Do We Expect Most AGN to Live in Disks?

Philip F. Hopkins1*, Dale D. Kocevski2, Kevin Bundy3

1TAPIR, Mailcode 350-17, California Institute of Technology, Pasadena, CA 91125, USA
2University of California Observatories/Lick Observatory, and Department of Astronomy and Astrophysics, University of California, Santa Cruz, CA 95064 USA
3Kavli Institute for the Physics and Mathematics of the Universe, Todai Institutes for Advanced Study, the University of Tokyo, Kashiwa, Japan 277-8583 (Kavli IPMU, WPI)

Submitted to MNRAS, July, 2013

ABSTRACT
Recent observations have indicated that a large fraction of the low to intermediate luminosity AGN population lives in disk-dominated hosts, while the more luminous quasars live in bulge-dominated hosts (that may or may not be major merger remnants), in conflict with some previous model predictions. We therefore build and compare a semi-empirical model for AGN fueling which accounts for both merger and non-merger “triggering.” In particular, we show that the “stochastic accretion” model – in which fueling in disk galaxies is essentially a random process arising whenever dense gas clouds reach the nucleus – provides a good match to the present observations at low/intermediate luminosities. However it falls short of the high-luminosity population. We combine this with models for major merger-induced AGN fueling, which lead to rarer but more luminous events, and predict the resulting abundance of disk-dominated and bulge-dominated AGN host galaxies as a function of luminosity and redshift. We compile and compare observational constraints from $z \sim 0 - 2$. The models and observations generically show a transition from disk to bulge dominance in hosts near the Seyfert-quasar transition, at all redshifts. “Stochastic” fueling dominates AGN by number (dominant at low luminosity), and dominates BH growth below the “knee” in the present-day BH mass function ($\gtrsim 10^7 M_\odot$). However it accounts for just ~10% of BH mass growth at masses $\gtrsim 10^8 M_\odot$. In total, fueling in disky hosts accounts for $\sim 30\%$ of the total AGN luminosity/ density/BH mass density. The combined model also accurately predicts the AGN luminosity function and clustering/bias as a function of luminosity and redshift; however, we argue that these are not sensitive probes of BH fueling mechanisms.

Key words: galaxies: formation — galaxies: evolution — galaxies: active — star formation: general — cosmology: theory

1 INTRODUCTION

The existence of tight correlations between black hole (BH) mass and properties of the host galaxy spheroid, including spheroid mass/luminosity (Kormendy & Richstone 1995; Magorrian et al. 1998; Kormendy et al. 2011), velocity dispersion (Ferrarese & Merritt 2000; Gebhardt et al. 2000), and binding energy/potential depth (Aller & Richstone 2007; Hopkins et al. 2007b; Feoli et al. 2010) have fundamental implications for the growth of BHs and – given the Soltan (1982) argument which implies that most BH mass was removed angular momentum from dense gas clouds as to produce such an event. Thus, it is commonly assumed that this necessitates an extreme violent galaxy-wide perturbation such as a major galaxy merger. And indeed, gas-rich galaxy mergers are observed to fuel at least a substantial fraction of bright quasars (see e.g. Guyon et al. 2006; Dasyra et al. 2007; Silverman et al. 2008; Bennert et al. 2008; Liu et al. 2009; Veilleux et al. 2009; Letawe et al. 2010; Koss et al. 2010; 2012 and references therein). Such encounters also convert disks into spheroids and further grow the bulge via centrally concentrated gas inflows in a merger-induced starburst (Mihos & Hernquist 1994; Hibbard & Yun 1999; Robertson et al. 2006a; Naab et al. 2006; Cox et al. 2006; Hopkins et al. 2008a,b; 2009a,b). As argued in Hopkins et al. (2007a, b, 2009a,b), Snyder et al. (2011), this deepens the central potential, so a merger both directly strips gas of angular momentum (providing a BH fuel source) and also increases the binding energy of that material (and bulge mass/velocity dispersion), meaning the BH will grow larger even if strong feedback “resists” inflows, before “catching up” to the BH-host relations and self-regulating.

However, recent observations of AGN host morphologies and colors have suggested that major mergers probably do not fuel most low and intermediate-luminosity AGN, as a large fraction appear in “normal” disks (Gabor et al. 2009; Schawinski et al. 2011; Cisternas et al. 2011; Kocevski et al. 2012; Rosario et al. 2011; Civano et al. 2012; Santini et al. 2012; Mullaney et al. 2012). This should perhaps not be surprising. Unlike a bright quasar, fueling a Seyfert (bolometric $L < 10^{12} L_\odot$ or $4 \times 10^{45} $ erg s$^{-1}$) for a typical ~10$^7$ yr episode (see Martini 2004) requires a gas supply within the range of just a single or a few giant molecular clouds (GMCs). There are many alternative mechanisms that could sufficiently disturb the gas in the central regions of the galaxy to as to produce such an event.

These include minor mergers (Hernquist & Mihos 1995; Woods et al. 2006; Woods & Geller 2007; Younger et al. 2008), secular angular momentum loss in bar/spiral arms (for a review, see Jogee 2006 or Toomre-unstable “clumpy” disks (Bournaud et al. 2011), steady-state accretion of diffuse (low-density) hot gas (see Allen et al. 2006; Best et al. 2007 and references therein), or multi-body interactions with nearby star clusters or other clouds (e.g. Genzel et al. 1994). All of these processes do occur in galaxies, and should at least indirectly contribute to AGN fueling insofar as they help remove angular momentum from dense gas.
Many models for the rates and luminosity functions of these processes have been proposed (see references above); however, as far as the central BH is concerned, they are all degenerate in the sense that none directly interacts with the BH. They instead all serve to drive gas into the galactic nucleus, whereupon some other mechanisms (including torus-scale gas-stellar disk processes and the “traditional” AGN accretion disk) must reduce the angular momentum of the gas by an additional six orders of magnitude before it can be accreted. This complicates any model for galactic-scale “fueling” considerably, as it is difficult to imagine any surviving one-to-one correlation between the current BH activity and the galactic state.

Therefore, Hopkins & Hernquist (2006) (hereafter Paper I) attempted to synthesize these processes into a general “stochastic accretion” model; rather than modeling every galaxy-scale event in a fully a priori manner (which involves large uncertainties), it is sufficient to know empirically their important effect for ultimate BH fueling, namely the (resulting) distribution of dense gas and its velocity dispersions in the central regions of the galaxy. Individual “episodes,” corresponding to the gravitational capture of dense gas (e.g. molecular clouds) by the BH directly, occur stochastically but with calculable statistical properties. Coupled to a simple model for AGN feedback, the total duty cycle of AGN as a function of luminosity from these “non-major merger” fueling modes can be estimated. Paper I argued that this can predict accurately many observed properties of $z \approx 0$ Seyferts, including their host galaxies, luminosity functions (LFs), and duty cycles.

One consequence of such models is the idea (discussed in detail in Hopkins & Hernquist 2009a; Draper & Ballantyne 2012; Santini et al. 2012) that there is some characteristic host bulge/BH mass (and corresponding quasar luminosity) below which these more ubiquitous mechanisms dominate AGN fueling (being more common and requiring less bulge growth to deepen the central potential in this mass regime). Above this division, less violent mechanisms are simply inefficient (they may still happen, but they do not sufficiently raise the bulge mass, so BHs quickly self-regulate and do not experience any significant lifetime of high-Eddington ratio growth) and the population requires more extreme mechanisms such as major mergers to build the most massive bulges and (corresponding) BHs.

Coupling these models to empirical estimates of the evolution of galaxy mass functions, gas fractions, and other quantities, Paper I attempted to extend the model predictions to high redshifts. The predicted LF from that paper at $z = 2$ is shown in Fig. 1. Qualitatively, we see the transition discussed above, with the stochastic mode dominant at low luminosities.

But the recent observations discussed above find that disk-dominated hosts (i.e. candidates for the “stochastic” mode, as opposed to post-major merger systems which may not, on average, appear as disks) dominate the population even at luminosities an order-of-magnitude larger than the “transition point” predicted in Fig. 1.

Clearly, there is something wrong with these models. However, the Paper I model remains a good description of some observations at $z = 0$, and captures many of the key processes from simulations which appear to be robust even as resolution and the treatment of AGN, star formation, feedback, and ISM physics has improved (see the comparisons in Debuhr et al. 2010; 2012; Johansson et al. 2009; Choi et al. 2012). We therefore, in this paper, re-visit these models for AGN fueling, but attempt to incorporate them into a modern, and observationally-constrained “population synthesis” model. This allows us to use more accurate assumptions and models for the evolution of the galaxy population with redshift (including galaxy mass/luminosity functions, merger rates, and gas fraction distributions), to define the “background” on which AGN fueling occurs. We also attempt to compile a range of observational constraints of the AGN host galaxy population, spanning redshifts $z \approx 0-2$, to develop the most rigorous constraints to date and so construct a better estimate of the integrated contribution of major merger vs. non-major merger mechanisms towards BH growth.

2 THE MODELS

The model we will present here supposes two independent AGN fueling populations. A “major merger-induced” population, and a “stochastic” population (which essentially includes all non major merger-induced events). We will make the same consistent assumptions about the background population and AGN behavior in fueling events in both cases, but treat the total AGN luminosity function as simply the sum of the predicted duty cycles from both sub-populations.

1 It is important to note that even major galaxy mergers can and do leave disk-dominated remnants under the right circumstances (when they are sufficiently gas rich and have favorable initial orbital parameters; see Springel & Hernquist 2005; Robertson et al. 2006a; Hopkins et al. 2009a). However, if major mergers were the dominant AGN fueling mechanism, any plausible distribution of orbital parameters (combined with the gas fractions estimated observationally in these populations) would at least produce a significant enhancement of bulge-dominated or “bulge-enhanced” galaxies relative to a control population at the same stellar mass (see Hopkins et al. 2009a). This is not observed except at higher AGN luminosities, as we will discuss further in the text.
2.1 Merger-Induced Fueling

The major merger-induced quasar fueling model here is taken directly from a series of papers: Hopkins et al. (2009d, 2010a,b,c), and Younger & Hopkins (2011). We use the most recent update to the model, presented in Hopkins et al. (2010c) (hereafter Paper II). There are three basic components of the model, for which all details are given in Paper II. Since we are only taking the results from that paper, we will only briefly summarize the key model elements here.

(1) At a given redshift, we begin with the observed galaxy mass functions and gas fraction distributions. This defines the empirical “background population” onto which we will add assumptions for AGN fueling. Of course, other types of models such as semi-analytic models and cosmological simulations attempt to predict these properties a priori, then further add assumptions about AGN fueling (see, e.g. Di Matteo et al. 2008; Somerville et al. 2008; Fanidakis et al. 2011 and references therein). But this adds considerable uncertainty. Since our focus here is on the AGN population alone, we prefer the Paper II “semi-empirical” model approach, which allows us to isolate the assumptions relevant to the AGN population. The actual mass function data are compiled from a range of sources.\footnote{Mass functions measurements are compiled from Bell et al. (2003); Arnouts et al. (2007); Ilbert et al. (2010); Pérez-González et al. (2006); Fontana et al. (2006); Marchesini et al. (2009);Kajisawa et al. (2009). Where different measurements overlap at the same redshift, we use the differences between them (added with the appropriate quoted error bars) to define the empirical uncertainty in the MF. The compilation is chosen such that there are always at least two overlapping measurements at each redshift. We then interpolate log-linearly between the median MF measurements at each redshift. We combine these with measurements of the mean and scatter in gas fractions as a function of stellar mass and redshift, from Bell & de Jong (2000); McGaugh (2005); Calura et al. (2008); Shapley et al. (2005); Erb et al. (2006); Puech et al. (2008); Mannucci et al. (2009); Cresci et al. (2009); Forster Schreiber et al. (2009); Erb (2008). For details, see Paper II.}

(2) Using a simple abundance-matching halo-occupation model (i.e. forcing the population to match observed number densities and clustering; see Conroy et al. 2006) each observed member of the galaxy population is assigned to a halo, from which the merger rate can be calculated from fits to the cosmological halo-halo merger rates. In other words, from a cosmological simulation,
all halo-halo mergers at a specific redshift of interest are identified. Each halo is then assigned a galaxy via abundance matching (and a dynamical friction time is assigned between the halo-halo and galaxy-galaxy merger). This leads to the galaxy-galaxy merger rate. Extensive discussion and tests of this methodology are presented in Hopkins et al. (2010b); we simply note here that taking the merger rate directly from observations gives a similar result, but with large uncertainties (comparisons with observations and semi-analytic models are in Stewart et al. 2009a, Jogee et al. 2009, Loz et al. 2011).

(3) For each such “semi-empirically” assigned merger, we then attach an AGN fueling model. Specifically, in a series of papers, Hopkins et al. (2006a,b, 2007a,b) use a simple model for AGN accretion rates and feedback to fit the resulting AGN lightcurves in galaxy-galaxy merger simulations as a function of galaxy mass, redshift, and gas fraction of the progenitors. Since we have this information in our mock population, we can then simply assign the corresponding fitted (bolometric) lightcurve (or equivalently, probability of being seen at a given luminosity) to every merger. The exact functional parameterization is given in Paper II; this is compared to observations (from Yu et al. 2005, Kauffmann & Heckman 2008) of the duty cycle distribution and alternative “synthesis models” (from Merloni & Heinz 2008, Shankar et al. 2009) in Hopkins & Hernquist (2009b). For our purposes here, the important conclusion in that paper is that the results are all similar, so (within the relatively large uncertainties) it makes relatively little difference which parameterization we adopt.

We stress that we are not presenting any modifications or revisions to this model; we take the predicted “merger-induced” AGN luminosity functions exactly as calculated in Paper II. Readers interested in how variations within that model affect the results presented here should see Paper II, Appendix B.

2.2 Stochastic Fueling

Paper I argued that AGN can and should also be triggered stochastically in non-merging systems via a variety of detailed mechanisms. We therefore crudely assign all “non-major merger” processes to the “stochastic fueling” category. In Paper I, however, this is “synthesized” into an estimate of the resulting luminosity function using very crude assumptions about the galaxy population and its redshift evolution. The methodology described above for the merger-induced population provides a much more well-motivated “background” onto which we apply the models from Paper I.

The two basic steps are as follows:

1. At a given redshift, we again begin with the observed galaxy mass functions and gas fraction distributions from Paper II, identical to the first step in § 2.1.

2. With this information, we apply the model from Paper I for the cumulative duty cycle of activity owing to non-major merger fueling mechanisms. This is the major model addition in this paper, to the model presented in Paper II.

We begin by assigning a BH mass to every galaxy in the

Figure 3. Predicted fraction of AGN in the stochastic mode, as a function of luminosity and redshift, from the models in Fig. 2. The mean of the model range is shown as the solid line with the ±1σ (dashed) and ±2σ range (dotted). Since the duty cycle of “stochastically” fueled systems is dominated by gas-rich disks, and most major merger-triggered systems that induce strong bulge and BH growth will appear as spheroids, we compare the observed fraction of disk-dominated AGN host galaxies in different luminosity/redshift intervals. At low redshift, measured from the PG quasar sample of Dunlop et al. (2003), Floyd et al. (2004) circles). At intermediate redshift, measured in low-luminosity AGN in COSMOS in Gabor et al. (2009), Cisternas et al. (2011) triangle), and in true (Type 2) quasars in Zakamska et al. (2006, 2008), Liu et al. (2009) inverted triangle). At z ∼ 2, measured in CANDELS AGN host galaxies in the low and high-luminosity sub-samples from Kocevski et al. (2012) squares).
model, at each redshift, according to the simple approximate observed relation:

\[ M_{\text{BH}} \approx 0.0014 (1 + z)^{0.5} f_{\text{fudge}} M_\star \]  

(1)

with a lognormal intrinsic scatter of \( \approx 0.3 \text{dex} \) in \( M_{\text{BH}} \). This is a purely empirical estimate of a best-fit to a range of observations (McLure \& Dunlop 2004; Peng et al. 2006; Woo et al. 2006; Adelberger \& Steidel 2005b; Shields et al. 2006; Salvatiener et al. 2007; Treu et al. 2007; Bennett et al. 2010) for a recent review see Koemendy \& Ho 2013). We stress that the relation and scatter are well-anchored at \( z = 0 \), but increasingly uncertain at high redshifts. But theoretical models give similar redshift evolution, mostly owing to the more gas-rich, compact nature of high-redshift hosts (see Hopkins et al. 2007a; Johansson et al. 2009; Choi et al. 2012). In any case, the results of varying the assumed redshift evolution are shown in Paper II (Figs. B1 \& B2): since it appears in almost identical form in the merger model, it will shift the normalization of both stochastic and merger-triggered AGNs in luminosity \( L_{\text{bol}} \), but not much alter their \( \text{relative} \) behavior, which is most interesting here. The scatter is observed, but has little effect – it is important for the abundance of the most massive BHs (above the “break” in the galaxy mass function, corresponding to luminosities well above the turnover in the LF; see Paper II Fig. B3), but we will show that the stochastic mode is sub-dominant in this regime in any case (so assuming any scatter \( \lesssim 1 \text{dex} \) makes little difference). Finally, \( f_{\text{fudge}} \) is estimated from our galaxy mass functions, but is formally degenerate with the normalization and redshift evolution of the relation; where (at high redshifts) it is poorly determined we simply assume \( f_{\text{fudge}} \approx 0.3 \), since this appears to give a good fit to observations of the relation between BH and total stellar mass (see references above).

With BH masses assigned, we need to assign luminosities. Since the triggering mode is “random” (on cosmological timescales), it is sufficient to simply assign a duty cycle (probability of observing a given luminosity). This is calculated for the stochastic mode in Paper I, assuming a triggering rate determined by capture of cold gas in the nucleus and subsequent regulation of accretion via feedback. It is shown there that this can be simply parameterized as:

\[ \frac{dP}{d \log L} = \alpha \left( \frac{f_{\text{fudge}}}{0.1} \right) \left( \frac{L}{L_{\text{Edd}}(M_{\text{BH}})} \right)^{-\beta} \]  

(2)

with \( \alpha \approx 0.003 \) and \( \beta \approx -0.6 \) (see Yu et al. 2005; Shankar et al. 2008; Kelly et al. 2009; Hickox et al. 2009; Bonoli et al. 2009; Kaufmann \& Heckman 2008; Trump et al. 2009). We truncate this at \( L > L_{\text{Edd}}(M_{\text{BH}}) \approx 3.3 \times 10^{44} (L_\odot / M_\odot) M_{\text{BH}} \). Note that in Hopkins \& Hernquist (2006b), this is compared to an extensive ensemble of observational constraints and measurements of the Eddington ratio distribution at \( z \approx 0 \), and shown to agree well (especially for moderate luminosity AGN), with relatively little allowed range in \( \alpha \) or \( \beta \) relative to the theoretically predicted values. Therefore, if we simply adopt a best-fit to the observed \( L / L_{\text{Edd}} \) at \( z = 0 \), we would obtain a nearly identical prediction. This duty cycle is simply convolved with the BH mass function to obtain the stochastic-mode LF.

We emphasize that the AGN-centric equations in step (2) were developed in Paper I. What distinguishes our predictions here from those therein is the model for the galaxy population. In Paper I, some very simple assumptions – many of which appear to be inaccurate in light of observations in the last several years – were made to extrapolate the model from \( z = 0 \). Implicitly, these would (for the same AGN fueling and feedback model) correspond to a very different distribution of galaxy masses and gas fractions, from that on which we develop here. The most important differences are: (1) observations of high-redshift galaxies indicate they are more gas-rich than assumed in Paper I, with gas fractions approaching \( \approx 50\% \) even in high-mass systems (see e.g. Tacconi et al. 2010); (2) high-mass galaxies are also more abundant at high redshift than was assumed in Paper I, indeed many cosmological simulations and semi-analytic models still underestimate the number density of galaxies with stellar masses \( \gtrsim 10^{11} M_\odot \) at \( z \gtrsim 2 \) (see Hayward et al. 2012). There was also no explicit model for the merger-induced population in Paper I; here we include that developed in Paper II.

3 RESULTS

Fig. 2 plots the predicted AGN luminosity functions from the models for both major merger-induced and “stochastic” (non-major merger) mechanisms, at several redshifts. As discussed in $\S$2, the empirical uncertainties in the galaxy number density, gas fractions, and merger rates at each redshift are added in quadrature to give the “total uncertainty” in the model predictions (shaded range in the plots). This should be thought of as the uncertainty owing to differences in the AGN fueling models (which might be quite large), but owing to un-related uncertainties in the background galaxy population.

To avoid uncertainties owing to obscuration, we compare the model predictions (which are really for the bolometric BH accretion rates and luminosities) with empirical estimates of the bolometric (obscuration and wavelength-corrected) AGN luminosity function presented in Hopkins et al. (2007a).
The sum of the stochastically fueled AGN and merger-induced AGN LFs agrees very well with the observed bolometric LF at most redshifts. This is reassuring, and it also suggests that some large additional fueling mechanism or driver is not needed to explain the observed demographics.

Clearly, stochastically fueled systems are predicted to dominate at the lowest luminosities, while merger-induced populations dominate at the highest luminosity. The transition between them occurs at a broadly similar luminosity \( \sim 10^{10} L_\odot \) (the traditional Seyfert-quasar divide) at all redshifts.

It is important to stress that we have not adjusted or “fine-tuned” any parameters in the model here to reproduce the observations. Moreover, the stochastic and merger-induced models are independent predictions, so it is encouraging that they appear to accurately sum to reproduce the total luminosity function. However, we should emphasize that the model presented here is not unique, and a combination of many observations is needed to fully break degeneracies in models. The AGN LF alone is a relatively poor constraint on fueling mechanisms: by allowing different AGN lightcurves, or including minor mergers, it is possible to fit the low-luminosity LF with only merger-induced fueling (see the models in Hopkins et al. 2005, 2006; Somerville 2009). On the other hand, by assuming a much stronger “secular” mode (in which traditional disk bar instabilities are assumed to channel 100% of the gas into the nucleus in a single burst), Fanidakis et al. (2011) show they can plausibly reproduce the high-luminosity LF. And at high redshifts and high-\( L_{\text{bol}} \), we see that the “allowed range” owing to uncertainties in galaxy number densities and merger rates is very large – this means that sufficient degeneracies exist such that the bright, high-redshift LF has little power to constrain fueling models.

In Fig. 1 we use this result to estimate the distribution of host population “type” versus mass. Specifically, we plot, at each redshift, the fraction of the population at each \( L_{\text{bol}} \) that are predicted to be fueled in the “stochastic” mode (as opposed to the major merger mode). At all redshifts, we see a continuous increase in the predicted merger-relic AGN population with luminosity, with the merger-mode being negligible at Seyfert luminosities but becoming dominant at QSO luminosities \( \gg 10^{12} L_\odot \). There is some quantitative increase in prevalence of mergers at intermediate luminosities at high redshifts, but the effect is small.

Very crudely, most “stochastically fueled” systems should be disk-dominated. To lowest order, this is simply a reflection of the background galaxy population (which, at lower masses where the fueling mode is dominant, is mostly disk-dominated). At second-order, at fixed mass, in the model we adopt (\( \sim 2 \) AGN activity does require a gas supply, so fueling is enhanced in gas-rich systems, which are overwhelmingly disk-dominated (though of course there will be some, albeit rarer, gas-rich spheroids). In contrast, most major merger-fueled systems should be bulge-dominated, since such mergers tend to build large bulges.

We therefore compare the predicted “stochastically fueled” fraction of AGN with the fraction of disk-dominated AGN hosts, as a function of luminosity and redshift. At low redshifts, we compare with the PG quasar sample of Dunlop et al. (2003) and Floyd et al. (2004) (we plot the fraction with best-fit bulge-to-total mass ratio \( B/T < 0.5 \) in bins of \( L_{\text{bol}} \), estimated from the observed nuclear V-band luminosities, with Poisson errors). At \( z \sim 0.5 - 1 \), we compare the low-luminosity sample from COSMOS studied in Gabòr et al. (2009) and Cisternas et al. (2011); we plot the “final” quoted fraction of disk-dominated galaxies in the sample (with the approximate \( \sim 10 \% \) systematic difference between classifiers quoted therein) and 90% range of \( L_{\text{bol}} \) estimated from the hard X-ray luminosities. We also compare the sample of true quasars in Zakamska et al. (2006, 2008); Liu et al. (2009) at \( z \sim 0.3 - 0.7 \). These are Type-II (obscured) objects whose host morphologies can be determined, of which 1/9 is a disk galaxy, and the remainder are clearly spheroid-dominated and/or visible late-stage mergers. At \( z \sim 2 \), we compare with the CANDELS sample from Kocevski et al. (2012), again using the quoted distribution of visual classifications for their low and high-luminosity samples (\( L_{\text{bol}} \) estimated here from the hard X-ray luminosities).

This is only a very rough comparison, to see whether the predictions are at all reasonable given present observational constraints on AGN host galaxy morphologies. Of course, as discussed in Trump (2011), considerable care is needed regarding the different selection in these samples. We have attempted to match in luminosity and redshift, but other aspects (color, AGN selection criteria, morphological classification method, imaging wavelength) must be investigated in more detail in future work before any rigorous, quantitative “best-fit” to these observations can be presented.

In Fig. 3 we use the models to predict the clustering amplitude of AGN populations as a function of redshift and luminosity. Recall that, in the model, every mock AGN has a known host galaxy stellar mass and (via abundance matching) assigned host halo mass. We can then simply adopt the expression for the clustering amplitude (bias) as a function of halo mass and redshift from Sheth et al. (2001), and use this to calculate the mean bias of the population in bins of AGN luminosity and redshift. We show the mean bias

---

6 As noted in § 1, we stress that a sizeable fraction of major mergers will produce disk-dominated galaxies, especially at high-redshifts where the disks are more gas-rich (see e.g. Springel & Hernquist 2005; Robertson et al. 2006a, Hopkins et al. 2009a, 2009b, 2009c, 2013; Governato et al. 2009). However, disk survival in mergers is most efficient at low galaxy masses, where the disks are actually gas-dominated; the large BH masses where the merger-induced mode is dominant imply bulge masses \( \gtrsim 10^{11} \odot \). Moreover, large surviving disks generically require conditions (gas distributions and orbits) that suppress strong inflows, the opposite of the regime we are interested in here where strong bulge growth and AGN fueling will result. As a result, this can be critical for the abundance of disks at low masses (Hopkins et al. 2009e; Somerville et al. 2008; Stewart et al. 2009b; Puch et al. 2012), but is probably not the dominant process in the mergers that produce bright quasar activity, of interest here.

7 For the clustering calculation, we adopt the WMAP5 cosmological pa-
as a function of redshift, for AGN in different luminosity intervals, and the bias as a function of luminosity at specific redshifts. We compare this to the compilation of observations in (Hopkins et al. 2007c) and Shen et al. [2012].

The agreement with observations is good. Unfortunately, it appears that clustering is not a strong constraint on models of AGN fueling mechanisms. For example, the trend of bias with redshift, either for AGN near the “knee” in the LF or weighted across the LF, is similar in models which assume only merger fueling (Lidz et al., 2006; Hopkin et al., 2007c), only secular (non-merger) fueling (Panidakis et al., 2011; Croton et al., in preparation), or which make no statement about fueling but only assume a random duty cycle independent of galaxy properties (Croton, 2009; Conroy & White, 2013). And we see here that the predicted “transition” between the stochastic mode and merger mode does not imprint any characteristic feature in the clustering as a function of luminosity (at a given redshift).

Finally, we should note (as discussed in Paper II), that since the synthesis model here essentially assumes the BH-host correlations observed in order to “populate” systems (and the simulations to which the AGN lightcurves are calibrated fall closely on these relations; Di Matteo et al., 2005; Hopkins et al., 2007a), it is automatically implicit that they also reproduce the local BH mass function. An explicit calculation and comparison with the mass function estimated in Marconi et al. (2004) or Shankar et al. (2009) confirms this. This is also implicit since the extended “continuity equation” version of the Soltan 1982 argument (see Yu & Lu, 2004; Yu & Tremaine, 2002; Merloni & Heinz, 2008) shows consistency between the quasar LF and BH mass function. Therefore this also has little power to constrain fueling models.

4 DISCUSSION

4.1 Overview

This paper presents a simple “semi-empirical” population synthesis model for AGN fueling that distinguishes between major merger-triggered and non-major merger triggered (“stochastic”) activity. We show that this can plausibly account for the bolometric AGN luminosity function from $z = 0 – 6$ and $L_{bol} \approx 10^{10} – 10^{13} L_{\odot}$, observations of the distribution of AGN host morphologies, and observed AGN clustering amplitudes as a function of redshift and luminosity.

Our model builds on the “semi-empirical” model approach from Paper II, which means that the “background” galaxy population properties are taken from observations. The theoretical “layer” added on top of this is the AGN fueling/feedback model. The non-major merger model is taken from Paper I; this model attempts to calculate the probability that cold, dense gas reaches an AGN and can be accreted, based on known empirical properties of galaxies (their distribution of gas fractions and the spatial distribution of that gas). The advantage of this model is that it makes no specific assumption about how this gas “gets into” the galaxy center in the first place – it can be contributed or torqued by minor mergers, disk instabilities (bars, spiral arms, massive clumps), directly fueled by “cold flows” or accretion streams, or simply random turbulent cloud-cloud scattering. Since these all contribute in a statistical sense to the distribution of gas fractions and dispersions, they are all accounted for implicitly. The merger model is taken from Paper II, using empirically-constrained merger rates convolved with a library of results from galaxy-galaxy merger simulations with simple prescriptions for BH growth and feedback.

We reach similar conclusions to those recently reached by (Hopkins et al. 2008 squares at $z \approx 0.5, 1.5, 2.0$), Myers et al. (2007a) triangles at $z \approx 0.5, 1.5, 2.0$, and Adelberger & Steidel (2005a) stars at $z \approx 2.0$. 

\textcopyright 0000 RAS, MNRAS 000, 000–000
4.2 The Role of Stochastic Fueling as a Function of Mass/Luminosity

Quantitatively, if we integrate the models here, we estimate that non-major merger AGN contribute about $\sim 30\%$ of the total AGN luminosity density and BH mass density of the Universe. This agrees well with some recent observational estimates ([Georgakakis et al. 2009][Koss et al. 2010]). But the predicted contribution of mergers is strongly BH mass and luminosity-dependent. Predicted low-mass BH growth is strongly dominated by non-major merger mechanisms, with nearly all the BH mass at $<10^7 M_\odot$ and most of the BH mass at $<10^8 M_\odot$ (at all redshifts) accreted in the “stochastic” mode. But above $M_{BH} \gtrsim 10^7 M_\odot$, most of the mass is accreted in the merger-induced mode. As argued in § 1, this seems physically reasonable. Growing a BH significantly above $\sim 10^8 M_\odot$ requires inflows that can channel a large fraction of an entire galaxy gas supply to $\sim 10$ pc in a Salpeter time – essentially a single galaxy dynamical time! Galaxy interactions represent one of the only well-established and sufficiently violent mechanisms to accomplish this.

There are a number of additional, indirect observational suggestions that there is a transition from essentially random fueling of AGN at Seyfert luminosities to merger-induced fueling in true quasars; some of these are summarized in [Hopkins & Hernquist 2009a]. This includes the fact that quasars exhibit excessive small-scale (sub-halo scale) clustering while Seyferts do not ([Serber et al. 2006][Myers et al. 2007][Serber et al. 2006][Hennawi et al. 2009][Shen et al. 2010]): quasar duty cycles rise more sharply with redshift (in agreement with observed merger rates), as opposed to Seyfert duty cycles which increase more slowly or less in agreement with galaxy gas fraction evolution (see the compilation in [Hopkins & Hernquist 2009a] and discussion in Draper & Ballantine 2012), and the much larger prevalence of post-staburst populations in true quasars (Brotherton et al. 1999[Vanden Berk et al. 2006][Lozzi et al. 2008][Wang et al. 2008][Shi et al. 2007][Kewley et al. 2006][Nandra et al. 2007][Silverman et al. 2008][Higdon et al. 2008]). A particularly compelling argument is the fact that, below the minimum BH mass required (by the Eddington limit) to power a quasar ($\sim 3 \times 10^7 M_\odot$), most BH host bulges are “pseudobulges,” generally believed to form via secular processes (or minor mergers); above this mass, essentially all the bulges are “classical,” and so formed (at least initially) in (major) mergers (see e.g. [Kormendy & Kennicutt 2004][Kormendy & Bender 2012][Fisher 2006][Fisher & Drory 2008][Hopkins & Hernquist 2009a][Balcells et al. 2007][Gadotti 2009]).

4.3 How Does This Relate to Star Formation?

This may mirror a predicted and increasingly observationally well-established distinction in what powers galactic star formation. At low star formation rates, “quiescent” star formation (steady consumption of gas in disks) dominates, but the highest star formation rate systems are essentially all major mergers. At low-redshifts, this has been well-known for $\sim 20$ years (with the transition occurring at IR luminosities of ULIRGs, see e.g. [Joseph & Wright 1985][Sanders & Mirabel 1996][Evans et al. 2009]). Models predict that the same should be true at high redshifts, but with a higher “transition” luminosity since all systems – mergers and quiescent galaxies – shift up to higher star formation rates at higher redshifts as all galaxies become more gas-rich (see e.g. Paper II, and references therein). Observations have now progressively mapped this transition from $z \sim 0 - 2$ (see e.g. [Tacconi et al. 2008][Younger et al. 2009][Casey et al. 2009][Melbourne et al. 2008][Dasyra et al. 2008][Sargent et al. 2012][Zamojski et al. 2011][Kartaltepe et al. 2012]. However, there are two critical differences between the star-forming and AGN populations. First, the “transition luminosity” $L_{ssf}$ for star-forming populations (between “quiescent” star formation and merger-induced bursts) increases rapidly with redshift, rising from $\sim 10^{11.5} L_\odot$ at $z = 0$ to $\sim 10^{13} L_\odot$ at $z > 2$ (see references above). The predicted evolution in the AGN transition luminosity $L_{bol}$ is much weaker (nearly constant at $10^{12} L_\odot$, in the model here). The rapid evolution in $L_{SS}$ is widely attributed to the fact that, as gas fractions systematically increase at high redshift (itself owing to more rapid cosmological gas inflow rates), the associated star formation rates rise super-linearly according to the (Kennicutt 1998) relation. However, in most models, the maximum AGN $L_{bol}$ is fundamentally limited by the BH mass (via the Eddington limit), not the galactic gas supply. Increasing gas fractions at high redshifts therefore tends to increase the AGN duty cycle in most models, but has relatively little effect on the characteristic luminosities of AGN (see e.g. [Hopkins et al. 2007][Johansson et al. 2009]). Since the mass at the “break” in the galaxy stellar mass function (hence implied BH masses, if the BH-host correlations still apply) does not evolve very strongly from $z \sim 0 - 2$, this implies that the AGN “transition” $L_{bol}$ should be more constant than the star formation transition $L_{SS}$. Second, it is increasingly clear in both models and observations that the integrated total of star formation in the Universe is dominated by the “quiescent” mode. However, the integrated BH growth (at least in the model here) is dominated by the merger-induced mode. In the model, this is closely related to the origin of galactic bulges. Most of the total stellar mass in bulges is in “classical” bulges, which a wide range of observational and theoretical constraints indicate formed in violent mergers (see references in § 4.2 above; for reviews, see [Kormendy & Kennicutt 2004][Kormendy & Bender 2012][Fisher & Drory 2008][Hopkins & Hernquist 2009a][Balcells et al. 2007][Gadotti 2009]). However, even if most of the bulge is formed in such an event, it is primarily via the transformation of pre-existing stars from a disk to a bulge via violent relaxation. A wide variety of independent observations (including e.g. stellar age and metallicity distributions, kinematics, phase-space density profiles, gas density and star formation properties in ongoing mergers, and more) indicate that only a small fraction ($\sim 10\%$ in an $\sim L_\odot$ spheroid) of the final stellar mass is actually formed in a nuclear starburst “driven by” the merger (for a rigorous discussion, see [Hernquist et al. 1993][Hopkins & Hernquist 2010]). However, these inflows can dominate the formation of stars at extremely high densities in galaxy nuclei (much larger than the densities at the center of disks). And since it is nuclear inflows that ultimately matter
for BH growth, these same inflows may dominate the growth of the BH population.

Empirically, if it is true that BH mass is correlated with bulge mass (at the masses \( \gtrsim 10^7 \) that contain most of the mass density), then it follows that most of the BH mass growth follows the mechanisms that build up most bulge mass (not necessarily the mechanisms that initially form those stars, if they are in disks). And most bulge mass is in classical (presumably merger-built) bulges. Though a subtle distinction, there is evidence that BH growth in luminous AGN is not strictly contemporaneous with most of the star formation, though they follow the same mean trends in a sufficiently time-averaged sense (as they must, for any linear BH-host mass relation); the sense is such that BH growth is biased towards more spheroid-dominated, and at high luminosities more obviously merging systems (see e.g. Zheng et al. 2009; Kartaltepe et al. 2010). In other words, this would say that most of the star formation is in low-mass, relatively low-luminosity galaxies, whereas most of the BH mass is in high-mass, bulge-dominated galaxies. However, it remains a critical, ultimately empirical question, to test whether BHs really do correlate with bulge (and not disk) properties, especially at higher redshifts (for a recent review, see Kormendy & Ho 2013).

ACKNOWLEDGMENTS

Support for PFH was provided by NASA through Einstein Postdoctoral Fellowship Award Number PF1-120083 issued by the Chandra X-ray Observatory Center, which is operated by the Smithsonian Astrophysical Observatory for and on behalf of the NASA under contract NAS8-03060. KB was supported by the World Premier International Research Center Initiative (WPI Initiative), MEXT, Japan.

REFERENCES

Adelberger, K. L., & Steidel, C. C. 2005a, ApJ, 630, 50
—. 2005b, ApJL, 627, L1
Aird, J. E., et al. 2009, MNRAS, 1743
Allen, S. W., Dunn, R. J. H., Fabian, A. C., Taylor, G. B., & Reynolds, C. S. 2006, MNRAS, 372, 21
Aller, M. C., & Richstone, D. O. 2007, ApJ, 665, 120
Armouts, S., et al. 2007, A&A, 476, 137
Balcells, M., Graham, A. W., & Peletier, R. F. 2007, ApJ, 665, 1104
Barger, A. J., Cowie, L. L., Mushotzky, R. F., Yang, Y., Wang, W.-H., Steffen, A. T., & Capak, P. 2005, AJ, 129, 578
Bell, E. F., & de Jong, R. S. 2000, MNRAS, 312, 497
Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D. 2003, ApJS, 149, 289
Bellovary, J., Brooks, A., Volonteri, M., Governato, F., Quinn, T., & Wadsley, J. 2013, ApJ, in press, arxiv:1307.0856
Bennert, N., Canalizo, G., Jungwiert, B., Stockton, A., Schweizer, F., Peng, C. Y., & Lacy. M. 2008, ApJ, 677, 846
Bennert, V. N., Treu, T., Woo, J., Malkan, M. A., Le Bris, A., Auger, M. W., Gallagher, S., & Blandford, R. D. 2010, ApJ, 708, 1507
Best, P. N., von der Linden, A., Kauffmann, G., Heckman, T. M., & Kaiser, C. R. 2007, MNRAS, 379, 894
Bonoli, S., Marulli, F., Springel, V., White, S. D. M., Branchini, E., & Moscardini, L. 2009, MNRAS, 396, 423
Bournaud, F., Dekel, A., Teyssier, R., Cacciato, M., Daddi, E., Juneau, S., & Shankar, F. 2011, ApJL, 741, L33
Brotherton, M. S., et al. 1999, ApJL, 520, L87
Brown, M. J. I., et al. 2006, ApJ, 638, 88
Calura, F., Jimenez, R., Panter, B., Matteucci, F., & Heavens, A. F. 2008, ApJ, 682, 252
Cappelluti, N., Ajello, M., Burlon, D., Krumpe, M., Miyaji, T., Bonoli, S., & Greiner, J. 2010, ApJL, 716, L209
Casey, C. M., Chapman, S. C., Beswick, R. J., Biggs, A. D., Blain, A. W., Hainline, L. J., Ivison, R. J., Muxlow, T., & Smail, I. 2009, MNRAS, 399, 121
Choi, E., Ostriker, J. P., Naab, T., & Johansson, P. H. 2012, ApJ, in press, arXiv:1205.2082
Cisternas, M., et al. 2011, ApJ, 726, 57
Civano, F., et al. 2012, ApJS, in press, arXiv:1205.5030
Conroy, C., Wechsler, R. H., & Kravtsov, A. V. 2006, ApJ, 647, 201
Conroy, C., & White, M. 2013, ApJ, 762, 70
Cox, T. J., Dutta, S. N., Di Matteo, T., Hernquist, L., Hopkins, P. F., Robertson, B., & Springel, V. 2006, ApJ, 650, 791
Cresci, G., et al. 2009, ApJ, 697, 115
Croom, S. M., Boyle, B. J., Shanks, T., Smith, R. J., Miller, L. D., Outram, P. J., Loaring, N. S., Hoyle, F., & da Angela, J. 2005, MNRAS, 356, 415
Croton, D. J. 2009, MNRAS, 394, 1109
da Angela, J., et al. 2008, MNRAS, 383, 565
Dasyra, K. M., Yan, L., Helou, G., Surace, J., Sajina, A., & Colbert, J. E. B. 2008, ApJ, 680, 232
Dasyra, K. M., et al. 2007, ApJ, 657, 102
Debuhr, J., Quataert, E., Ma, C., & Hopkins, P. 2010, MNRAS, 406, L55
Debuhr, J., Quataert, E., & Ma, C.-P. 2012, MNRAS, 420, 2221
Di Matteo, T., Colberg, J., Springel, V., Hernquist, L., & Sijacki, D. 2008, ApJ, 676, 33
Di Matteo, T., Springel, V., & Hernquist, L. 2005, Nature, 433, 604
Draper, A. R., & Ballantyne, D. R. 2012, ApJ, 751, 72
Dunlop, J. S., McLure, R. J., Kukula, M. J., Baum, S. A., O’Dea, C. P., & Hughes, D. H. 2003, MNRAS, 340, 1095
Erb, D. K. 2008, ApJ, 674, 151
Erb, D. K., Steidel, C. C., Shapley, A. E., Pettini, M., Reddy, N. A., & Adelberger, K. L. 2006, ApJ, 646, 107
Evans, D. A., et al. 2009, ApJ, submitted
Fakhouri, O., & Ma, C.-P. 2008, MNRAS, 386, 577
Fanidakis, N., Baugh, C. M., Benson, A. J., Bower, R. G., Cole, S., Done, C., & Frenk, C. S. 2011, MNRAS, 410, 53
Feoli, A., Mancini, L., Marulli, F., & van den Bergh, S. 2010, General Relativity and Gravitation, 57
Ferrarese, L., & Merritt, D. 2000, ApJL, 539, L9
Fisher, D. B. 2006, ApJL, 642, L17
Fisher, D. B., & Drory, N. 2008, AJ, 136, 773
Floyd, D. J. E., Kukula, M. J., Dunlop, J. S., McLure, R. J., Miller, L., Percival, W. J., Baum, S. A., & O’Dea, C. P. 2004, MNRAS, 355, 196
Fontana, A., et al. 2006, A&A, 459, 745
Forster Schreiber, N. M., et al. 2009, ApJ, 706, 1364
Gabor, J. M., et al. 2009, ApJ, 691, 705
Gadotti, D. A. 2009, MNRAS, 393, 1531
Gebhardt, K., et al. 2000, ApJL, 539, L13
Genzel, R., Hollenbach, D., & Townes, C. H. 1994, Reports on Progress in Physics, 57, 417
Georgakakis, A., et al. 2009, MNRAS, 397, 623
Governato, F., et al. 2009, MNRAS, 398, 312
Guyon, O., Sanders, D. B., & Stockton, A. 2006, ApJS, 166, 89
Hasinger, G., Miyaji, T., & Schmidt, M. 2005, A&A, 441, 417
Hayward, C. C., Narayanan, D., Kereš, D., Jonsson, P., Hop-
