussed models for the accretion rate as a function of time or distribution of accretion rates given the observed $z = 0$ distribution and find in all cases that the accumulated mass should be dominated by mass gained in efficient, high-$\dot{m}$ phases. It is only possible to gain significant mass in low-$\dot{m}$ phases in these models with a high degree of fine-tuning of the quasar lifetime, giving values inconsistent with both theory and observations. Furthermore, scaling with redshift is not expected to change our results significantly, and it is increasingly difficult for black holes to gain any mass in low-$\dot{m}$ states if a previous, high-efficiency mass gain period is suppressed.

We do not expect that selection effects will significantly change our results. We have compared directly the sample of Ho (2002) and the combined samples of Ho (2002) and Marchesini et al. (2004) and find identical qualitative results. We have considered separately different classes of black holes measured by different techniques, namely reverberation-mapping and stellar kinematics, and find that the two methods yield similar qualitative results despite having essentially opposite biases. However, because the two methods have opposite biases, current measurements with either method only minimally overlap at bolometric luminosities $L \sim 10^9 L_\odot$, allowing the possibility of a bias against sources with accretion rates $\dot{m} \sim 10^{-3}$ to 10^{-1}. The addition of the Marchesini et al. (2004) sample spans this gap in luminosity, while preserving the deficit of objects at these accretion rates.

Based on the $z = 0$ distribution of accretion rates, the integrated mass of the black hole population must have been gained in earlier, radiatively efficient, high-$\dot{m}$ ($\dot{m} \gtrsim 0.1$) phases. Likewise, the mass growth of most individual black holes must be dominated by high-$\dot{m}$ rapid growth. This is in good agreement with a range of inferred accretion histories from observations of quasar populations (e.g., Soltan 1982; Salucci et al. 1999; Yu & Tremaine 2002; Marconi et al. 2004; Shankar et al. 2004) and direct observations of quasar accretion rates and their evolution with redshift (Vestergaard 2004; Heckman et al. 2004; McLure & Dunlop 2004; Yu et al. 2005). It similarly agrees well with theoretical models of the relation between black hole mass and host galaxy mass or velocity dispersion (Silk & Rees 1998; Fabian 1999; Ciotti & Ostriker 2001; Wyithe & Loeb 2003; Di Matteo et al. 2005) and semianalytical modeling of galaxy and quasar formation (e.g., Kauffmann & Haehnelt 2000). Recently, more detailed modeling of quasar light curves and the implied rate of formation of sources from the observed luminosity functions (Hopkins et al. 2005a, 2005b, 2005c, 2005d, 2006a, 2006b, 2006c) also predicts the black hole mass function allowing for luminosity and host galaxy-dependent quasar lifetimes in which mass growth is dominated by a short peak accretion phase of high-$\dot{m}$. Unlike all of these previous observational and theoretical constraints, however, our present result is model independent and does not involve large uncertainties in, e.g., the evolution of quasars and their fueling mechanisms.

Because the radiative efficiency of black hole accretion is expected and observed to decrease at low-$\dot{m}$, this argument is amplified when considering the radiative luminosity of accreting black holes. That is, the radiative output is even more dominated by high-$\dot{m}$ phases of accretion. This is consistent with the cosmic IR or X-ray backgrounds, synthesis models of which suggest black hole growth dominated by high-$\dot{m}$ phases (e.g., Comastri et al. 1995; Gilli et al. 1999; Ueda et al. 2003; Hopkins et al. 2006a) and relatively short time spent at $\dot{m} \sim \dot{m}_{\rm crit}$ (Cao 2005), and also with simple arguments that suggest radiative efficiencies equal to or even larger than $\epsilon_r = 0.1$ (Elvis et al. 2002; Merloni 2004). Thus, models in which quasars spend a very large fraction of their time at low luminosities or low accretion rates (e.g., Hopkins et al. 2005a, 2005b, 2005c, 2005d, 2006a, 2006b, 2006c) are completely consistent with both the observed $z = 0$ distribution of accretion rates and the above observations, which all suggest that mass growth and radiative output are dominated by short, high-$\dot{m}$ phases.

Finally, we have not ruled out the possibility that some small fraction of *individual* objects may gain significant mass in low-$\dot{m}$ states. Based on the observed $\dot{m}$ distribution, this fraction is small, $\lesssim 5\%$, and we have shown that it does not contribute significantly to the integrated black hole mass density. However, it could have important implications for the energetics of these objects and their relations to their host galaxies, as feedback may couple differently to the surrounding medium (in, for example, the form of jets instead of integrated disk radiation, given the low accretion rate state; e.g., Narayan & Yi 1995a). Likewise, it is not clear whether the physical mechanisms responsible for the M-σ relation in most systems would apply (or whether alternative mechanisms might have the same effect) in low-$\dot{m}$ systems (although see, e.g., Churazov et al. 2005).

To summarize, the observed $z = 0$ distribution of black hole accretion rates constrains the total time black holes spend in different states and the relative contribution of these states to the integrated final black hole mass and luminosity output.

1. Supermassive black holes spend most *time* in a low-efficiency, low accretion rate state. The $z = 0$ Eddington ratio distribution implies that the bright quasar lifetime $t_Q \lesssim t_{\rm S}$. This means that $\sim 1\%$ of a supermassive black hole's lifetime is spent in a bright, high-$\dot{m}$ quasar phase and $\sim 99\%$ is spent in a dim low-$\dot{m}$ phase.

2. Nevertheless, the growth of black hole *mass* is dominated by the short, high-$\dot{m}$ phase of evolution. The Eddington ratio distribution implies an effective mass-weighted $\langle \dot{m} \rangle \sim 0.3$, in good agreement with estimates from a wide range of other observations of quasars.

3. Likewise, the integrated *luminosity* radiated is dominated by the high-$\dot{m}$, high radiative efficiency growth phase. The effective luminosity-weighted radiative efficiency implied by the observations, coupled with simple models of the quasar evolution, is $\langle \epsilon_r \rangle \gtrsim 0.8\epsilon_r^0 - 0.9\epsilon_r^0$, where $\epsilon_r^0 \sim 0.1$ is the radiative efficiency at $\dot{m} = 1$. Again, this is consistent with a range of constraints from observations.

We thus confirm current wisdom regarding points (1) and (3), and we obtain a robust result in the case of (2).

REFERENCES

Abramowicz, M. A. 2005, in Growing Black Holes: Accretion in a Cosmo-
    logical Context, ed. A. Merloni, S. Nayakshin, & R. A. Sunyaev (Berlin:
    Springer), 257
Aller, M. C., & Richstone, D. 2002, AJ, 124, 3035
Barger, A. J., et al. 2005, AJ, 129, 578
Barnes, J. E., & Hernquist, L. 1991, ApJ, 370, L65
———. 1992, ARA&A, 30, 705
———. 1996, ApJ, 471, 115
Barret, D., & Olive, J. 2002, ApJ, 576, 391
Blandford, R. D., & Begelman, M. C. 1999, MNRAS, 303, L1
Blandford, R. D., & McKee, C. F. 1982, ApJ, 255, 419
Bower, G. C., Wright, M. C. H., Falcke, H., & Backer, D. C. 2003, ApJ, 588,
    331
Cao, X. 2005, ApJ, 631, L101
Cao, X., & Xu, Y.-D. 2005, ApJ, submitted (astro-ph/0504662)
Chen, X., Abramowicz, M. A., Lasota, J., Narayan, R., & Yi, I. 1995, ApJ, 443,
    L61
Churazov, E., Sazonov, S., Sunyaev, R., Forman, W., Jones, C., & Böhringer, H.
    2005, MNRAS, 363, L91
Ciotti, L., & Ostriker, J. P. 2001, ApJ, 551, 131
Comastri, A., Setti, G., Zamorani, G., & Hasinger, G. 1995, A&A, 296, 1
Cox, T. J., Di Matteo, T., Hernquist, L., Hopkins, P. F., Robertson, B., &
    Springel, V. 2006, ApJ, in press (astro-ph/0504156)
de Vaucouleurs, G., de Vaucouleurs, A., Corwin, H. G., Jr., Buta, R. J., Paturel, G.,
    & Fouqué, R. 1991, Third Reference Catalogue of Bright Galaxies (New York:
    Springer)
Di Matteo, T., Carilli, C. L., & Fabian, A. C. 2001, ApJ, 547, 731
Di Matteo, T., Croft, R. A. C., Springel, V., & Hernquist, L. 2003, ApJ, 593, 56
———. 2004, ApJ, 610, 80
Di Matteo, T., Quataert, E., Allen, S. W., Narayan, R., & Fabian, A. C. 2000,
    MNRAS, 311, 507
Di Matteo, T., Springel, V., & Hernquist, L. 2005, Nature, 433, 604
Eckart, A., Genzel, R., Ott, T., & Schödel, R. 2002, MNRAS, 331, 917
Elvis, M., Risaliti, G., & Zamorani, G. 2002, ApJ, 565, L75
Elvis, M., et al. 1994, ApJS, 95, 1
Esin, A. A., McClintock, J. E., & Narayan, R. 1997, ApJ, 489, 865
Fabian, A. C. 1999, MNRAS, 308, L39
Fabian, A. C., & Canizares, C. R. 1988, Nature, 333, 829
Ferrarese, L., & Merritt, D. 2000, ApJ, 539, L9
Gebhardt, K., et al. 2000, ApJ, 539, L13
George, I. M., et al. 1998, ApJS, 114, 73
Ghez, A. M., Salim, S., Hornstein, S. D., Tanner, A., Lu, J. R., Morris, M.,
    Becklin, E. E., & Duchêne, G. 2005, ApJ, 620, 744
Gilli, R., Risaliti, G., & Salvati, M. 1999, A&A, 347, 424
Granato, G. L., De Zotti, G., Silva, L., Bressan, A., & Danese, L. 2004, ApJ,
    600, 580
Green, P. J., et al. 1995, ApJ, 450, 51
Greenhill, L. J., Gwinn, C. R., Antonucci, R., & Barvainis, R. 1996, ApJ, 472,
    L21
Greenhill, L. J., Moran, J. M., & Herrnstein, J. R. 1997, ApJ, 481, L23
Greenhill, L. J., et al. 2001, in IAU Symp. 205, Galaxies and Their Constituents
    at the Highest Angular Resolutions, ed. R. Schilizzi et al. (San Francisco:
    ASP), 334
Haiman, Z., & Menou, K. 2000, ApJ, 531, 42
Hao, L., et al. 2005, AJ, 129, 1795
Heckman, T. M., Kauffmann, G., Brinchmann, J., Charlot, S., Tremonti, C., &
    White, S. D. M. 2004, ApJ, 613, 109
Hernquist, L. 1990, ApJ, 356, 359
Hernquist, L., Spergel, D. N., & Heyl, J. S. 1993, ApJ, 416, 415
Ho, L. C. 1999a, in Observational Evidence for Black Holes in the Universe,
    ed. S. K. Chakrabarti (Dordrecht: Kluwer), 157
———. 1999b, ApJ, 516, 672
———. 2002, ApJ, 564, 120
Ho, L. C., & Peng, C. Y. 2001, ApJ, 555, 650
Ho, L. C., et al. 2000, ApJ, 541, 120
Hopkins, P. F., Hernquist, L., Cox, T. J., Robertson, B., Di Matteo, T., Martini, P.,
    & Springel, V. 2005a, ApJ, 630, 705
Hopkins, P. F., Hernquist, L., Cox, T. J., Robertson, B., Di Matteo, T., &
    Springel, V. 2005b, ApJ, 630, 716
———. 2005c, ApJ, 632, 81
———. 2006a, ApJS, 163, 1
———. 2006b, ApJS, 163, 50
———. 2006c, ApJ, 632, 700
Hopkins, P. F., Hernquist, L., Martini, P., Cox, T. J., Robertson, B., Di Matteo, T.,
    & Springel, V. 2005d, ApJ, 625, L71
Jester, S. 2005, ApJ, 625, 667

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
Kollmeier, J. A., et al. 2005, ApJ, submitted (astro-ph/0508657)
Kormendy, J., & Gebhardt, K. 2001, in AIP Conf. Proc. 586, 20th Texas
    Symposium on Relativistic Astrophysics, ed. J. C. Wheeler & H. Martel
    (Melville: AIP), 363
Kormendy, J., & Richstone, D. O. 1995, ARA&A, 33, 581
Krolik, J. H. 2001, ApJ, 551, 72
Loewenstein, M., Mushotzky, R. F., Angelini, L., Arnaud, K. A., & Quataert, E.
    2001, ApJ, 555, L21
Maccarone, T. J. 2003, A&A, 409, 697
Maccarone, T. J., & Coppi, P. S. 2003, MNRAS, 338, 189
Maccarone, T. J., Gallo, E., & Fender, R. 2003, MNRAS, 345, L19
Magorrian, J., et al. 1998, AJ, 115, 2285
Mahadevan, R. 1997, ApJ, 477, 585
Marchesini, D., Celotti, A., & Ferrarese, L. 2004, MNRAS, 351, 733
Marconi, A., & Hunt, L. K. 2003, ApJ, 589, L21
Marconi, A., Risaliti, G., Gilli, R., Hunt, L. K., Maiolino, R., & Salvati, M.
    2004, MNRAS, 351, 169
Martini, P. 2004, in Coevolution of Black Holes and Galaxies, ed. L. C. Ho
    (Cambridge: Cambridge Univ. Press), 169
McClintock, J. E., & Remillard, R. A. 2006, in Compact Stellar X-Ray Sources,
    ed. W. H. G. Lewin & M. van der Klis (Cambridge: Cambridge Univ. Press),
    in press (astro-ph/0306213)
McLure, R. J., & Dunlop, J. S. 2001, MNRAS, 327, 199
———. 2002, MNRAS, 331, 795
———. 2004, MNRAS, 352, 1390
Meier, D. L. 2001, ApJ, 548, L9
Merloni, A. 2004, MNRAS, 353, 1035
Merloni, A., Rudnick, G., & Di Matteo, T. 2004, MNRAS, 354, L37
Meyer, F., Liu, B. F., & Meyer-Hofmeister, E. 2000, A&A, 354, L67
Mihos, J. C., & Hernquist, L. 1994, ApJ, 437, 611
———. 1996, ApJ, 464, 641
Miyoshi, M., Moran, J., Herrnstein, J., Greenhill, L., Nakai, N., Diamond, P., &
    Inoue, M. 1995, Nature, 373, 127
Narayan, R. 1996, ApJ, 462, 136
———. 2006, in From X-Ray Binaries to Quasars: Black Hole Accretion on
    All Mass Scales, ed. T. Maccarone, R. Fender, & L. Ho (Dordrecht: Kluwer),
    in press (astro-ph/0411385)
Narayan, R., & Yi, I. 1994, ApJ, 428, L13
———. 1995a, ApJ, 444, 231
———. 1995b, ApJ, 452, 710
Narayan, R., Yi, I., & Mahadevan, R. 1995, Nature, 374, 623
———. 1996, A&AS, 120, 287
Newman, J. A., Eracleous, M., Filippenko, A. V., & Halpern, J. P. 1997, ApJ,
    485, 570
Pellegrini, S. 2005, ApJ, 624, 155
Perna, R., Narayan, R., Rybicki, G., Stella, L., & Treves, A. 2003, ApJ, 594,
    936
Perola, G. C., et al. 2002, A&A, 389, 802
Quataert, E. 2001, in ASP Conf. Ser. 224, Probing the Physics of Active
    Galactic Nuclei, ed. B. M. Peterson, R. W. Pogge, & R. S. Polidan (San
    Francsico: ASP), 71
Richstone, D., et al. 1998, Nature, 395, A14
Robertson, B., Cox, T. J., Hernquist, L., Franx, M., Hopkins, P. F., Martini, P.,
    & Springel, V. 2006, ApJ, 641, 21
Salpeter, E. E. 1964, ApJ, 140, 796
Salucci, P., Szuszkiewicz, E., Monaco, P., & Danese, L. 1999, MNRAS, 307,
    637
Sanders, D. B., et al. 1988, ApJ, 325, 74
Schödel, R., et al. 2002, Nature, 419, 694
Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 24, 337
Shankar, F., Salucci, P., Granato, G. L., De Zotti, G., & Danese, L. 2004,
    MNRAS, 354, 1020
Shuder, J. M. 1981, ApJ, 244, 12
Silk, J., & Rees, M. J. 1998, A&A, 331, L1
Soltan, A. 1982, MNRAS, 200, 115
Spinrad, H., Marr, J., Aguilar, L., & Djorgovski, S. 1985, PASP, 97, 932
Springel, V., Di Matteo, T., & Hernquist, L. 2005a, ApJ, 620, L79
———. 2005b, MNRAS, 361, 776
Springel, V., & Hernquist, L. 2002, MNRAS, 333, 649
———. 2003, MNRAS, 339, 289
Strateva, I., Brandt, N., Schneider, D. P., Vanden Berk, D. G., & Vignali, C.
    2005, AJ, 130, 387
Telfer, R. C., Zheng, W., Kriss, G. A., & Davidsen, A. F. 2002, ApJ, 565, 773

Toomre, A. 1977, in Evolution of Galaxies and Stellar Populations, ed. B. M. Tinsley & R. B. Larson (New Haven: Yale Univ. Obs), 401

Toomre, A., & Toomre, J. 1972, ApJ, 178, 623

Tremaine, S., et al. 2002, ApJ, 574, 740

Ueda, Y., Akiyama, M., Ohta, K., & Miyaji, T. 2003, ApJ, 598, 886

Vanden Berk, D. E., et al. 2001, AJ, 122, 549

———. 2006, AJ, 131, 84

van der Marel, R. P. 1999, in IAU Symp. 186, Galaxy Interactions at Low and High Redshift, ed. J. E. Barnes & D. B. Sanders (Dordrecht: Kluwer), 333

Vestergaard, M. 2004, ApJ, 601, 676

Vignali, C., Brandt, W. N., & Schneider, D. P. 2003, AJ, 125, 433

Wandel, A., Peterson, B. M., & Malkan, M. A. 1999, ApJ, 526, 579

Wilkes, B. J., Tananbaum, H., Worrall, D. M., Avni, Y., Oey, M. S., & Flanagan, J. 1994, ApJS, 92, 53

Wyithe, J. S. B., & Loeb, A. 2003, ApJ, 595, 614

Yee, H. K. C. 1980, ApJ, 241, 894

Yu, Q., Lu, Y., & Kauffmann, G. 2005, ApJ, 634, 901

Yu, Q., & Tremaine, S. 2002, MNRAS, 335, 965

Yuan, F., & Narayan, R. 2004, ApJ, 612, 724