The Astrophysical Journal, 639:700–709, 2006 March 10

THE EVOLUTION IN THE FAINT-END SLOPE OF THE QUASAR LUMINOSITY FUNCTION

Philip F. Hopkins,[1] Lars Hernquist,[1] Thomas J. Cox,[1] Brant Robertson,[1]
Tiziana Di Matteo,[2] and Volker Springel[3]
Received 2005 August 5; accepted 2005 October 31

ABSTRACT

Using a new model for quasar lifetimes and light curves derived from numerical simulations of galaxy mergers that incorporate black hole growth, we study the faint-end slope of the quasar luminosity function (QLF) and its evolution with redshift. Our model motivates a new interpretation of the QLF in which the bright end consists of quasars radiating near their peak luminosities but the faint end is mostly made up of brighter peak luminosity quasars seen in less luminous phases of evolution. The faint-end slope of the QLF is then set by the behavior of the lifetime (light curve) of quasars with peak luminosities near the observed break when they are in less luminous stages of evolution. We determine the faint-end slope of the QLF from the quasar lifetime, based on a set of simulations that encompass a wide range of host galaxy, merger, black hole, and interstellar gas properties. Brighter peak luminosity (higher black hole mass) systems undergo more violent evolution, and gas is expelled and heated more rapidly in the final stages of quasar evolution, resulting in a flatter faint-end slope (as these objects fall below the observed break in the QLF more rapidly). Therefore, as the QLF break luminosity moves to higher luminosities with increasing redshift, implying a larger typical quasar peak luminosity, the faint-end QLF slope flattens. From our model, we predict the evolution of the faint-end slope of the QLF and find good agreement with observations. Although black holes grow in an antihierarchical manner (with lower mass black holes formed primarily at lower redshifts), in our picture the observed change in slope and differential or "luminosity-dependent density evolution" in the QLF is determined by the nontrivial, luminosity-dependent quasar lifetime and physics of quasar feedback, and not by changes in the shape of the underlying peak luminosity or active black hole mass distributions.

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

Online material: color figures

1. INTRODUCTION

The shape and evolution of the quasar luminosity function (QLF) is fundamental to cosmology, constraining theories of galaxy and supermassive black hole (BH) formation, accretion models, the X-ray, UV, and infrared backgrounds, the role of galaxy mergers and interactions in galaxy formation and evolution, and reionization. Until recently, it has not been possible to reliably measure the faint-end slope of the QLF even at low redshifts, but this has begun to change with the advent of large, uniformly selected quasar samples in surveys such as the Sloan Digital Sky Survey (SDSS) and the Two Degree Field (2dF). Furthermore, a growing number of observations at various redshifts, in radio, optical, and soft and hard X-ray bands, indicate that the faint-end slope evolves, becoming flatter at higher redshift (e.g., Page et al. 1997; Miyaji et al. 2000, 2001; La Franca et al. 2002; Cowie et al. 2003; Ueda et al. 2003; Fiore et al. 2003; Hunt et al. 2004; Cirasuolo et al. 2005; Hasinger et al. 2005).

This evolution, parameterized either as direct evolution in the faint-end slope or as luminosity-dependent density evolution (LDDE), has been the subject of much speculation, as it implies that the density of lower luminosity quasars peaks at lower redshift. In traditional models of the quasar lifetime and light curve, this evolution is directly related to the properties of quasar hosts, implying, e.g., significant and rapid evolution in the *shape* of the distribution of host galaxy masses, which cannot be accounted for in either semianalytical models or numerical simulations and is not consistent with a wide range of galaxy observations. Models that adopt these idealized prescriptions for quasar evolution have had some success predicting the evolution of the *bright* end of the QLF, generally succeeding at high redshift but failing to account for the decrease in counts of bright quasars at low redshift (e.g., Kauffmann & Haehnelt 2000; Wyithe & Loeb 2003; Enoki et al. 2003), without invoking either feedback mechanisms to suppress growth of high-mass spheroids (e.g., Scannapieco & Oh 2004) or evolution in the BH accretion efficiency with redshift (e.g., Haiman & Menou 2000) or characteristic gas density (e.g., Cattaneo et al. 2005).

In either case, such models do not have a natural explanation for the faint-end slope or its evolution and thus cannot be extrapolated to low luminosities or to redshifts where the slope is undetermined. Observations at high redshifts are uncertain, and no large, uniformly selected samples yet exist that measure the faint-end slope at both low ($z \lesssim 1$) and high ($z \gtrsim 3$) redshifts. The high-z, faint-end slope is important in determining the early formation history of BHs, and especially their contribution to reionization, as well as possible connections between quasars and, e.g., the low-luminosity Seyfert galaxies seen at $z \sim 0$.

Without a physical model of quasar evolution, attempts to reconcile observations of evolution in the faint-end QLF slope and BH populations (e.g., Merloni 2004) have generally fitted the QLF and BH mass population to somewhat arbitrary distributions of lifetimes/duty cycles and accretion rates as a function of redshift. Even so, this phenomenological approach has elucidated the antihierarchical nature of BH growth, with smaller mass BHs formed at lower redshift as an implication of this evolution in the QLF. Similarly, Cattaneo (2001) fitted these distributions with different forms for the quasar light curve, suggesting that

[1] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138.
[2] Department of Physics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213.
[3] Max-Planck-Institut für Astrophysik, Karl-Schwarzchild-Strasse 1, 85740 Garching bei München, Germany.

the quasar lifetime must increase at lower luminosities (following, e.g., a power-law decay in the quasar light curve) to reproduce the faint-end QLF if the quasar host galaxy mass function decreases at low masses, as suggested by observations (e.g., Bahcall et al. 1997; Hamilton et al. 2002). With this argument, he concluded that faint quasars are mostly fading massive BHs accreting at low rates, rather than low-mass BHs accreting close to the Eddington limit. But an actual prediction of the faint-end slope requires a more detailed model for both quasar lifetimes and the QLF.

Recently, BH growth and feedback have been incorporated into numerical simulations of galaxy mergers (Springel et al. 2005a, 2005b). In these simulations, gravitational torques drive inflows of gas into the nuclei of merging galaxies (e.g., Barnes & Hernquist 1991, 1996), triggering starbursts (e.g., Mihos & Hernquist 1996) and feeding the growth of central supermassive BHs (Di Matteo et al. 2005). As the BHs accrete, some of the radiated energy couples to the surrounding gas, and the growth eventually stalls when this feedback energy is sufficient to unbind the reservoir of gas. These results connect galaxy evolution, the formation of supermassive BHs, and the self-regulated nature of quasar activity, and they provide quantitative predictions that agree well with observations of, e.g., the $M_{\rm BH}$-σ relation (Di Matteo et al. 2005; Robertson et al. 2006), quasar lifetimes (Hopkins et al. 2005a, 2005d), and the QLF in various wave bands (Hopkins et al. 2005b, 2005c, 2006b). These simulations provide a self-consistent description of the light curve of quasar activity and the implied quasar "lifetime"—i.e., the duration of time that a quasar spends at a given luminosity—and its dependence on the properties of the merging galaxies.

In the simulations, the quasar lifetime depends on both the instantaneous *and* peak luminosities of a given quasar, such that the lifetime is longer at lower luminosities; i.e., a given quasar spends more time (and is more likely to be observed) at a luminosity well below its peak luminosity. This differs from previous models, which generally assume that quasars radiate at a fixed luminosity for some characteristic lifetime or adopt idealized exponential light curves, but it is suggested by both recent observations of quasar clustering (Adelberger & Steidel 2005; Lidz et al. 2006) and comparison of semianalytical (e.g., Granato et al. 2004) and empirical (e.g., Cattaneo & Bernardi 2003) models to the observed QLF.

The complex character of quasar lifetimes in our picture motivates a new interpretation of the QLF in which the bright end consists of quasars radiating near their peak luminosities, while the faint end consists mainly of quasars evolving toward or declining from peak activity. The "break" in the QLF corresponds directly to the maximum in the intrinsic distribution of *peak* luminosities, which falls off at both brighter and fainter luminosities. The faint-end slope of the QLF is then determined by the faint-end slope of the luminosity-dependent quasar lifetime (i.e., the differential time the quasar spends in any given luminosity interval) for quasars with peak luminosity near the observed break. In other words, the number of quasars in near-peak phases of evolution needed to account for the observed bright end of the QLF is already sufficient to mostly account for the observed faint-end population, as any given quasar is much more likely to be observed at luminosities well below its peak, and thus the probability of seeing such brighter sources at these lower luminosities determines the shape of the faint-end QLF.

This model of the quasar lifetime and its implications for the QLF thus provide a physical interpretation for the location of the break luminosity and the faint-end slope of the QLF. Furthermore, Hopkins et al. (2006a, 2006b) showed that this description reproduces well many quasar and galaxy observables that are difficult to account for in more idealized modeling, including differences in the QLF in different bands and redshifts, Eddington ratio and column density distributions, the X-ray background spectrum, and relic red/elliptical galaxy population colors and distributions.

In this paper, we use our model of quasar lifetimes and light curves to study the faint-end QLF slope and its redshift evolution. We are able to reproduce the observed evolution in the slope and corresponding LDDE based on our simulations, as a unique consequence of our interpretation of the QLF and without the need to invoke evolution in the shape of, e.g., the quasar host or galaxy mass distribution. We further develop a simple analytical model of quasar feedback that gives similar predictions and demonstrates the scale-dependent physics that drive the evolution of the faint-end slope.

Throughout, we define the bolometric luminosity $L = L_{\rm bol}$. We denote the faint-end slope γ by $\phi = d\Phi/d\log(L) \propto L^{-\gamma}$, where ϕ is the differential *bolometric* QLF, and we are considering luminosities $L \ll L_*$, where L_* is the break in the QLF. Note that $d\Phi/dL \propto L^{-\gamma-1}$, as, e.g., in the standard "double power law" form of the QLF. The sign choice (i.e., $+\gamma$) follows conventional usage. We adopt a $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, $H_0 = 70$ km s^{-1} Mpc^{-1} cosmology.

2. THE FAINT-END SLOPE AS A FUNCTION OF PEAK LUMINOSITY

If quasars spend a differential time $dt_Q/d\log(L)$ per logarithmic interval in luminosity, then the observed QLF is (at times when the quasar lifetime is short compared to the Hubble time)

$$\phi(L) = \int \frac{dt_Q}{d\log(L)}\,\dot{n}(L_{\rm peak})\,d\log L_{\rm peak}, \qquad (1)$$

where $\dot{n}(L_{\rm peak})$ is the birthrate of quasars of a given peak luminosity $L_{\rm peak}$ (per unit time per unit comoving volume per logarithmic interval in $L_{\rm peak}$). In Hopkins et al. (2005b, 2006b), we used this and our model of quasar lifetimes to determine $\dot{n}(L_{\rm peak})$ and found that it does not have the same shape as the observed QLF, as is expected for idealized models in which a quasar turns "on/off" as a step function or follows a pure exponential light curve. Instead, $\dot{n}(L_{\rm peak})$ traces the shape of the observed QLF at the bright end (above L_*), peaks at $L_{\rm peak} \sim L_*$, then falls off below this. Therefore, the slope of the faint-end QLF, γ, is dominated by the faint-end slope of $dt_Q/d\log(L)$ for quasars with $L_{\rm peak} \sim L_*$; i.e., γ is determined by the integrated probability of seeing the population of brighter $L_{\rm peak}$ sources, dominated by sources with $L_{\rm peak} \sim L_*$, at lower L.

Given this interpretation of the QLF, γ should be calculable from the $L \ll L_{\rm peak}$ slope of $dt_Q/d\log(L)$ (since the probability of seeing such a source at L is proportional to its lifetime at L) for sources with $L_{\rm peak} \approx L_*$. If the quasar lifetime is a function of L and $L_{\rm peak}$ only (i.e., not affected systematically by other host galaxy properties), then γ can be predicted by knowing $L_*(z)$, which directly gives the peak in the $\dot{n}(L_{\rm peak})$ distribution as a function of redshift. There may be some curvature introduced because of the nonzero contributions to the faint end of the QLF from sources with $L_{\rm peak} \neq L_*$, but as such corrections depend on the exact shape of $\dot{n}(L_{\rm peak})$ and are small, as $\dot{n}(L_{\rm peak})$ drops off rapidly (as the bright-end QLF does) away from $L = L_*$, we ignore them here.

2.1. *The Quasar Lifetime from Simulations*

We first consider the $L \ll L_{\rm peak}$ behavior of the quasar lifetime determined from hydrodynamic simulations. We employ

several hundred simulations, described in detail in Robertson et al. (2006) and Hopkins et al. (2006b), performed with the new parallel TreeSPH code GADGET-2 (Springel 2005), which uses an entropy-conserving formulation of smoothed particle hydrodynamics (SPH; Springel & Hernquist 2002) and includes a subresolution, multiphase model of the dense interstellar medium (ISM) to describe star formation (Springel & Hernquist 2003). The multiphase gas is pressurized by feedback from supernovae, allowing us to stably evolve even pure gas disks (see, e.g., Springel et al. 2005b; Robertson et al. 2004; Springel & Hernquist 2005). We generate two stable, isolated disk galaxies, each with an extended dark matter halo having a Hernquist (1990) profile, an exponential disk, and a bulge. The galaxies are then set to collide from a zero energy orbit.

BHs are represented by "sink" particles that accrete gas, with an accretion rate $\dot{M}$ estimated using a Bondi-Hoyle-Lyttleton parameterization $\dot{M} = 4\pi\alpha(G M_{\rm BH})^2 \rho (v_{\rm BH}^2 + c_s^2)^{-3/2}$ with an upper limit equal to the Eddington rate, where the local gas density ρ, sound speed c_s, and bulk velocity $v_{\rm BH}$ are determined from the local gas properties with the SPH formalism (Springel et al. 2005b). We deliberately set α to a high (constant) value of $\alpha = 100$ in order to roughly translate the coarse-averaged mean density of the ISM obtained in our subresolution model to the expected higher densities close in to the BH. Otherwise, the initial Bondi growth phase of the small seed BH mass we adopted ($M_{\rm seed} = 10^5\ M_\odot$) would be artificially long, while with our choice the BH can grow to sizes where feedback can become important within ~200–300 Myr. For a different choice of α, a timing offset in the quasar light curve can occur, but our results are insensitive to this and essentially degenerate in the product $\alpha M_{\rm seed}$, provided the seed mass is much smaller than the final BH mass (see, e.g., Fig. 6 of Hopkins et al. [2006b], which shows that the simulation quasar lifetimes are independent of seed mass and therefore, equivalently, α). The bolometric luminosity of the BH is then $L = \epsilon_r \dot{M} c^2$, where $\epsilon = 0.1$ is the radiative efficiency. We allow a small fraction (typically ≈5%) of L to couple dynamically to the gas as thermal energy. This fraction is a free parameter, determined in Di Matteo et al. (2005) by fitting the $M_{\rm BH}$-σ relation. For a detailed description of the accretion prescription, see Springel et al. (2005b), and for a discussion of the effects of changing different feedback mechanisms, see Hopkins et al. (2006b). For our purposes, changing the normalization of $\dot{M}$ or ϵ_r will only alter the relation between final BH mass (peak luminosity) and galaxy mass, not the scaling (e.g., power-law slope) of the quasar lifetime relevant for our predictions. As discussed in § 2.2, any feedback mechanism that involves the injection of large amounts of energy in the central regions of the merger in a short time should result in a similar scaling.

In our simulations, we vary the masses and virial velocities of the initial galaxies, halo concentrations, ISM equation of state, parameters describing feedback from supernovae and BH growth, presence or absence of bulges in the host galaxies, initial BH seed masses, numerical resolution (typically $\sim 2 \times 10^5$ particles per galaxy, but we consider up to 128 times as many), disk inclinations and pericenter separation of the initial orbits, and initial disk gas fractions. We further scale all galaxy properties appropriately to resemble galaxies at redshifts $z = 0$–6 for a large subset of our simulations, as described in Robertson et al. (2006). Our simulations produce quasars with $L,\ L_{\rm peak}$ from $\sim 10^8$ to $10^{15}\ L_\odot$, spanning the entire observed range at all redshifts. Hopkins et al. (2006b) used the simulations to determine the quasar lifetime $dt_Q/d\log(L)$ and found that $dt_Q/d\log(L)$ shows no systematic dependence and little scatter when the parameters are varied, when quantified as a function of L and $L_{\rm peak}$. The predictions in Hopkins et al. (2006b) did not depend sensitively on the (much more uncertain) faint-end slope, and therefore the authors did not consider a detailed fit to the faint-end behavior of $dt_Q/d\log(L)$, but rather parameterized the lifetime as an exponential, $dt_Q/d\log(L) = t_0\ \exp(-L/L_0)$, where $L_0 \approx 0.2 L_{\rm peak}$ and t_0 depends weakly on $L_{\rm peak}$, which provides an acceptable fit to the simulation results for $L \gtrsim 10^{-2}$ to $10^{-1} L_{\rm peak}$. Hopkins et al. (2005a) considered the $L \ll L_{\rm peak}$ behavior of $dt_Q/d\log(L)$ in more detail and found that the lifetime more closely resembles a power law at these L, with a power-law slope α as a function of $L_{\rm peak}$. The combination of these results suggests that the quasar lifetime is best parameterized as a Schechter function with slope α, normalization $t_0(L_{\rm peak})$, and turnover $L_0(L_{\rm peak})$, for purposes where the faint-end slope is important; i.e. (to clarify our conventions),

$$\frac{dt_Q}{d\log(L)} = t_0 \left(\frac{L}{L_0}\right)^{-\alpha} \exp\left(-\frac{L}{L_0}\right). \quad (2)$$

For a QLF with break $L_*(z)$, which implies an $\dot{n}(L_{\rm peak})$ distribution peaked at $L_{\rm peak} \sim L_*$, the observed faint-end QLF slope is then $\gamma \approx \alpha(L_{\rm peak} = L_*)$.

For each simulation, we fitted a Schechter function to the quasar lifetime as a function of luminosity and quantify α as a function of $L_{\rm peak}$. These faint-end slopes are subject to several uncertainties in our modeling. One is the finite time duration of our simulations, which may flatten $dt_Q/d\log(L)$ at low L, as the BH cannot completely relax in the time interval simulated. A complementary means to determine γ is then to consider the rate at which L falls off after $L = L_{\rm peak}$ at time $t = t_p = t(L_{\rm peak})$. We take $L(t)/L_{\rm peak}$ to be a function of $t - t_p$ and find that it falls off in approximate power-law fashion for all our simulations,

$$L(t) \propto (t - t_p)^{-\beta}. \quad (3)$$

This implies

$$\frac{dt_Q}{d\log(L)} \propto L^{-1/\beta}; \quad (4)$$

i.e., $\gamma = \alpha = 1/\beta$. Although these fits are not affected by the finite time duration of the simulations, they do not include the time spent at different L before $L = L_{\rm peak}$ and thus are entirely accurate only at $t \gg t_p$ or for a symmetric rise/fall in L and may therefore overestimate the steepness of γ, so we consider both fitting methods below.

The uncertainties in this modeling at low luminosities are discussed in greater detail by Hopkins et al. (2005b, 2006b). Hopkins et al. (2005a) examined fits to the power-law behavior of $dt_Q/d\log(L)$ at $L \ll L_{\rm peak}$, considering both the entire simulation and only times after the merger [similar to our fits for the decay of $L(t)$]. Furthermore, Hopkins et al. (2005b) considered the application of an advection-dominated accretion flow (ADAF)-type correction for radiatively inefficient accretion flows at low accretion rates (following, e.g., Narayan & Yi 1995) to account for more detailed variation in the radiative efficiency and spectrum as a function of accretion rate. We apply this correction in our fitting as well. Although these previous works considered a much smaller subset of the simulations that we fitted, they found similar results, and we compare their fits and ours below to demonstrate both the general agreement and range of uncertainty introduced by these different determinations of γ.

Figure 1 shows the results of this fitting for several representative simulations. We show the quasar light curve after peak

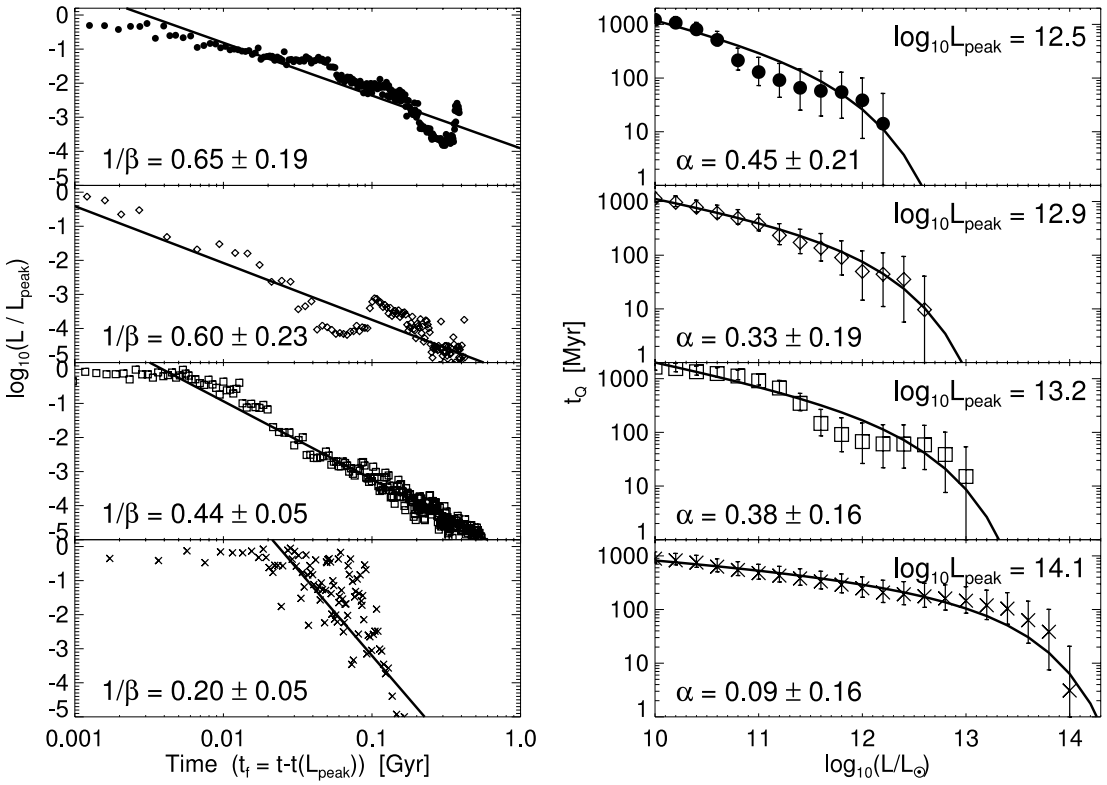

FIG. 1.—*Left*: Bolometric luminosity over peak luminosity ($L/L_{\rm peak}$) as a function of time [$t_f = t - t(L_{\rm peak})$], with the best-fit power law of slope β, for representative mergers with peak luminosities $L_{\rm peak}/L_\odot = 3.1 \times 10^{12}$ (*circles*), 7.3×10^{12} (*diamonds*), 1.7×10^{13} (*squares*), and 1.4×10^{14} (*crosses*). *Right*: Quasar lifetime (integrated time above a given luminosity) for the same simulations, with the best-fit Schechter functions (*lines*) of slope α. The predicted faint-end QLF slopes $\gamma = \alpha = 1/\beta$ are also shown. [*See the electronic edition of the Journal for a color version of this figure.*]

luminosity (*left*), $L(t)/L_{\rm peak}$ as a function of $t - t_p$, for simulations with $L_{\rm peak}/L_\odot = 3.1 \times 10^{12}$ (*circles*), 7.3×10^{12} (*diamonds*), 1.7×10^{13} (*squares*), and 1.4×10^{14} (*crosses*), with the best-fit power-law slope β shown (*solid line*). The implied faint-end QLF slope $1/\beta$ is shown for each, and we compare to the slope α determined from fitting to the quasar lifetime as a function of L [*right*, same notation; for clarity, the integrated lifetime above a given L, $t_Q(L' > L)$, is shown] to the best-fit Schechter function (*solid lines*) with slope α. From the resulting faint-end slope γ for each $L_{\rm peak}$ and each fitting method, there is a clear decrease of γ with $L_{\rm peak}$.

We fitted for $\gamma(L_{\rm peak})$ in each of all our simulations with both methods described above. We found that γ decreases with $L_{\rm peak}$ in a manner similar to that in Figure 1. Given the scatter, the resulting slopes $\gamma(L_{\rm peak})$ are well fitted by a loglinear relation,

$$\gamma(L_{\rm peak}) = \gamma_{12} + \frac{d\gamma}{d\log(L_{\rm peak})}\left[\log(L_{\rm peak}/L_\odot) - 12\right], \quad (5)$$

where $\gamma_{12} \sim 0.5$ and $d\gamma/d\log(L_{\rm peak}) \sim -0.3$. This is illustrated in Figure 2, where the value of γ determined from fitting Schechter functions to the quasar lifetime is plotted as a function of the peak luminosity of those simulations (*circles with error bars*), with the best-fit loglinear relation (*line*). The results for both fitting methods are summarized in Table 1. We also show the results of Hopkins et al. (2005a, 2005b), described previously. As discussed in Hopkins et al. (2006b), we find no systematic dependence on redshift, host galaxy properties, or other varied parameters in the simulations.

In the simulations, gas is expelled when feedback from accretion rapidly unbinds it in the gravitational potential of the remnant. Around more massive BHs (i.e., higher $L_{\rm peak}$ quasars), this "blowout" event is progressively more violent; the higher luminosity of the quasar heats the gas more rapidly and to higher temperatures, resulting in more efficient, more rapid expulsion of the gas (see also Cox et al. [2005], who showed that the gas mass fraction expelled increases with BH mass).

2.2. *A Simple Model of Quasar Feedback*

To understand this trend better, we construct an analytical description of quasar blowout. This model reproduces the late-time (post–peak activity) quasar evolution, although the early phases

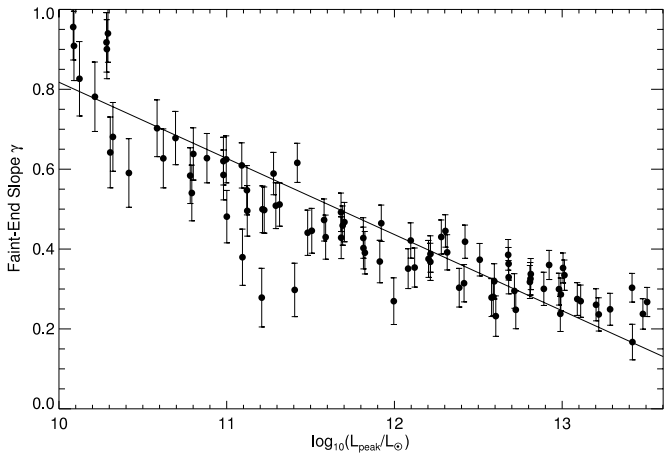

FIG. 2.—Faint-end QLF/quasar lifetime slope $\gamma = \alpha$ determined from fits to the quasar lifetime as a function of peak luminosity $L_{\rm peak}$ in our simulations. The best-fit loglinear relation (see Table 1) is shown (*solid line*).

TABLE 1
FAINT-END SLOPE $\gamma(L_{\rm peak})$

Determination[a]	$\gamma_{12} = \gamma(L_{\rm peak} = 10^{12}\ L_\odot)$	$d\gamma/d\log(L_{\rm peak})$
H05a, all times	0.95 ± 0.25	-0.32 ± 0.11
H05a, $t \gtrsim t(L_{\rm peak})$	0.76 ± 0.20	-0.30 ± 0.07
H05b, all times	0.94 ± 0.20	-0.33 ± 0.08
H05b, $t \gtrsim t(L_{\rm peak})$	0.50 ± 0.15	-0.50 ± 0.15
Schechter fitting	0.44 ± 0.02	-0.21 ± 0.02
$L(t) \propto t^{\beta}$ decay	0.69 ± 0.05	-0.30 ± 0.04
Cumulative best fit	0.55 ± 0.12	-0.25 ± 0.04
Blast wave model	0.35 ± 0.15	-0.22 ± 0.06

[a] H05a and H05b (Hopkins et al. 2005a and 2005b, respectively) use a small subset of the simulations here, with H05b applying an ADAF-type correction at low accretion rates.

(which contribute significantly to the lifetime at low-L) are determined by the complex merger process and are more difficult to describe analytically. We note that this problem has been considered in previous works (e.g., Barkana & Loeb 2001; Furlanetto & Loeb 2001; Scannapieco & Oh 2004), but these authors have considered the large-scale structure of the blowout and impact of these outflows on the outer regions of halos and clusters, the surrounding IGM, and reionization; we are instead interested in the small-scale, rapid expulsion of gas from the innermost BH-dominated regions of the galaxy. Interestingly, these other works suggest that feedback on large scales is critical in regulating the slope of the *bright* end of the QLF, which highlights the importance of quasar feedback over the entire QLF.

In our simulations (Di Matteo et al. 2005) and most analytical models of the $M_{\rm BH}$-σ relation (e.g., Silk & Rees 1998; Fabian 1999; Ciotti & Ostriker 2001; Wyithe & Loeb 2002), feedback energy from BH growth heats the surrounding gas (or momentum from coupling of the radiation field to dust drives a wind). In early merger stages, this is a small luminosity, and the gas is able to cool efficiently and reradiate this energy. However, in peak merger stages, the BH growth is exponential (Eddington limited), eventually reaching a critical threshold where the surrounding gas can no longer cool efficiently in a local dynamical time. Consequently, the gas is rapidly expelled, and what remains is heated to the virial temperature, preventing it from being easily accreted, shutting down further accretion. The threshold for this behavior is determined by the local gas properties and gravitational potential of the host, but essentially all models of feedback-driven self-regulation (owing to exceeding a critical energy input in a dynamical time) predict a relation between final BH mass and virial velocity or spheroid velocity dispersion of roughly $M_{\rm BH} \propto V_{\rm vir}^5$ or $M_{\rm BH} \propto \sigma^4$, respectively.

Because the gas in the inner regions is able to cool efficiently during early stages and is then suddenly heated and driven out in a short time (in our simulations, this occurs over a timescale $\sim 10^7$ yr in massive mergers) by an exponentially increasing luminosity, it is reasonable to model the outflow as a Sedov-Taylor–type blast wave (Sedov 1946, 1959; Taylor 1950), with energy injection from a point explosion with energy $E = \eta_L L_{\rm Edd}(M_{\rm BH}^f)\, t_{\rm dyn}$ (η_L describes the efficiency of feedback coupling, $\approx 5\%$ in our simulations and similar in most analytical models of the $M_{\rm BH}$-σ relation). A detailed examination of this problem (e.g., Furlanetto & Loeb 2001) shows that this is a reasonable approximation to a full solution including radiative cooling, the pressure of the external medium, magnetic fields, and further effects.

Modeling the accretion with a Bondi-Hoyle (Bondi & Hoyle 1944; Bondi 1952) parameterization (as we are considering times when the accretion rate falls below the Eddington limit), considered at the radius of influence of the BH, $R_{\rm BH} = GM_{\rm BH}\sigma^{-2}$ (σ being the spheroid velocity dispersion), gives $L \propto M_{\rm BH}^2 \rho(R_{\rm BH}) c_s^{-3}(R_{\rm BH})$, where c_s is the isothermal sound speed (in our simulations, the effective sound speed of the multiphase ISM is used; see Springel & Hernquist 2003). As the accretion rate is rapidly falling from its peak during this stage, we can reasonably approximate $M_{\rm BH} = M_{\rm BH}^f = $ constant, and thus $R_{\rm BH} = $ constant, and we only need to calculate the time evolution of the gas density ρ and sound speed c_s at fixed $r = R_{\rm BH}$ to determine the time evolution of luminosity L in this stage. We are only interested in the scaling of $L(t) \propto t^{-\beta}$ (where t is the time since the peak quasar luminosity or beginning of this blowout stage), so we could theoretically pick any fixed radius to evaluate these quantities, and thus the choice of the radius of influence of the BH as opposed to the transonic radius, for example, is not significant.

We first consider the simplest, scale-invariant solution to this problem, in which we neglect the gravitational field of the BH and halo, following Ostriker & McKee (1988) in our derivation. We also ignore the consequences of radiation as noted above, but this should not be a large effect in the very inner regions of the blast wave expansion in which we are interested (see also Murray et al. [2005], who considered in detail the coupling of BH radiation to dust and gas opacity and showed that it produces a similar $M_{\rm BH}$-σ relation and blowout behavior).

As we need to describe $\rho(r, t)$ at small radii (relative to the size of the galaxy and the final radius of the blast wave), we account for density gradients by adopting a one-power approximation (OPA) $\rho \propto r^{-k_\rho}$ for the ambient (preshocked) medium. In principle, there could be departures from spherical symmetry in the surrounding gas or the energy input from the quasar, but the complex gas flows and short dynamical times in the central regions of the merger should isotropize the blast wave.

For a self-similar OPA Sedov-Taylor solution, the shock radius R_s expands as $R_s \propto t^\eta$ with the postshock (internal) density profile $\rho = \rho_1 (r/R_s)^{l_\rho}$, where $\rho_1 \propto \rho(R_s)$ is the density just inside the shock front. At fixed $r = R_{\rm BH}$, we then have

$$\rho(R_{\rm BH}, t) \propto t^{-(k_\rho + l_\rho)\eta}. \tag{6}$$

For a self-similar, energy-conserving blast wave, $\eta = 2/(5 - k_\rho)$, and mass conservation (coupled with the strong shock jump conditions) requires $l_\rho = \left[6 - (\gamma_{\rm EOS} + 1)k_\rho\right]/(\gamma_{\rm EOS} - 1)$. Here, we denote the ratio of specific heats as $\gamma_{\rm EOS}$ to distinguish it from the faint-end QLF slope γ. We obtain

$$(k_\rho + l_\rho)\eta = \frac{4}{\gamma_{\rm EOS} - 1}\left(\frac{3 - k_\rho}{5 - k_\rho}\right). \tag{7}$$

The pressure $P = \rho c_s^2$ also follows a power-law scaling (with internal power-law slope l_P and external slope k_P), but with a more complicated result for l_P,

$$l_P = \frac{3\gamma_{\rm EOS}^2 + 20\gamma_{\rm EOS} + 1 - (\gamma_{\rm EOS} + 1)(3\gamma_{\rm EOS} + 1)k_\rho}{2(\gamma_{\rm EOS}^2 - 1)}. \tag{8}$$

Thus, the final scaling of $L(t) \propto \rho\, c_s^{-3} \propto \rho^{5/2} P^{-3/2}$ in this phase is given by $L \propto t^{-\beta}$, with

$$\beta = \frac{10}{\gamma_{\rm EOS} - 1}\left(\frac{3 - k_\rho}{5 - k_\rho}\right) - 3\left(\frac{k_P + l_P}{5 - k_\rho}\right). \tag{9}$$

This formalism has an exact, self-consistent solution corresponding to, e.g., a blast wave in an isothermal sphere or wind,

which should be a reasonable approximation to our simulations, with $k_\rho = 2$ and $\gamma_{\rm EOS} = 5/3$ (yielding $l_\rho = 1$, $l_P = 3$). This solution gives $\beta = 2$, implying a faint-end QLF slope $\gamma = 1/2$. But the above equations should be a good approximation (Ostriker & McKee 1988) for a range of $\gamma_{\rm EOS} \sim 1-5/3$ (corresponding to variability in the equation of state owing to the density and temperature dependence of star formation and radiative cooling) and $k_\rho \sim 1-3$ (corresponding to, e.g., the inner regions of a Hernquist [1990] spheroid profile with a central point mass). In fact, it is still exact for the equilibrium structure of a filled blast wave with the most natural $v \propto r$ velocity structure, which yields

$$k_\rho = k_{\rho,\,\rm crit} = \frac{7-\gamma_{\rm EOS}}{\gamma_{\rm EOS}+1}, \tag{10}$$

$$\beta = \frac{1}{3\gamma_{\rm EOS}-1}\left[20-\frac{9}{2}\,(+1)\right]. \tag{11}$$

For the reasonable range of equations of state $1 \le \gamma_{\rm EOS} \le 5/3$, this results in $2 \le \beta \le 11/2$; i.e., $0.2 \lesssim \gamma \lesssim 0.5$, in good agreement with the range of γ observed and determined in our simulations.

The scale-invariant Sedov-Taylor solution then accounts for the typical values of γ implied by our simulations and the rapid falloff of $L(t)$ during the blowout phase of quasar evolution. However, this does not immediately describe the dependence of γ on $L_{\rm peak}$. This dependence is not contained, as might be presumed from the above derivation, in $\gamma_{\rm EOS}$, as essentially all our simulations have $\gamma_{\rm EOS}$ in the entire range $\sim$1–5/3 in their central regions, and we specifically find no significant systematic dependence of γ or the quasar lifetime on the ISM gas equation of state (see also Hopkins et al. 2006b; their Figs. 4 and 5). Rather, the dependence of γ on $L_{\rm peak}$ is driven, naturally, by the fact that this problem is *not* precisely scale-invariant or self-similar.

To illustrate the dependence on $L_{\rm peak}$, consider a small change in the initial logarithmic slope of the density in the central regions, $k_\rho = k_\rho^0 + \delta k_\rho$. For simplicity we expand about the $k_\rho^0 = 2$, $\gamma_{\rm EOS} = 5/3$ exact solution, but our result is similar regardless of the choice of these parameters (within the above ranges). This gives

$$\beta = 2 - 13\delta k_\rho/3, \tag{12}$$

$$\gamma = \frac{1}{2}\left(1+\frac{13}{6}\,\delta k_\rho\right); \tag{13}$$

i.e., $d\gamma/dk_\rho = 13/12$. Thus, if the initial logarithmic density gradient in the inner regions is *steeper*, i.e., has a larger k_ρ ($\rho_0 \propto r^{-k_\rho}$), then β is actually *shallower*. In detail, two competing effects occur. First, the density falls off more rapidly, allowing the blast wave to propagate more rapidly as it builds up more of its mass earlier and encounters less mass farther from the central regions, but this effect is relatively weak; i.e., $\eta \rightarrow \eta + 2\delta k_\rho/9$.

The dominant effect is the alteration of the postshock density profile. Because the density gradient is steeper, the propagation of the blast wave builds up a less pronounced "snowplow"; i.e., the mass buildup at the front is less pronounced, implying a *flatter* postshock density gradient by mass conservation (recall that ρ increases with radius within the shocked region). Essentially, with the gas mass concentrated in the center, the shocked density profile evolves less dramatically, because less mass is added and the early blast is able to effectively redistribute more of the ultimately acquired mass. Given that less massive halos typically have higher concentrations, we can estimate the magnitude of this effect on the evolution of γ with $L_{\rm peak}$. Estimating $dk_\rho/dc \sim 0.1$ (where c is the concentration index and we estimate a variation $\sim$10 in c changes the inner logarithmic slope at fixed r by $\sim$1), we can determine $dc/d\log(L_{\rm peak})$, using $L_{\rm peak} \propto M_{\rm BH} \propto M_{\rm vir}$ (e.g., Marconi & Hunt 2003) and $c \approx 9(M_{\rm vir}/10^{13}\ M_\odot)^{-0.13}$ (Bullock et al. 2001), and we find

$$\frac{d\gamma}{d\log(L_{\rm peak})} \sim -0.2\left(\frac{L_{\rm peak}}{10^{12}\ L_\odot}\right)^{-0.13}, \tag{14}$$

in reasonable agreement with our measured dependence of $\gamma(L_{\rm peak})$ in our simulations and only weakly dependent on $L_{\rm peak}$. This actually predicts that the magnitude of $d\gamma/d\log L_{\rm peak}$ should be larger at low $L_{\rm peak}$ (~-0.28 at $L_{\rm peak} \sim 10^{11}\ L_\odot$) and smaller at high $L_{\rm peak}$ (~-0.16 at $L_{\rm peak} \sim 10^{13}\ L_\odot$); this may occur (see Fig. 2) but is most likely a coincidence, as our modeling of the blowout in a scale-invariant fashion and adopting a simple concentration parameter in such a chaotic period in the evolution of the merger are rough approximations at best.

Incorporating the weak, but nonnegligible, effects of the change in concentration with mass breaks the scale invariance of the standard Sedov-Taylor solution, as lower mass systems have steeper inner density profiles, which flatten the evolution of the accretion rate in time and produce steeper γ. Of course, many other effects will break the self-similarity of this problem as well—a realistic gravitational potential will imply a characteristic scale length, and the physics of radiative cooling will likewise define fundamental physical scales. Even the scale-invariant solution incorporating radiative energy loss depends on the logarithmic slope of the cooling function versus density and temperature, but the values of these slopes themselves depend on the characteristic temperature of the blast waves and change quite significantly over the mass scale of our simulations (heating to virial temperatures $c_s^2 \sim \sigma^2$ implies temperatures $T \sim 10^5$–10^7 K over the mass range shown in, e.g., Fig. 2). Although we do not model the chaotic interactions and evolving BH mass of the early merger stages, the more violent torquing associated with more massive mergers can explain the more rapid, peaked BH evolution over a larger range in BH mass, even in early merger stages, generating a flatter quasar lifetime that spans a wider range in luminosity.

Given these various scalings, it is possible that our inferred trend of $\gamma(L_{\rm peak})$ could change or even reverse at low $L_{\rm peak}$, as in models of stellar winds in dwarf elliptical galaxies for which lower mass objects (lower $M_{\rm BH}$) are more easily unbound (e.g., Mac Low & Ferrara [1999], although this is more concerned with the large-scale binding of gas, as opposed to evolution in the inner accretion regions of interest in our modeling), but the masses/luminosities where this is likely to become important ($M_{\rm gal} \lesssim 10^8\ M_\odot$, i.e., $L_{\rm peak} \lesssim 10^9\ L_\odot$) are well below the break luminosity at any redshift and thus will not affect our results. Likewise, this could occur at large radius $r \gg a$ in any $L_{\rm peak}$ system, but again we are not attempting to model the large-scale blast wave but only the evolution relevant to the inner regions.

2.3. *Parameterizations of the Quasar Light Curve*

The fits and analytical modeling above imply a simple prescription for the quasar light curve as applied in semianalytical and other approaches that cannot resolve the detailed time history of individual objects. Generally, the quasar light curve is characterized by two "modes": a "growing mode," characterized by a high-Eddington ratio, rapid BH growth, and a "decaying mode," characterized by the nearly self-similar power-law falloff of the quasar luminosity as gas is heated or expelled. The

growing mode can be parameterized by exponential growth at a constant Eddington ratio $\dot{m}$, with an exponential light curve $L = L(t = 0)\exp(t/t_Q)$, where $t_Q = t_S/\dot{m}$ is the e-folding time and $t_S = 4.2 \times 10^7$ yr is the Salpeter time. Such BH growth is expected in essentially all models of quasar activity in which a plentiful fuel supply enables high accretion rates.

Once the quasar reaches a critical luminosity or mass (e.g., in accord with the M-σ relation), it begins to heat and expel the surrounding gas, and the accretion rate rapidly falls off: i.e., the light curve can be roughly parameterized as entering the decay mode described above. Our fits to the quasar light curves after blowout and analytical modeling of this phase of evolution as a driven blast wave suggest a power-law decline in the quasar light curve, which can be approximated as

$$L(t)/L_{\rm peak} \approx \dot{m} = \frac{1}{1 + (t/t_Q)^{\beta}}. \quad (15)$$

Here, $L_{\rm peak}$ is the peak bolometric luminosity, just as the quasar enters the blowout phase, and $L(t)$ is the subsequent bolometric luminosity at time t after the peak ($L = L_{\rm peak}$ at $t = 0$). It is a reasonable approximation to use this to describe both the light curve and accretion rate with $L(t)/L_{\rm peak} \approx \dot{m}$, because not much total BH mass is accumulated in this mode. The equation given assumes $\dot{m} \approx 1$ at peak luminosity, as is true in most of our simulations, but this can be renormalized to any assumed $\dot{m}$ in the constant Eddington ratio growing mode. The functional form of this equation is chosen such that it joins continuously with the constant Eddington ratio exponential light curve at $t = 0$ ($L = L_{\rm peak}$, i.e., at the beginning of the blowout stage) and behaves as our fitted power laws at times large compared to the duration of the blowout.

From our analysis, we determine the power-law decay of the quasar light curve at late times, $L(t) \propto (t/t_Q)^{-\beta}$. To lowest order, a canonical value of $\beta \approx 2$ is indicated by both our fits to the simulation light curves and the self-similar Sedov-Taylor solution for a quasar-driven blast wave and Bondi accretion. But we have also explicitly determined β as a function of peak luminosity, which we inverted to determine $\gamma(L_{\rm peak})$ above. Our fits to the simulations yield

$$\beta \approx \beta_{12} + \frac{d\beta}{d\log L_{\rm peak}}\left[\log_{10}(L_{\rm peak}/10^{12}\ L_{\odot})\right], \quad (16)$$

where $\beta_{12} \approx 1.7 \pm 0.3$ and $d\beta/d\log L_{\rm peak} \approx 0.7 \pm 0.1$. There is some ambiguity in β depending on whether we are fitting to a given blowout or a whole quasar lifetime that includes the "growing" phases (giving shallower effective β), as can be seen directly by the fact that different values of the faint-end quasar lifetime slope γ are suggested by our fitting of the quasar light curves and our fitting of the quasar lifetime to a Schechter function. However, for most theoretical models that require a time-dependent quasar light curve, the parameterization desired is that of an individual blowout, and so these fits are used to determine equation (16) above. Furthermore, the "effective" β from fitting to the entire quasar lifetime is $1/\alpha$, which can be directly determined from the values of $\gamma(L_{\rm peak})$ given in Table 1. There is a significant amount of scatter around this mean relation in β (in contrast to the quite small scatter in α) at a given $L_{\rm peak}$, $\Delta\beta \sim 1$, which can easily be incorporated in theoretical models of quasar populations if desired. Clearly, this formula for $\beta(L_{\rm peak})$ cannot be extrapolated to arbitrary luminosities, as by definition $\beta > 0$ always. In fact, most of the simulations that are well fitted by a low β are so because they never reach a high Eddington ratio and cause a significant blowout event, but rather they quiescently accrete for $\sim$Gyr at moderate Eddington ratios ($\sim 0.1 L_{\rm Edd}$). In either case, this is only important for $L_{\rm peak} \lesssim$ a few times $10^9\ L_{\odot}$, which implies BH masses $\lesssim 10^5\ M_{\odot}$, well below the regimes where the processes we have modeled should be important (and below the limits to which our simulation light curves can be reliably estimated).

There is, unfortunately, significant ambiguity in the appropriate value of t_Q to use in this parameterization of the quasar light curve. If we use equation (16) to estimate the *cumulative* quasar lifetime at low luminosities ($L \ll L_{\rm peak}$, low enough so that the contribution from the growing mode is negligible) and compare this with the Schechter function fits or direct calculation of the quasar lifetime from our simulations at these luminosities, there is a well-defined approximately constant $t_Q \approx t_S = 4.2 \times 10^7$ yr (or, if we allow t_Q to vary with β and demand that this match to the low-luminosity limit of the fitted Schechter function with the same $L_{\rm peak}$, $t_Q \approx \beta t_S \sim 10^8$ yr). However, there are several minor blowout events in our simulations and usually at least two major ones (one at first passage, the other after the final merger). The t_Q that we have constrained by fitting the integrated quasar lifetime is approximately the sum of these contributions. If we instead fit to t_Q in an *individual* blowout, we obtain a significantly smaller value $t_Q \approx 10^7$ yr (as is also suggested by direct examination of Fig. 1). The "appropriate" value of t_Q is then somewhat determined by the ability of a given model to resolve separate starbursts triggered by, e.g., passage of galaxies as opposed to the final merger. However, our simulations constrain the range of reasonable t_Q to $\sim 10^7$–10^8 yr, a significantly narrower range than estimated from observations (e.g., Martini 2004). More important, to first order, t_Q only controls the normalization of various predicted quantities and should be easily calibrated in a given theoretical model by comparison to, e.g., quasar number counts and the normalization of the luminosity function. Even with these uncertainties, the implied appropriate effective t_Q for a given model is an interesting constraint on, e.g., radiative efficiencies of quasars (as the t_Q we fitted scales with the e-folding time for BH growth and hence the radiative efficiency) and coupling of quasar feedback. For example, a shorter t_Q, especially in a model that is summing over several different blowout events, implies more rapid expulsion of gas and more efficient coupling of quasar energy or momentum to the surrounding gas in the blowout phase.

The quasar light curve is, of course, much more complex than we have modeled here. Not only are there several such growing, blowout, and decaying modes in a given merger event, but also each follows a light curve that is not trivial in detail, and our mixed exponential and power-law light curves are not always a good fit. Whenever possible, theoretical models should adopt the most accurate approximations to the quasar light curve available, attempting to resolve the detailed time history where it is important (for example, in estimating the properties and luminosity function of the faint quasar population). However, in many cases this is not possible or feasible, and an analytical approximation for the quasar light curve is necessary or convenient. The above approximation provides a physically motivated parameterization of growing and decaying modes of the quasar light curve, which capture the dependence of the quasar lifetime on both instantaneous and peak luminosities. This does not include the effects of obscuration, which are not important (at least along most sight lines) in the late blowout stages, but can dramatically change the *observed* quasar light curve in various bands near peak luminosity, especially before the blowout phase (e.g.,

Hopkins et al. 2005a, 2006b). These effects can be estimated from the parameterization of column density as a function of instantaneous and peak luminosity in Hopkins et al. (2006b), but these fits are only approximate and do not accurately describe the blowout stage in which luminosity and column density both vary rapidly.

3. DISCUSSION AND CONCLUSIONS

Luminosity-dependent quasar lifetimes, in which quasars spend a longer time at lower L, imply a novel interpretation of the QLF, in which the faint-end slope γ is determined by the slope of the quasar lifetime versus L. For quasars with $L_{\rm peak} \sim L_*$, the break in the observed QLF corresponds directly to the *peak* in the distribution of quasar peak luminosities at any redshift. We have determined γ from both a large range of hydrodynamic simulations (varying BH and host galaxy properties) and from an analytical model of quasar feedback and the $M_{\rm BH}$-σ relation, and we find that it is a monotonic decreasing function of $L_{\rm peak}$ over the range of observed L_*.

As the break luminosity $L_*(z)$ is estimated from observations of the QLF, we use our modeling to predict $\gamma(z)$. Figure 3 shows our result for γ in the redshift interval $z = 0$–4. Above $z \sim 2$–3, the location of L_* is uncertain, and we consider both pure density evolution ($L_* =$ constant or increasing, *solid line*) and pure peak luminosity evolution (declining L_*, *dashed line*) at higher redshifts. The shaded region shows the range predicted by different fitting methods from Table 1, and lines show the cumulative best fit. The form of $L_*(z)$ is taken from Hopkins et al. (2006a, 2006b) but is based directly on fits to the observed QLFs of Ueda et al. (2003), Richards et al. (2005), and Hasinger et al. (2005). We compare this to observations in the hard X-ray (Ueda et al. 2003, *squares*), soft X-ray (Hasinger et al. 2005, *circles*), optical (Wolf et al. 2003, *diamonds*; Hunt et al. 2004, *star*; Richards et al. 2005, *crosses*; Pei 1995, *plus signs*), and radio (Sadler et al. 2002, *asterisks*; Cirasuolo et al. 2005, *triangle*). We convert these to bolometric luminosities and refit or rescale γ with the bolometric corrections of Marconi et al. (2004), which are also discussed in detail in Hopkins et al. (2005c, 2006b) and are similar to those in, e.g., Richards et al. (2005). Although the uncertainties in the observed faint-end slope γ are large, we reproduce its value at all redshifts.

Figure 4 demonstrates how our prediction for the evolution of γ with redshift translates to a LDDE. In each panel, the integrated number density Φ (per comoving volume) of quasars in each of several luminosity intervals is plotted as a function of redshift for $z < 5$. The top left panel shows Φ for intervals in bolometric luminosity L. The effects traditionally fitted to LDDE forms are clear: the density of larger L systems rises more rapidly to a peak at higher redshift, then falls off more slowly. The difference in evolution between two luminosity intervals becomes less dramatic with increasing L, as observed.

Using the bolometric conversions of Marconi et al. (2004) described above, we also consider the observed density evolution in other bands. The top right panel shows density evolution in the B band in intervals of B-band magnitude. This modeling demonstrates why observations in the optical have not found the dramatic LDDE seen in X-ray samples, as the effects do not become pronounced until low luminosities not usually probed in B-band samples (as low as $M_B \gtrsim -18$), and the bolometric corrections also slightly blur out the effect.

The bottom left panel shows our prediction for the hard X-ray, in three intervals of $L(2-10\ {\rm keV}) = L_{\rm HX}$. In order to directly compare with the observations of Ueda et al. (2003), we adopt centimeter gram second (cgs) units. For each interval, we show

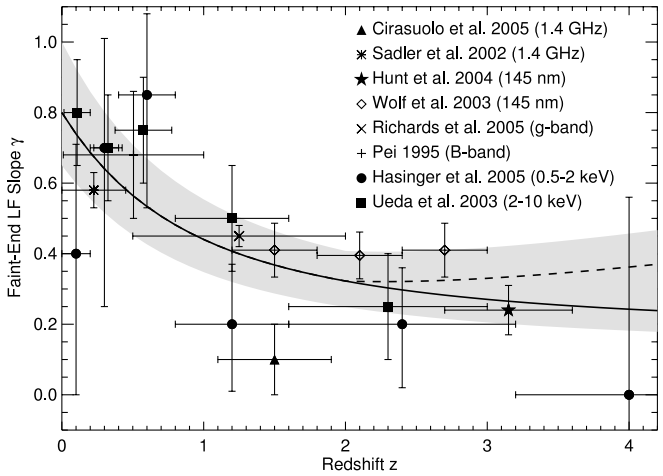

FIG. 3.—Predicted faint-end QLF slope γ as a function of redshift z from the fits in Table 1. The shaded area shows the $\sim$1 σ range depending on fitting method; lines show the mean prediction. Above $z \sim 2$, the solid line assumes pure density evolution for the QLF, and the dashed line assumes pure peak luminosity evolution. Observations are shown (converted to bolometric QLF slopes) from different wavelengths as noted. [*See the electronic edition of the Journal for a color version of this figure.*]

the observed best-fit LDDE model with approximate 1 σ errors at each observed redshift (*circles*). Likewise, the bottom right panel shows our prediction for the soft X-ray in intervals of $L(0.5\text{–}2\ {\rm keV}) = L_{\rm SX}$, where we compare to Hasinger et al. (2005). Again, we show the observed best-fit LDDE model in each observed z interval (*circles*), but we have multiplied the QLF normalization of Hasinger et al. (2005) by a factor of 10 to account for the mean obscured fraction (Hopkins et al. 2006b). For all the above, we adopt pure density evolution for $z \gtrsim 2$. We do not show the errors in our predictions, as these are actually dominated by the uncertainty in the fitted $\dot{n}(L_{\rm peak})$ distribution, not the uncertainty in $\gamma(L_{\rm peak})$.

The results in Figure 4 show that our modeling of the quasar lifetime reproduces the LDDE seen in X-ray samples. We may slightly underpredict the low-luminosity, high-redshift number of sources, but this is not surprising, as it is both where the observations are most uncertain and where, for a fixed low L, the break luminosity becomes much larger than L, meaning that our modeling is based on the quasar lifetime at L well below $L_{\rm peak}$, where it is not well constrained. The $\dot{n}(L_{\rm peak})$ distribution that yields the LDDE seen in Figure 4 has a simple functional form: a lognormal, with constant narrow width and normalization, and a center/peak directly related to the observed break $L_*(z)$ in the QLF. The observed faint end is dominated by brighter peak luminosity sources in dimmer stages of their evolution, and we have demonstrated that, as a result, the *observed LDDE is entirely accounted for by the quasar lifetime as a function of peak luminosity, not by a change in the shape of the peak luminosity or BH mass distribution*.

A more detailed analysis of the faint-end slope should account for several effects. Sample selection from reddening and extinction can be significant, especially if the fraction of obscured objects is a function of luminosity (e.g., Ueda et al. 2003; Hopkins et al. 2006b). Our results agree best with the hard X-ray data, where attenuation is least important, but accurate estimates of γ in other wave bands should account for the joint evolution of obscuration and luminosity. For example, although the observed faint-end B-band QLF is generally too bright to be dominated by obscured sources, extinction of moderately obscured quasars from optical samples means that the B-band QLF will consist of

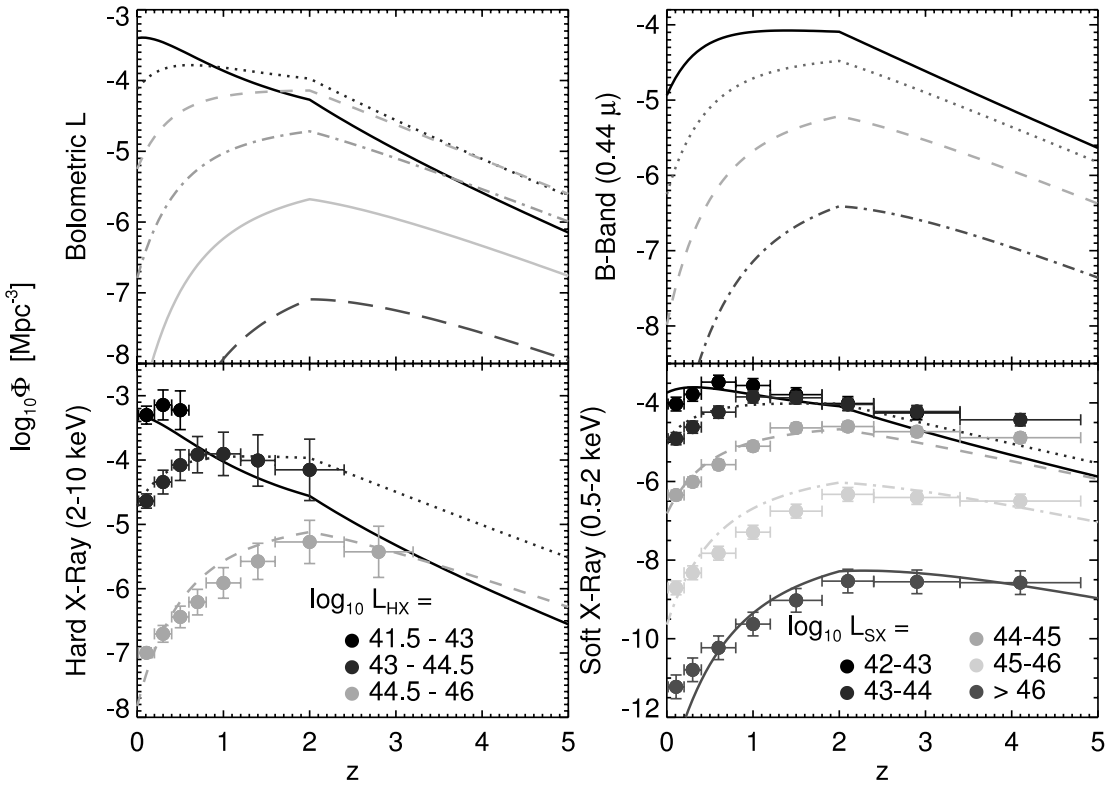

FIG. 4.—Predicted comoving number density in different luminosity intervals as a function of redshift. *Top left*: Bolometric, $\log(L/L_{\odot}) = 9-10$ (*solid line*); 10–11 (*dotted line*); 11–12 (*short-dashed line*); 12–13 (*dot-dashed line*); 12–14 (*triple-dot–dashed line*); and 14–15 (*long-dashed line*). *Top right*: *B* band, $-20 > M_B > -22.5$ (*solid line*); $-22.5 > M_B > -25$ (*dotted line*); $-25 > M_B > -27.5$ (*dashed line*); and $-27.5 > M_B > -30$ (*dot-dashed line*). *Bottom left*: Hard X-ray, $\log(L_{\rm HX}/{\rm ergs\ s^{-1}}) = 41.5-43$ (*solid line*); 43–44.5 (*dotted line*); and 44.5–46 (*dashed line*), compared to observations of Ueda et al. (2003, *circles*). *Bottom right*: Soft X-ray, $\log(L_{\rm SX}/{\rm ergs\ s^{-1}}) = 42-43$ (*solid line*); 43–44 (*dotted line*); 44–45 (*dashed line*); 45–46 (*dot-dashed line*); and >46 (*triple-dot–dashed line*), compared to observations of Hasinger et al. (2005, *circles*, normalization adjusted for obscuration). [*See the electronic edition of the Journal for a color version of this figure.*]

sources in or following the blowout phase (as opposed to objects in earlier, growing phases), introducing corrections in the faint-end slope, as discussed in § 2.3. To probe low accretion rates $L \ll 10^{-2}L_{\rm peak}$, higher resolution simulations with more sophisticated models for low-efficiency accretion (rates well below Bondi or Eddington) and spectral modeling of the corresponding radiatively inefficient accretion flows are needed. Furthermore, the contributions to the faint-end QLF from quasars with $L_{\rm peak} \neq L_*$ (see Hopkins et al. 2006b) will introduce overall curvature into the QLF (weakening the observed break, as observed by, e.g., Wolf et al. [2003] and Richards et al. [2005]), which may aid in constraining the $\dot{n}(L_{\rm peak})$ distribution. Finally, at low enough L (typical of low-ionization nuclear emission-line regions [LINERs] and low-L Seyfert galaxies), we expect different, perhaps stochastic, fueling mechanisms to contribute, which we do not account for.

Nevertheless, our modeling provides physical motivation for both the break luminosity L_* and faint-end slope γ. The break L_* corresponds to the *peak* in the birthrate of quasars with a given peak luminosity (final BH mass), and γ is determined (to first order) by the differential lifetime of these objects, as they spend substantial time at low $L \ll L_{\rm peak}$. The observed γ and its evolution with redshift are a consequence of L_* and its evolution. At high z, L_* is larger, implying that most quasars have higher peak luminosity. From our simulations and analytical modeling of quasar feedback, we expect higher $L_{\rm peak}$ objects to both grow and expel gas more rapidly, when they reach their final mass. Thus, brighter $L_{\rm peak}$ objects "die" more quickly, resulting in a flatter γ as they spend less relative time in any given $L < L_{\rm peak}$.

The observed values and evolution of γ provide a direct test of the model of quasar lifetimes and the QLF proposed in Hopkins et al. (2005a, 2005b, 2005c, 2005d, 2006a, 2006b) and are not predicted by models that invoke idealized on/off or pure exponential quasar light curves. We have also provided analytical forms for the quasar light curve, for use in semianalytical and other theoretical models that require the time-dependent quasar light curve (not merely the statistical properties contained in the quasar lifetime fits we have previously calculated) and that cannot resolve the detailed time dependence of quasar activity in individual mergers and interactions.

Our present results suggest that observations of γ can be used to constrain the differential quasar lifetime, even at $L \ll L_{\rm peak}$, as the faint-end QLF effectively traces the quasar *lifetime* of $L_{\rm peak} \sim L_*$ sources. These observations will further limit models of the distribution of quasar masses and host properties [through $\dot{n}(L_{\rm peak})$] and models of quasar fueling mechanisms and accretion (through the form of the lifetime/slope at low-L). As is clear in Figure 3, the QLF faint-end slope as a function of redshift is still only poorly constrained by observations. Samples that cover the relevant redshift range in a uniform rest-frame wavelength would be particularly valuable, and measurements in different wavelengths can provide different constraints. For example, the faint-end hard X-ray QLF is relatively less affected by obscuration and so provides a better indicator of the bolometric luminosity function, tracing, e.g., relatively low luminosity stages hidden by circumnuclear starbursts. The optical QLF, on the other hand, is in our modeling more closely associated with the peak quasar luminosity and blowout phase, meaning that the faint-end optical QLF (measurement of which would require extending the completeness of current high-redshift optical surveys by several magnitudes) can constrain the underlying peak luminosity distribution,

a particularly valuable quantity in itself because it determines the QLF in all other bands and reflects much more directly the underlying cosmological context, such as merger rates as a function of host galaxy mass. The combination of the two, yielding a reliable bolometric faint-end luminosity function slope and underlying peak luminosity distribution, would allow measurements of $\gamma(z)$ to be translated reliably into $\gamma(L_{\rm peak})$ and could be deconvolved to construct an entirely observational determination of the quasar lifetime as a function of luminosity.

We thank an anonymous referee for comments that significantly improved the text, and also Gordon Richards and Rachel Somerville for helpful discussion and suggestions for the content of this paper. This work was supported in part by NSF grants ACI 96-19019, AST 00-71019, AST 02-06299, and AST 03-07690; and NASA ATP grants NAG 5-12140, NAG 5-13292, and NAG 5-13381. The simulations were performed at the Center for Parallel Astrophysical Computing at the Harvard-Smithsonian Center for Astrophysics.

REFERENCES

Adelberger, K. L., & Steidel, C. C. 2005, ApJ, 630, 50
Bahcall, J. N., Kirhakos, S., Saxe, D. H., & Schneider, D. P. 1997, ApJ, 479, 642
Barkana, R., & Loeb, A. 2001, Phys. Rep., 349, 125
Barnes, J. E., & Hernquist, L. 1991, ApJ, 370, L65
———. 1996, ApJ, 471, 115
Bondi, H. 1952, MNRAS, 112, 195
Bondi, H., & Hoyle, F. 1944, MNRAS, 104, 273
Bullock, J. S., Dekel, A., Kolatt, T. S., Kravtsov, A. V., Klypin, A. A., Porciani, C., & Primack, J. R. 2001, ApJ, 555, 240
Cattaneo, A. 2001, MNRAS, 324, 128
Cattaneo, A., & Bernardi, M. 2003, MNRAS, 344, 45
Cattaneo, A., Blaizot, J., Devriendt, J., & Guiderdoni, B. 2005, MNRAS, 364, 407
Ciotti, L., & Ostriker, J. P. 2001, ApJ, 551, 131
Cirasuolo, M., Magliocchetti, M., & Celotti, A. 2005, MNRAS, 357, 1267
Cowie, L. L., Barger, A. J., Bautz, M. W., Brandt, W. N., & Garmire, G. P. 2003, ApJ, 584, L57
Cox, T. J., Di Matteo, T., Hernquist, L., Hopkins, P. F., Robertson, B., & Springel, V. 2005, ApJ, submitted (astro-ph/0504156)
Di Matteo, T., Springel, V., & Hernquist, L. 2005, Nature, 433, 604
Enoki, M., Nagashima, M., & Gouda, N. 2003, PASJ, 55, 133
Fabian, A. C. 1999, MNRAS, 308, L39
Fiore, F., et al. 2003, A&A, 409, 79
Furlanetto, S. R., & Loeb, A. 2001, ApJ, 556, 619
Granato, G. L., De Zotti, G., Silva, L., Bressan, A., & Danese, L. 2004, ApJ, 600, 580
Haiman, Z., & Menou, K. 2000, ApJ, 531, 42
Hamilton, T. S., Casertano, S., & Turnshek, D. A. 2002, ApJ, 576, 61
Hasinger, G., Miyaji, T., & Schmidt, M. 2005, A&A, 441, 417
Hernquist, L. 1990, ApJ, 356, 359
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, ApJ, in press (astro-ph/0508167)
———. 2006b, ApJS, in press (astro-ph/0506398)
Hopkins, P. F., Hernquist, L., Martini, P., Cox, T. J., Robertson, B., Di Matteo, T., & Springel, V. 2005d, ApJ, 625, L71
Hunt, M. P., Steidel, C. C., Adelberger, K. L., & Shapley, A. E. 2004, ApJ, 605, 625
Kauffmann, G., & Haehnelt, M. 2000, MNRAS, 311, 576
La Franca, F., et al. 2002, ApJ, 570, 100
Lidz, A., Hopkins, P. F., Cox, T. J., Hernquist, L., & Robertson, B. 2006, ApJ, in press (astro-ph/0507361)
Mac Low, M., & Ferrara, A. 1999, ApJ, 513, 142
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
Merloni, A. 2004, MNRAS, 353, 1035
Mihos, J. C., & Hernquist, L. 1996, ApJ, 464, 641
Miyaji, T., Hasinger, G., & Schmidt, M. 2000, A&A, 353, 25
———. 2001, A&A, 369, 49
Murray, N., Quataert, E., & Thompson, T. A. 2005, ApJ, 618, 569
Narayan, R., & Yi, I. 1995, ApJ, 452, 710
Ostriker, J. P., & McKee, C. F. 1988, Rev. Mod. Phys., 60, 1
Page, M. J., Mason, K. O., McHardy, I. M., Jones, L. R., & Carrera, F. J. 1997, MNRAS, 291, 324
Pei, Y. C. 1995, ApJ, 438, 623
Richards, G. T., et al. 2005, MNRAS, 360, 839
Robertson, B., Hernquist, L., Cox, T. J., Di Matteo, T., Hopkins, P. F., Martini, P., & Springel, V. 2006, ApJ, in press (astro-ph/0506038)
Robertson, B., Yoshida, N., Springel, V., & Hernquist, L. 2004, ApJ, 606, 32
Sadler, E. M., et al. 2002, MNRAS, 329, 227
Scannapieco, E., & Oh, S. P. 2004, ApJ, 608, 62
Sedov, L. I. 1946, Prikl. Mat. Mekh., 10, 241
———. 1959, Similarity and Dimensional Methods in Mechanics (New York: Academic Press)
Silk, J., & Rees, M. J. 1998, A&A, 331, L1
Springel, V. 2005, MNRAS, 364, 1105
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
———. 2005, ApJ, 622, L9
Taylor, G. I. 1950, Proc. R. Soc. London A, 201, 175
Ueda, Y., Akiyama, M., Ohta, K., & Miyaji, T. 2003, ApJ, 598, 886
Wolf, C., Wisotzki, L., Borch, A., Dye, S., Kleinheinrich, M., & Meisenheimer, K. 2003, A&A, 408, 499
Wyithe, J. S. B., & Loeb, A. 2002, ApJ, 581, 886
———. 2003, ApJ, 595, 614