Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY
MNRAS **445,** 823–834 (2014) doi:10.1093/mnras/stu1736

Do we expect most AGN to live in discs?

Philip F. Hopkins,[1]★ Dale D. Kocevski[2] and Kevin Bundy[3]

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

Accepted 2014 August 22. Received 2014 August 21; in original form 2013 September 15

ABSTRACT

Recent observations have indicated that a large fraction of the low- to intermediate-luminosity AGN population lives in disc-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 fuelling which accounts for both merger and non-merger 'triggering'. In particular, we show that the 'stochastic accretion' model – in which fuelling in disc 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 disc-dominated and bulge-dominated AGN host galaxies as a function of luminosity and redshift. We compile and compare observational constraints from $z \sim 0$ to 2. The models and observations generically show a transition from disc to bulge dominance in hosts near the Seyfert-quasar transition, at all redshifts. 'Stochastic' fuelling dominates AGN by number (dominant at low luminosity), and dominates black hole (BH) growth below the 'knee' in the present-day BH mass function ($\lesssim 10^7\,\mathrm{M}_{\odot}$). However, it accounts for just ~10 per cent of BH mass growth at masses $\gtrsim 10^8\,\mathrm{M}_{\odot}$. In total, fuelling in discy hosts accounts for ~30 per cent 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 fuelling mechanisms.

Key words: galaxies: active – galaxies: evolution – galaxies: formation – 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, Bender & Cornell 2011), velocity dispersion (Ferrarese & Merritt 2000; Gebhardt et al. 2000), and binding energy/potential depth (Aller & Richstone 2007; Hopkins et al. 2007d; Feoli et al. 2011) have fundamental implications for the growth of BHs and – given the Soltan (1982) argument which implies that most BH mass was assembled in luminous quasar phases (e.g. Salucci et al. 1999; Yu & Tremaine 2002; Hopkins, Richards & Hernquist 2007a; Shankar, Weinberg & Miralda-Escudé 2009) – corresponding active galactic nucleus (AGN) activity.

Fuelling the most luminous quasars at a level required to grow the BH significantly involves channelling an entire typical galaxy's supply of gas ($\gtrsim 10^9$–$10^{10}\,\mathrm{M}_{\odot}$) into the central few pc, probably requiring $\sim 10^{11}\,\mathrm{M}_{\odot}$ worth of gas in the central ~100 pc, on a time-scale comparable to the galaxy dynamical time. 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, Sanders & Stockton 2006; Dasyra et al. 2007; Bennert et al. 2008; Silverman et al. 2008; Liu et al. 2009; Veilleux et al. 2009; Letawe, Letawe & Magain 2010; Koss et al. 2010, 2012, and references therein). Such encounters also convert discs into spheroids and further grow the bulge via centrally concentrated gas inflows in a merger-induced starburst (Mihos & Hernquist 1994; Hibbard & Yun 1999; Cox et al. 2006; Naab, Jesseit & Burkert 2006; Robertson et al. 2006a; Hopkins, Cox & Hernquist 2008c; Hopkins et al. 2008a, 2009a,b). As argued in Hopkins et al.

★E-mail: phopkins@caltech.edu

(2007c), Hopkins & Hernquist (2009a), and Snyder, Hopkins & Hernquist (2011a), 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.

Unfortunately, uniquely identifying observational signatures of ongoing mergers in AGN is incredibly difficult and has been controversial for decades. This is because tidal features are extremely faint and further suppressed by surface-brightness dimming (meaning even the most 'obvious' mergers are very easily classified as relaxed galaxies; see e.g. Lotz et al. 2008; Younger et al. 2009a; Puech et al. 2012; Snyder et al. 2013), mergers are rare so control samples and good statistics are difficult, and the models themselves (almost without exception) predict that the gas inflow rates into the nucleus and subsequent AGN duty cycle peak in the *post-merger* phases where the galaxy can easily look like a 'relaxed' bulge down to optical surface brightnesses $\mu \gtrsim 30$ mag arcsec^{-2} (see Di Matteo, Springel & Hernquist 2005; Li et al. 2008; Johansson, Burkert & Naab 2009; Hopkins & Quataert 2010; Hopkins 2011; Snyder et al. 2011b; Hopkins et al. 2012).

Nevertheless, recent observations of AGN host morphologies and colours have suggested that major mergers probably do not fuel most low- and intermediate-luminosity AGN, as a large fraction appear in 'normal' discs (Gabor et al. 2009; Cisternas et al. 2011; Schawinski et al. 2011; Civano et al. 2012; Kocevski et al. 2012; Mullaney et al. 2012; Santini et al. 2012; Treister et al. 2012; Rosario et al. 2013). This should perhaps not be surprising. Unlike a bright quasar, fuelling a Seyfert (bolometric $L < 10^{12}\,\mathrm{L_{\odot}}$ or 4×10^{45} erg s^{-1}) for a typical $\sim 10^7$ yr episode (see Martini 2004) requires a gas supply within the range of just a single or a few giant molecular clouds. 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, Geller & Barton 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' discs (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 multibody interactions with nearby star clusters or other clouds (e.g. Genzel, Hollenbach & Townes 1994). All of these processes do occur in galaxies, and should at least indirectly contribute to AGN fuelling in so far as they help remove angular momentum from dense gas.

Many models for the rates and luminosity functions (LFs) 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 disc processes and the 'traditional' AGN accretion disc) 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 'fuelling' 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 modelling 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 fuelling, 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' fuelling modes can be estimated. Paper I argued that this can predict accurately many observed properties of $z \approx 0$ Seyferts, including their host galaxies, 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; Treister et al. 2012) that there is some characteristic host bulge/BH mass (and corresponding quasar luminosity) below which these more ubiquitous mechanisms dominate AGN fuelling (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 (MF), 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 disc-dominated hosts (i.e. candidates for the 'stochastic' mode, as opposed to post-major merger systems which may not, on average,

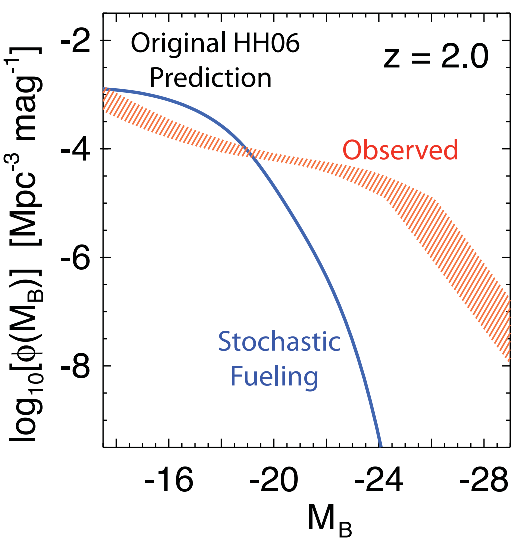

Figure 1. Original predicted $z = 2$ LF for 'stochastically' fuelled AGN from the models of Paper I, compared to the observed LF at the same redshift fit in Ueda et al. (2003). 'Stochastic fuelling' refers to any non-major merger triggered accretion of cold gas by AGN (typically in gas-rich disc-dominated galaxies, as opposed to fuelling associated with a major merger and substantial bulge growth). The Paper I model predicted 'non-major merger' fuelling dominated below luminosities $L_{\rm bol} \approx 4 \times 10^{10}\,\mathrm{L_{\odot}}$ ($M_{\rm B} \gtrsim -19$). However, Kocevski et al. (2012) and others (see Section 1) find disc-dominated hosts dominate the population up to at least a factor $\sim$10 higher-$L_{\rm bol}$ (close to the 'knee' in the LF).

appear as discs)[1] 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 have improved (see the comparisons in Johansson et al. 2009; Debuhr et al. 2010; Choi et al. 2012; Debuhr, Quataert & Ma 2012). We therefore, in this paper, re-visit these models for AGN fuelling, 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/LFs, merger rates, and gas fraction distributions), to define the 'background' on which AGN fuelling 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 versus non-major merger mechanisms towards BH growth.

2 THE MODELS

The model we will present here supposes two independent AGN fuelling 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 behaviour parametrization in fuelling events in both cases, but treat the total AGN LF as simply the sum of the predicted duty cycles from both subpopulations.

2.1 Merger-induced fuelling

The major merger-induced quasar fuelling model here is taken directly from a series of papers: Hopkins et al. (2009c, 2010a,b,c), and Younger & Hopkins (2011). We use the most recent update to the model, presented in Hopkins et al. (2010a, 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 MFs and gas fraction distributions. This defines the empirical 'background population' on to which we will add assumptions for AGN fuelling. 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 fuelling (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 MF data are compiled from a range of sources.[2]

(2) Using a simple abundance-matching halo-occupation model (i.e. forcing the population to match observed number densities and clustering; see Conroy, Wechsler & Kravtsov 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.[3] 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 rates. Extensive discussion and tests of this methodology are presented in Hopkins et al. (2010c); 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. 2009b, Jogee et al. 2009, and Lotz et al. 2011).

(3) For each such 'semi-empirically' assigned merger, we then attach an AGN fuelling model. Specifically, in a series of papers, Hopkins et al. (2006a,b, 2007c,d) use a simple model for AGN accretion rates and feedback to fit the resulting AGN light curves 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) light curve (or equivalently, probability of being seen at a given luminosity) to every merger. The exact functional parametrization is given in Paper II; this is compared to observations (from Yu, Lu & Kauffmann 2005; Kauffmann & Heckman 2009) 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 parametrization we adopt.

[1] It is important to note that even major galaxy mergers can and do leave disc-dominated remnants under the right circumstances (when they are sufficiently gas rich and have favourable initial orbital parameters; see Springel & Hernquist 2005; Robertson et al. 2006b; Hopkins et al. 2009d). However, if major mergers were the dominant AGN fuelling 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. 2009c). This is not observed except at higher AGN luminosities, as we will discuss further in the text.

[2] MF measurements are compiled from Bell et al. (2003), Arnouts et al. (2007), Ilbert et al. (2010), Pérez-González et al. (2008), Fontana et al. (2006), Marchesini et al. (2009), and 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), and Erb (2008). For details, see Paper II.

[3] The specific results here use the halo mergers from the Millennium simulation in Fakhouri & Ma (2008). However, in Hopkins et al. (2010c) we compare this to a wide variety of other simulations with varied numerical methods, cosmological parameters, and post-processing method for halo and merger identification; we use this to define a 'theoretical uncertainty' in the halo merger rate. In the model here, this is added in quadrature to the 'empirical uncertainty' in the number density of galaxies, to define the total uncertainty in the final merger rate. These uncertainties and comparison of the predicted rates to observations are in Hopkins et al. (2010c).

We stress that we are not presenting any modifications or revisions to this model; we take the predicted 'merger-induced' AGN LFs 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 fuelling

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 fuelling' category. In Paper I, however, this is 'synthesized' into an estimate of the resulting LF 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' on to 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 MFs and gas fraction distributions from Paper II, identical to the first step in Section 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 fuelling 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 model, at each redshift, according to the simple approximate observed relation:

$$M_{\rm BH} \approx 0.0014\,(1+z)^{0.5}\, f_{\rm bulge}\, M_{*} \quad (1)$$

with a lognormal intrinsic scatter of ≈0.3 dex in $M_{\rm BH}$. This is a purely empirical estimate of a best fit to a range of observations (McLure & Dunlop 2004; Adelberger & Steidel 2005a; Peng et al. 2006; Shields et al. 2006; Woo et al. 2006; Salviander et al. 2007; Treu et al. 2007; Bennert et al. 2010; for a recent review see Kormendy & 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. 2007c; 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 and 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 $L_{\rm bol}$, but not much alter their *relative* behaviour, 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 MF, corresponding to luminosities well above the turnover in the LF; see Paper II fig. B3), but we will show that the stochastic mode is subdominant in this regime in any case (so assuming any scatter $\lesssim$1 dex makes little difference). Finally, $f_{\rm bulge}$ is estimated from our galaxy MFs, but is formally degenerate with the normalization and redshift evolution of the relation; where (at high redshifts) it is poorly determined we simply assume $f_{\rm bulge} \approx 0.3$, 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 time-scales), 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 parametrized as

$$\frac{{\rm d}P}{{\rm d}\log L} = \alpha\left(\frac{f_{\rm gas}}{0.1}\right)\left(\frac{L}{L_{\rm Edd}(M_{\rm BH})}\right)^{-\beta} \quad (2)$$

with $\alpha \approx 0.003$ and $\beta \approx 0.6$ (see Yu, Lu & Kauffmann 2005; Bonoli et al. 2009; Hickox et al. 2009; Kauffmann & Heckman 2009; Kelly, Vestergaard & Fan 2009; Trump et al. 2009; Shankar et al. 2010, as well as Paper I). Here, the parameter α physically represents the duty cycle at high accretion rates, given by the rate at which a collection of Jeans-length sized clouds (on isotropic, virial equilibrium orbits) in a galactic nucleus would intersect the BH (then multiplied by the time required to accrete the captured mass, i.e. $\alpha \sim n_{\rm clouds}\,\pi\,R_{\rm cloud}^2\,\sigma_{v,\,{\rm cloud}}\,(\epsilon_r\,M_{\rm cloud}\,c^2/L)$). The parameter β represents the relative amount of time at each accretion rate – for a simple power-law light curve this corresponds to $L \propto t^{-1/\beta}$, with the value here representing the typical behaviour of this 'decay' in each accretion event in simulations. We truncate equation (2) at $L > L_{\rm Edd}(M_{\rm BH}) \approx 3.3 \times 10^4\,({\rm L}_{\odot}/{\rm M}_{\odot})\,M_{\rm BH}$. Note that in Hopkins & Hernquist (2009b), this is compared to an extensive ensemble of observational constraints and measurements of the Eddington ratio distribution at $z \approx 0$–1, and shown to agree well (especially for moderate luminosity AGN), with relatively little allowed range in α or β relative to the theoretically predicted values. Therefore, if we simply adopted a best fit to the observed $L/L_{\rm Edd}$ distribution at $z = 0$, we would obtain a nearly identical prediction.[4] This duty cycle is simply convolved with the BH MF 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 fuelling and feedback model) correspond to a very different distribution of galaxy masses and gas fractions, from that which we develop here. The most important differences are as follows: (1) observations of high-redshift galaxies indicate that they are more gas-rich than assumed in Paper I, with gas fractions approaching ∼50 per cent 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 underpredict the number density of galaxies with stellar masses $\gtrsim 10^{11}\,{\rm M}_{\odot}$ 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 LFs from the models for both major merger-induced and 'stochastic' (non-major merger) mechanisms,

[4] For simplicity, in the plots in this paper we use exactly the formula given in equation (2); in Paper I a more complicated convolution is presented, to which this is a simplified fitting function. Repeating all of our calculations with the more detailed formulation makes a completely negligible difference to the predictions. Also we should formally introduce a lower limit to equation (2), which corresponds to the luminosity above which the duty cycle integrates to unity. However, this occurs at such low luminosities that it is irrelevant to any of our calculations.

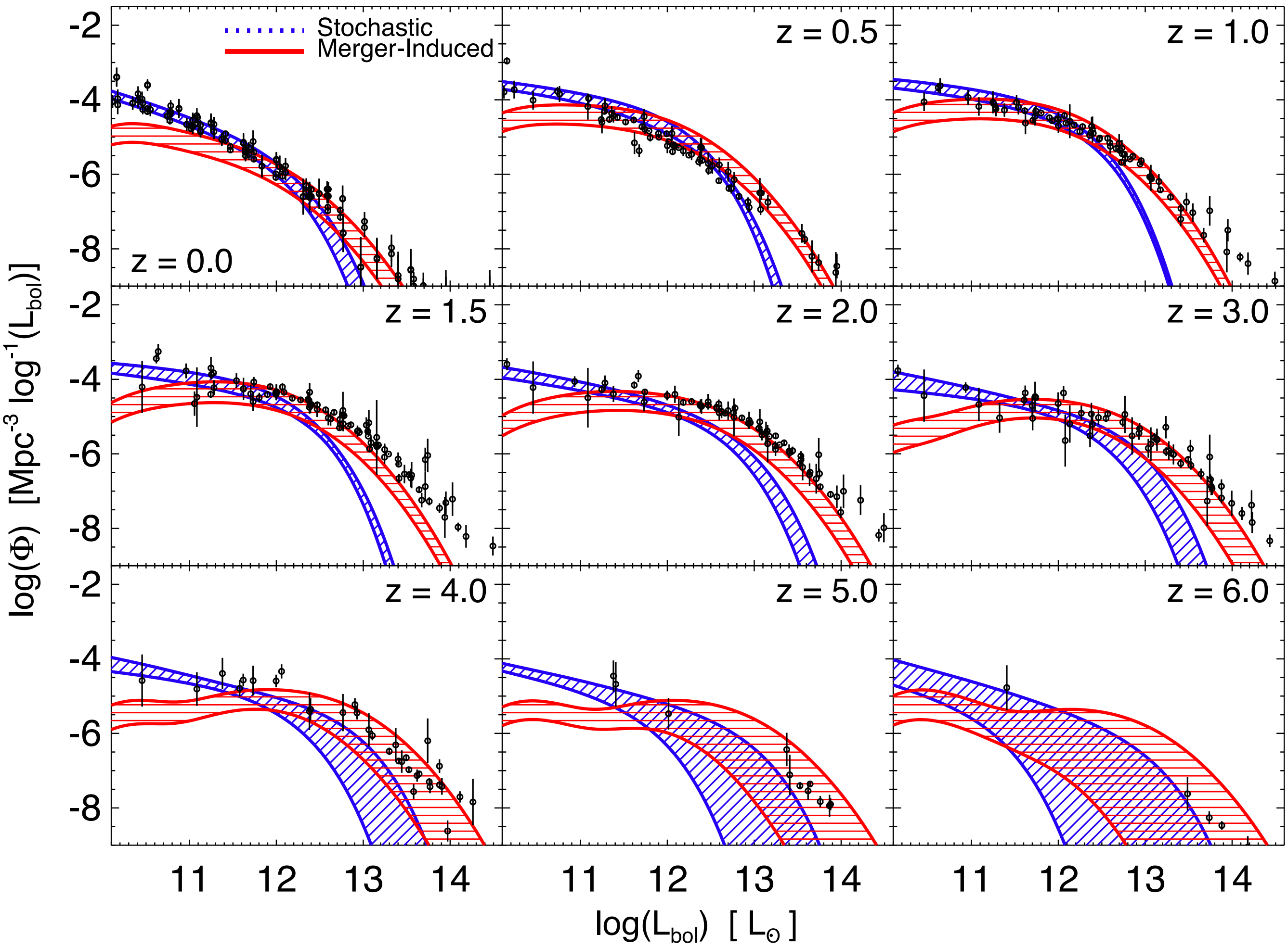

Figure 2. Bolometric AGN LFs as a function of redshift. We show the predicted LF of major merger-induced AGN from Paper II (red), and LF of non-major merger 'stochastically' fuelled systems from Paper I (blue), with updated observational inputs (stellar MFs and gas fractions used to construct the model) matching those from Paper II. Shaded ranges reflect the uncertainty from different stellar MF observations used in constructing the model. Black points show the compilation of observational data used to derive bolometric AGN LFs in Hopkins et al. (2007a).

at several redshifts. As discussed in Section 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 *not* to differences in the AGN fuelling models (which might be quite large), but owing to unrelated uncertainties in the background galaxy population.

Of course, there are significant uncertainties in the AGN models themselves. However – within the context of the models we use here – these actually contribute surprisingly little to the 'total' prediction uncertainties. Consider the merger model: a major merger converts an order-unity fraction of the galaxy stellar mass into a bulge, and in any model we consider that the final BH must lie near the appropriate $M_{\rm BH}(M_{\rm bulge})$ relation (at that redshift). Since the fuelling time-scale is short (inflow rates are large), but we cap accretion at the Eddington limit, the contribution to the LF from a merger is dominated by the last Salpeter time near the Eddington limit for $M_{\rm BH}(M_{\rm bulge})$. Thus, if the Eddington limit applies, the uncertainty in the 'merger mode' duty cycle is dominated by the merger rate, and the uncertainty in the 'merger mode' characteristic luminosity by the background population (mass, bulge-to-disc, and $M_{\rm BH}(M_{\rm bulge})$ distributions). For explicit demonstrations of this, we refer to Hopkins et al. (2005, 2006a), Merloni & Heinz (2008), Shankar et al. (2010), Bonoli et al. (2009), and Somerville (2009). Of course we could, in principle, assume that post-merger systems lie far off the relation for some reason, but we will not consider such models here. For the 'stochastic mode', the assumed duty cycle (α in equation 2) simply enters linearly in the predicted AGN LF normalization. Theoretically, it is highly uncertain, but empirically, there is not much room for it to vary without violating the constraint that we must match the abundance of low-luminosity AGN. Changing the prediction at high-L without violating this constraint would require a *qualitative* change in the model, namely invoking a duty cycle which increases substantially (by a factor of ∼100 over the dynamic range plotted) in more massive, bulge-dominated, and gas-poor galaxies – the opposite of what is seen in simulations and numerical models of secular evolution. That is, not to say there are no uncertainties, just that the uncertainties in the parameters of these models do not have a large effect on our conclusions; we discuss more radical alternative models below.

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 LF presented in Hopkins et al. (2007a).[5]

The sum of the stochastically fuelled 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* fuelling mechanism or driver is not needed to explain the observed demographics.[6]

Clearly, stochastically fuelled 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 $\sim 10^{12}\,\mathrm{L}_{\odot}$ (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 LF. 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 fuelling mechanisms: by allowing very different AGN light curves, or including minor mergers (and assuming they produce a long duty cycle of low-luminosity activity), it is possible to fit the low-luminosity LF with *only* merger-induced fuelling (see the models in Hopkins et al. 2005, 2006b; Somerville 2009). On the other hand, by assuming a much stronger 'secular' mode (in which traditional disc bar instabilities are assumed to channel 100 per cent of the galaxy gas into the nucleus in a single burst – essentially mimicking a major merger), Fanidakis et al. (2011) show they can plausibly reproduce the high-luminosity LF. And at high redshifts and high-L_{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 bight, high-redshift LF has little power to constrain fuelling models.

In Fig. 3, 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_{bol} that are predicted to be fuelling 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 $\gg 10^{12}\,\mathrm{L}_{\odot}$. There is some quantitative increase in prevalence of mergers at intermediate luminosities at high redshifts, but the effect is small.

Very crudely, most 'stochastically fuelling' systems should be disc dominated. To lowest order, this is simply a reflection of the background galaxy population (which, at lower masses where the fuelling mode is dominant, is mostly disc dominated). At second-

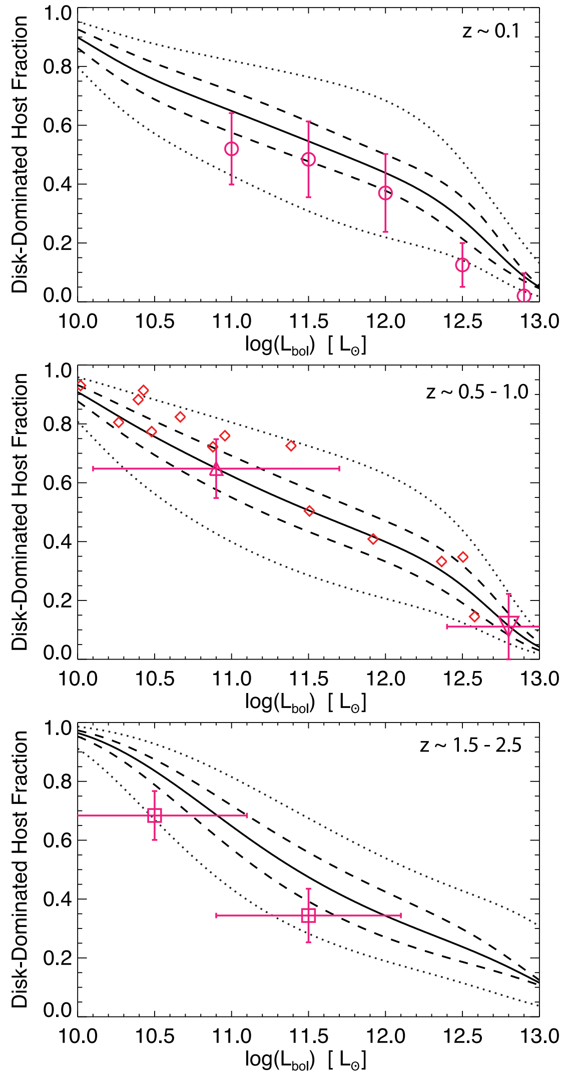

Figure 3. Predicted fraction of AGN in the stochastic mode, as a function of luminosity and redshift, from the models in Fig. 2. The mean of the model range is shown as the solid line with the $\pm 1\sigma$ (dashed) and ($\pm 2\sigma$) range (dotted). Since the duty cycle of 'stochastically' fuelled systems is dominated by gas-rich discs, and most major merger-triggered systems that induce strong bulge and BH growth will appear as spheroids, we compare the observed fraction of disc-dominated AGN host galaxies in different luminosity/redshift intervals. At low redshift, measured from the PG quasar sample of Dunlop et al. (2003) and Floyd et al. (2004, circles). At intermediate redshift, measured in low-luminosity AGN in COSMOS in Gabor et al. (2009) and Cisternas et al. (2011, triangle), and in true (type 2) quasars in Zakamska et al. (2006, 2008), Zakamska et al. (2008), and Liu et al. (2009, inverted triangle). At $z \sim 2$, measured in CANDELS AGN host galaxies in the low- and high-luminosity subsamples from Kocevski et al. (2012, squares). Treister et al. (2012) independently considered a (partially overlapping) compilation which agrees well with ours (red diamonds in centre show $1 - f_{\mathrm{m}}$, where f_{m} is the fraction of AGN they specifically identify as merger induced).

[5] This is based on a compilation of observations at wavelengths from the IR through optical, soft and hard X-rays (see e.g. Ueda et al. 2003; Hunt et al. 2004; Barger et al. 2005; Hasinger, Miyaji & Schmidt 2005; La Franca et al. 2005; Richards et al. 2005; Brown et al. 2006; Richards et al. 2006; Siana et al. 2008). An alternative but similar bolometric compilation is presented in Shankar et al. (2009), and bolometric LFs from hard X-ray LFs with appropriate corrections are in Aird et al. (2010), and Yencho et al. (2009); the differences are generally smaller than the model uncertainties in Fig. 2. Additional observations have been developed since these papers; however, they generally overlap with the plotted points except at the highest redshifts ($z \gtrsim 5$) where they extend the dynamic range significantly (see McGreer et al. 2013). However, at these redshifts the newer data lie well within the (very large) model uncertainties.

[6] At $z \sim 1$–3, the total LF at the very highest luminosities $L_{\mathrm{bol}} \gtrsim 10^{14}\,\mathrm{L}_{\odot}$ does appear to fall short of the bolometric LF estimates. This is discussed in Paper II (since the predictions are dominated by the merger-induced contribution at these luminosities). It is certainly worth considering that this owes to a deficiency in the AGN fuelling/light-curve models. However, we caution against reading too much into the discrepancy. These are extreme populations with number densities of just a few per cubic Gpc, and so systematic uncertainties in e.g. the relevant bolometric corrections and contributions from lensing are very large.

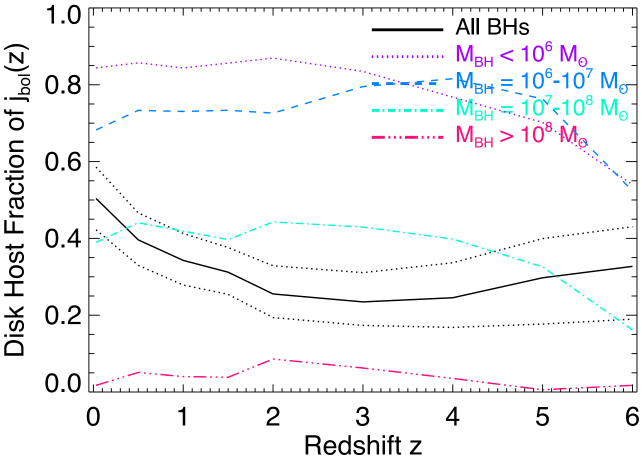

Figure 4. Fractional contribution of 'stochastic' fuelling to the AGN luminosity density/integrated BH accretion as a function of redshift. We show the total contribution (solid black), with the uncertainties (within the context of the model) from Fig. 3 (dotted black); we also compare the contribution to BH growth in different intervals in BH mass at each redshift (coloured, as labelled). The model here is the best fit to both the bolometric LF and the observed disc/spheroid fractions at each L, z. The best-fitting model predicts ~30 per cent of the luminosity density from non-major merger fuelling modes (increasing at the lowest redshifts). The non-merger modes completely dominate at low BH masses $\ll 10^7\ \mathrm{M}_\odot$, while merger modes dominate at high BH masses $\gg 10^7\ \mathrm{M}_\odot$.

order, at fixed mass, in the model we adopt (Section 2) AGN activity does require a gas supply, so fuelling is enhanced in gas-rich systems, which are overwhelmingly disc dominated (though of course there will be some, albeit rarer, gas-rich spheroids). In contrast, most major merger-fuelled systems should be bulge dominated, since such mergers tend to build large bulges.[7]

We therefore compare the predicted 'stochastically fuelled' fraction of AGN with the fraction of disc-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-fitting bulge-to-total mass ratio $B/T < 0.5$ in bins of $L_{\rm bol}$, 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 Gabor et al. (2009) and Cisternas et al. (2011); we plot the 'final' quoted fraction of disc-dominated galaxies in the sample (with the approximate ~10 per cent systematic difference between classifiers quoted therein) and 90 per cent range of $L_{\rm bol}$ estimated from the hard X-ray luminosities. We also compare the sample of true quasars in

[7] As noted in Section 1, we stress that a sizeable fraction of major mergers will produce disc-dominated galaxies, especially at high-redshifts where the discs are more gas rich (see e.g. Springel & Hernquist 2005; Robertson et al. 2006b; Hopkins et al. 2008b, 2009d, 2013; Governato et al. 2009). However, disc survival in mergers is most efficient at low galaxy masses, where the discs are actually gas dominated; the large BH masses where the merger-induced mode is dominant imply bulge masses $\gtrsim 10^{11}\ \mathrm{M}_\odot$. Moreover large surviving discs 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 fuelling will result. As a result, this can be critical for the abundance of discs at low masses (Somerville et al. 2008; Hopkins et al. 2009c; Stewart et al. 2009a; Puech et al. 2012), but is probably not the dominant process in the mergers that produce bright quasar activity, of interest here.

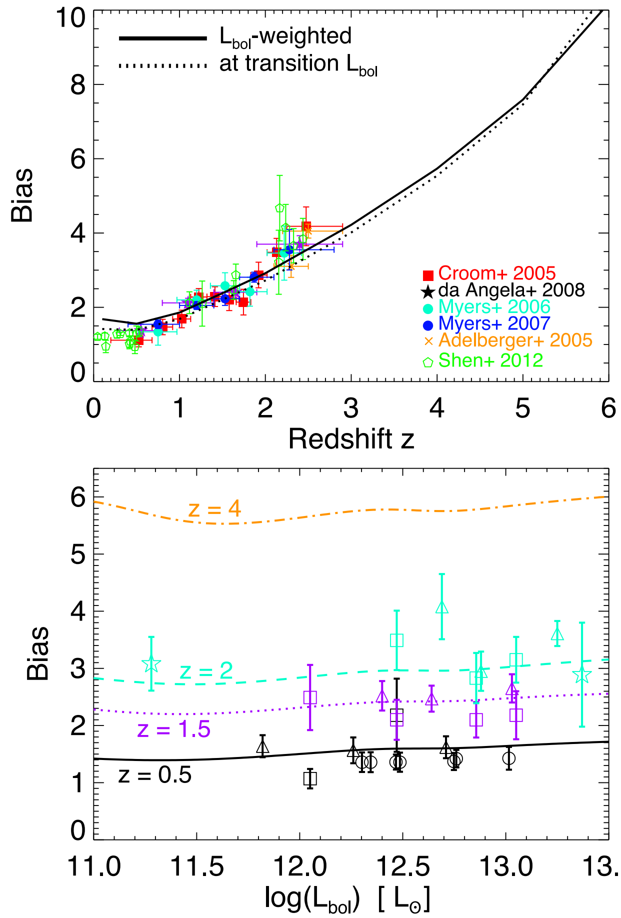

Figure 5. Predicted clustering amplitude (linear bias) of AGN populations from the model here. Top: mean bias as a function of redshift. We plot the luminosity-density-weighted bias (integrated over the LF; solid), and the mean bias at the 'transition' luminosity where the contribution from stochastic and merger fuelling is equal (dotted). We compare to compiled observations of quasar clustering from Hopkins et al. (2007b) and Shen et al. (2013); see footnote 9. The two agree well, with bias similar to ~1–4 × $10^{12}\ \mathrm{M}_\odot$ haloes at each redshift. Bottom: mean bias as a function of luminosity, at fixed redshifts (specific values shown). We compare observations in narrow luminosity intervals at approximately the same redshifts (denoted by the same colours), from Shen et al. (2013, circles), da Angela et al. (2008, squares), Myers et al. (2007a, triangles), and Adelberger & Steidel (2005b, stars). The clustering amplitude predicted is a very weak function of luminosity at all redshifts, in agreement with the observations. In particular, there is no feature or trend marking the 'transition' in Fig. 3.

Zakamska et al. (2006, 2008) and 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 disc 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 ($L_{\rm bol}$ estimated here from the hard X-ray luminosities). Note that Treister et al. (2012) consider a similar compilation (partially overlapping the sources we have compiled here); the results from their compilation (using different data sets, bolometric corrections, and classification criteria) agree extremely well with what we plot in Fig. 3 at each redshift.

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 (colour, 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.

Fig. 4 plots the fractional contribution from the 'stochastic' mode (predicted from the model), integrated over the LF, to BH growth in different mass intervals and different redshifts.

In Fig. 5, 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, Mo & Tormen (2001), and use this to calculate the mean bias of the population in bins of AGN luminosity and redshift.[8] We show the mean bias 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. (2007b) and Shen et al. (2013).[9]

The agreement with observations is good. Unfortunately, it appears that clustering is not a strong constraint on models of AGN fuelling 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 fuelling (Lidz et al. 2006; Hopkins et al. 2007b), only secular (non-merger) fuelling (Fanidakis et al. 2011, Croton et al., in preparation), or which make no statement about fuelling 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 light curves are calibrated fall closely on these relations; Di Matteo et al. 2005; Hopkins et al. 2007c), it is automatically implicit that they also reproduce the local BH MF. An explicit calculation and comparison with the MF 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 & Tremaine 2002; Yu & Lu 2004; Merloni & Heinz 2008) shows consistency between the quasar LF and BH MF. Therefore, this also has little power to constrain fuelling models.

[8] For the clustering calculation, we adopt the *WMAP*5 cosmological parameters. However, within reasonable uncertainties this only has a small systematic effect on the normalization of the bias in Fig. 5.

[9] Hopkins et al. (2007b) compile the observations from Croom et al. (2005), Adelberger & Steidel (2005b), Myers et al. (2006), Myers et al. (2007a), Porciani & Norberg (2006), and da Angela et al. (2008). Shen et al. (2013) compile the results from Shen et al. (2009), Hickox et al. (2009), Cappelluti et al. (2010), Hickox et al. (2011), White et al. (2012), and Krumpe et al. (2012). The measurements of clustering amplitude as a function of luminosity at fixed redshift are compiled from Shen et al. (2013, circles at $z \approx 0.5$), da Angela 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 (2005b, stars at $z \approx 2.0$).

4 DISCUSSION

4.1 Overview

This paper presents a simple 'semi-empirical' population synthesis model for AGN fuelling that distinguishes between major merger-triggered and non-major merger-triggered ('stochastic') activity. We show that this can plausibly account for the bolometric AGN LF from $z = 0$ to 6 and $L_{\rm bol} \sim 10^{10}$ to $10^{14}\,{\rm L_{\odot}}$, 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 fuelling/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 centre in the first place – it can be contributed or torqued by minor mergers, disc instabilities (bars, spiral arms, massive clumps), directly fuelled 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 Draper & Ballantyne (2012), using an independent BH population synthesis approach with very different methods used to model the merger and non-merger triggering rates. In short, the models predict that 'stochastic' fuelling, with no specific preference for large-scale 'triggering phenomena' in discy, secularly evolving systems should dominate the population at Seyfert and lower luminosities, while mergers dominate fuelling of bright quasars. As argued in Bellovary et al. (2013), this means that no new 'direct' large-scale fuelling mechanisms (such as cold flows somehow penetrating directly to the BH) need to exist at high redshift – and in fact, there is little room for such mechanisms in this model. And recent observations in e.g. Treister et al. (2012, and references therein) may have begun to map this transition – the authors there see a strikingly similar trend with luminosity to that predicted here, with only a weak secondary dependence on redshift.

4.2 The role of stochastic fuelling as a function of mass/luminosity

Quantitatively, if we integrate the models here, we estimate that non-major merger AGN contribute about ∼30 per cent 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 $<10^6\,{\rm M_{\odot}}$ and most of the BH mass at $<10^7\,{\rm M_{\odot}}$ (at all redshifts) accreted in the 'stochastic' mode. But above $M_{\rm BH} \gtrsim 10^7\,{\rm M_{\odot}}$, most of the mass is accreted in the merger-induced mode. As argued in Section 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 $\lesssim 10$ 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.

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 discs) dominates, but the highest star formation rate systems are essentially all major mergers. At low redshifts, this has been well known for ~20 yr (with the transition occurring at IR luminosities of ULIRGs, see e.g. Joseph & Wright 1985; Sanders & Mirabel 1996). 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 $z \sim 0$ to 2 (see e.g. Dasyra et al. 2008; Melbourne et al. 2008; Tacconi et al. 2008; Casey et al. 2009; Younger et al. 2009b; Zamojski et al. 2011; Kartaltepe et al. 2012; Sargent et al. 2012).

However, there are two critical differences between the star-forming and AGN populations. First, the 'transition luminosity' $L_{\rm SF}$ 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 $L_{\rm bol}$ is much weaker (nearly constant at $10^{12}\,L_\odot$, in the model here). The rapid evolution in $L_{\rm SF}$ 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 superlinearly according to the (Kennicutt 1998) relation.[10] However, in most models, the maximum AGN $L_{\rm bol}$ 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. 2007c; Johansson et al. 2009). Since the mass at the 'break' in the galaxy stellar MF (hence implied BH masses, if the BH-host correlations still apply) does not evolve very strongly from $z \sim 0$ to 2, this implies that the AGN 'transition' $L_{\rm bol}$ should be more constant than the star formation transition $L_{\rm SF}$.

Secondly, 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 Section 4.2 above; for reviews, see Kormendy & Kennicutt 2004; Balcells, Graham & Peletier 2007; Fisher & Drory 2008; Gadotti 2009; Hopkins & Hernquist 2009a; Kormendy & Bender 2012). 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 disc 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 (~10 per cent in an $\sim L_*$ spheroid) of the final stellar mass is actually formed in a nuclear starburst 'driven by' the merger (for a rigorous discussion, see Hernquist, Spergel & Heyl 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 centre of discs). 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 discs). 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; Santini et al. 2012). 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 disc) properties, especially at higher redshifts (for a recent review, see Kormendy & Ho 2013).

4.4 Observational predictions

We have compared some of the lowest order predictions of this model to observations, and reach a couple of important conclusions. First, the LF itself, and the clustering (or environments) of AGN are *not* sensitive probes of the AGN fuelling mechanism (see also Bonoli et al. 2009). Even when our model predicts a sharp transition in the fuelling mode as a function of luminosity, no signature appears in these data. But we do predict a strong trend of the 'post-merger' (bulge-like) versus disc population of AGN hosts as a function of luminosity, as discussed above. This is primarily a function of luminosity, and only weakly depends on redshift. At the moment, the statistical evidence for this is tantalizing, but not strong – larger samples of systems with reliable morphological classifications, especially at high luminosities and redshifts, would be tremendously useful.

If this trend of bulge versus disc-like AGN host morphologies is, in fact, due to the role of merger-induced fuelling as we have proposed, there should be a number of corollaries. The next step would be to examine the bulge-dominated hosts and look for evidence that they are not simply uniformly subsampling the 'normal' bulge-dominated galaxy population. Indeed, there are already a number of additional, indirect observational suggestions that there is a transition from essentially random fuelling of AGN at Seyfert luminosities to merger-induced fuelling in true quasars; some of these are summarized in Hopkins & Hernquist (2009a). This includes the fact that quasars exhibit excessive small-scale (subhalo-scale) clustering while Seyferts do not (Serber et al. 2006; Myers et al. 2007b; Hennawi et al. 2009; Shen et al. 2010); future observations

[10] This is the dominant effect driving evolution in the Paper II models for the IR LFs of star-forming galaxies, which appear to accurately describe the evolution in $L_{\rm SF}$.

should confirm that this difference appears even considering bulge-dominated galaxies. Quasar duty cycles rise more sharply with redshift (in agreement with observed merger rates), as opposed to Seyfert duty cycles which increase more slowly more or less in agreement with galaxy gas fraction evolution (see the compilation in Hopkins & Hernquist 2009a and discussion in Draper & Ballantyne 2012); this is qualitatively *opposite* the trend in galaxy bulge-to-disc ratios (galaxies become less bulge dominated, even at fixed mass, at high redshift) – it is therefore very difficult to predict a trend purely from secular or stochastic fuelling in which the ratio of disc-to-bulge dominance is nearly redshift independent (as we predict here). We also expect a much larger prevalence of 'post-starburst' (or recently star-forming, K+A or E+A type) populations in true quasars compared specifically to normal *bulges* (not discs) of the same mass, as many observations have suggested (Brotherton et al. 1999; Kewley et al. 2006; Vanden Berk et al. 2006; Nandra et al. 2007; Shi et al. 2007; Higdon et al. 2008; Lutz et al. 2008; Silverman et al. 2008; Wang et al. 2008; Glikman et al. 2012; Cales et al. 2013).

The most naively obvious prediction to search for in the observations of these systems is evidence of tidal interactions, double nuclei, and other morphological merger signatures; in future work, we will make specific predictions for the prevalence of these signatures from the simulations used to inform the models here. However, as we cautioned in Section 1, most models predict the quasars appear in the late stages of the merger (or even a few hundred million years after the coalescence of the galactic nuclei), since it takes many dynamical times for material to make its way to the centre of the galaxy. And indeed, recent observations of local mergers in different stages have confirmed that the incidence of AGN is highly biased towards the post-merger phases (during which almost all commonly used morphological classifiers at high redshift would identify the galaxies as 'normal'; see e.g. Satyapal et al. 2014, and references therein). This is because the time-scale for tidal features, asymmetries, and other features to relax out to beyond the galaxy effective radius is relatively short; thus even most quasars triggered by major mergers are predicted to only have incredibly faint tidal features. For example, there are number of known nearby systems which are <0.5 Gyr post-merger and have tidal features of surface brightness $\mu \gtrsim 28$–30 mag arcsec^{-2} at $z = 0$ (see Schweizer 1998; Schweizer & Seitzer 2007); Bennert et al. (2008) and Canalizo & Stockton (2013) have identified a number of analogous cases among the nearest quasars (and indeed all of these cases were initially misclassified based on early *Hubble Space Telescope* images as 'relaxed' bulges). So it is almost impossible, unfortunately, for present morphological observations of bulges at $z > 0$ to rule *out* late-stage or post-merger fuelling.

On the other hand, a particularly compelling corollary of our model 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 'pseudo-bulges', 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; Fisher 2006; Balcells et al. 2007; Fisher & Drory 2008; Gadotti 2009; Hopkins & Hernquist 2009a; Kormendy & Bender 2012). Verifying that this holds for the AGN hosts themselves (rather than simply for their 'relics' at $z = 0$, would represent a direct and very powerful confirmation of the models. Thus far, this has only been confirmed for the very most local of samples, the PG quasars (see Dunlop et al. 2003; Floyd et al. 2004). If bright quasar hosts at high redshift were instead all pseudo-bulges, a new form of fuelling beyond what we model here may be required.

ACKNOWLEDGEMENTS

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, 627, L1
Adelberger K. L., Steidel C. C., 2005b, ApJ, 630, 50
Aird J. et al., 2010, MNRAS, 401, 2531A
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
Arnouts 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, 779, 136
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, ApJ, 741, L33
Brotherton M. S. et al., 1999, ApJ, 520, L87
Brown M. J. I. et al., 2006, ApJ, 638, 88
Cales S. L. et al., 2013, ApJ, 762, 90
Calura F., Jimenez R., Panter B., Matteucci F., Heavens A. F., 2008, ApJ, 682, 252
Canalizo G., Stockton A., 2013, ApJ, 772, 132
Cappelluti N., Ajello M., Burlon D., Krumpe M., Miyaji T., Bonoli S., Greiner J., 2010, ApJ, 716, L209
Casey C. M. et al., 2009, MNRAS, 399, 121
Choi E., Ostriker J. P., Naab T., Johansson P. H., 2012, ApJ, 754, 125
Cisternas M. et al., 2011, ApJ, 726, 57
Civano F. et al., 2012, ApJS, 201, 30
Conroy C., White M., 2013, ApJ, 762, 70
Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201
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. et al., 2005, MNRAS, 356, 415
Croton D. J., 2009, MNRAS, 394, 1109
da Angela J. et al., 2008, MNRAS, 383, 565
Dasyra K. M. et al., 2007, ApJ, 657, 102
Dasyra K. M., Yan L., Helou G., Surace J., Sajina A., Colbert J., 2008, ApJ, 680, 232
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., Springel V., Hernquist L., 2005, Nature, 433, 604

Di Matteo T., Colberg J., Springel V., Hernquist L., Sijacki D., 2008, ApJ, 676, 33
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
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., 2011, Gen Relativ Gravit, 43, 1007F
Ferrarese L., Merritt D., 2000, ApJ, 539, L9
Fisher D. B., 2006, ApJ, 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, ApJ, 539, L13
Genzel R., Hollenbach D., Townes C. H., 1994, Rep. Prog. Phys., 57, 417
Georgakakis A. et al., 2009, MNRAS, 397, 623
Glikman E. et al., 2012, ApJ, 757, 51
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., Hopkins P. F., Cox T. J., Hernquist L., 2012, MNRAS, 428, 2529
Hennawi J. F. et al., 2009, ApJ, 719, 1672
Hernquist L., Mihos J. C., 1995, ApJ, 448, 41
Hernquist L., Spergel D. N., Heyl J. S., 1993, ApJ, 416, 415
Hibbard J. E., Yun M. S., 1999, ApJ, 522, L93
Hickox R. C. et al., 2009, ApJ, 696, 891
Hickox R. C. et al., 2011, ApJ, 731, 117
Higdon J. L., Higdon S. J. U., Willner S. P., Brown M. J., Stern D., Le Floc'h E., Eisenhardt P., 2008, preprint (arXiv:0806.2138)
Hopkins P. F., 2011, MNRAS, 420, L8
Hopkins P. F., Hernquist L., 2006, ApJS, 166, 1 (Paper I)
Hopkins P. F., Hernquist L., 2009a, ApJ, 694, 599
Hopkins P. F., Hernquist L., 2009b, ApJ, 698, 1550
Hopkins P. F., Hernquist L., 2010, MNRAS, 402, 985
Hopkins P. F., Quataert E., 2010, MNRAS, 407, 1529
Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Robertson B., Springel V., 2005, ApJ, 630, 716
Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Robertson B., Springel V., 2006a, ApJS, 163, 1
Hopkins P. F., Somerville R. S., Hernquist L., Cox T. J., Robertson B., Li Y., 2006b, ApJ, 652, 864
Hopkins P. F., Richards G. T., Hernquist L., 2007a, ApJ, 654, 731
Hopkins P. F., Lidz A., Hernquist L., Coil A. L., Myers A. D., Cox T. J., Spergel D. N., 2007b, ApJ, 662, 110
Hopkins P. F., Hernquist L., Cox T. J., Robertson B., Krause E., 2007c, ApJ, 669, 45
Hopkins P. F., Hernquist L., Cox T. J., Robertson B., Krause E., 2007d, ApJ, 669, 67
Hopkins P. F., Hernquist L., Cox T. J., Dutta S. N., Rothberg B., 2008a, ApJ, 679, 156
Hopkins P. F., Hernquist L., Cox T. J., Younger J. D., Besla G., 2008b, ApJ, 688, 757
Hopkins P. F., Cox T. J., Hernquist L., 2008c, ApJ, 689, 17
Hopkins P. F., Cox T. J., Dutta S. N., Hernquist L., Kormendy J., Lauer T. R., 2009a, ApJS, 181, 135
Hopkins P. F., Lauer T. R., Cox T. J., Hernquist L., Kormendy J., 2009b, ApJS, 181, 486
Hopkins P. F. et al., 2009c, MNRAS, 397, 802
Hopkins P. F., Cox T. J., Younger J. D., Hernquist L., 2009d, ApJ, 691, 1168
Hopkins P. F., Younger J. D., Hayward C. C., Narayanan D., Hernquist L., 2010a, MNRAS, 402, 1693 (Paper II)
Hopkins P. F. et al., 2010b, ApJ, 715, 202
Hopkins P. F. et al., 2010c, ApJ, 724, 915
Hopkins P. F., Hernquist L., Hayward C. C., Narayanan D., 2012, MNRAS, 425, 1121
Hopkins P. F., Cox T. J., Hernquist L., Narayanan D., Hayward C. C., Murray N., 2013, MNRAS, 430, 1901
Hunt M. P., Steidel C. C., Adelberger K. L., Shapley A. E., 2004, ApJ, 605, 625
Ilbert O. et al., 2010, ApJ, 709, 644
Jogee S., 2006, in Alloin D., ed., Lecture Notes in Physics, Vol. 693, Physics of Active Galactic Nuclei at all Scales. Springer-Verlag, Berlin, p. 143
Jogee S. et al., 2009, ApJ, 697, 1971
Johansson P. H., Burkert A., Naab T., 2009, ApJ, 707, L184
Joseph R. D., Wright G. S., 1985, MNRAS, 214, 87
Kajisawa M. et al., 2009, ApJ, 702, 1393
Kartaltepe J. S. et al., 2010, ApJ, 709, 572
Kartaltepe J. S. et al., 2012, ApJ, 757, 23
Kauffmann G., Heckman T. M., 2009, MNRAS, 397, 135
Kelly B. C., Vestergaard M., Fan X., 2009, ApJ, 692, 1388
Kennicutt R. C., Jr, 1998, ApJ, 498, 541
Kewley L. J., Groves B., Kauffmann G., Heckman T., 2006, MNRAS, 372, 961
Kocevski D. D. et al., 2012, ApJ, 744, 148
Kormendy J., Bender R., 2012, ApJS, 198, 2
Kormendy J., Ho L. C., 2013, ARA&A, 51, 511
Kormendy J., Richstone D., 1995, ARA&A, 33, 581
Kormendy J., Kennicutt R. C., Jr, 2004, ARA&A, 42, 603
Kormendy J., Bender R., Cornell M. E., 2011, Nature, 469, 374
Koss M., Mushotzky R., Veilleux S., Winter L., 2010, ApJ, 716, L125
Koss M., Mushotzky R., Treister E., Veilleux S., Vasudevan R., Trippe M., 2012, ApJ, 746, L22
Krumpe M., Miyaji T., Coil A. L., Aceves H., 2012, ApJ, 746, 1
La Franca F. et al., 2005, ApJ, 635, 864
Letawe Y., Letawe G., Magain P., 2010, MNRAS, 403, 2088
Lidz A., Hopkins P. F., Cox T. J., Hernquist L., Robertson B., 2006, ApJ, 641, 41
Liu X., Zakamska N. L., Greene J. E., Strauss M. A., Krolik J. H., Heckman T. M., 2009, ApJ, 702, 1098
Li Y. et al., 2008, ApJ, 678, 41
Lotz J. M., Jonsson P., Cox T. J., Primack J. R., 2008, MNRAS, 391, 1137
Lotz J. M., Jonsson P., Cox T. J., Croton D., Primack J. R., Somerville R. S., Stewart K., 2011, ApJ, 742, 103
Lutz D. et al., 2008, ApJ, 684, 853
McGaugh S. S., 2005, ApJ, 632, 859
McGreer I. D. et al., 2013, ApJ, 768, 105
McLure R. J., Dunlop J. S., 2004, MNRAS, 352, 1390
Magorrian J. et al., 1998, AJ, 115, 2285
Mannucci F. et al., 2009, MNRAS, 398, 1915
Marchesini D., van Dokkum P. G., Förster Schreiber N. M., Franx M., Labbé I., Wuyts S., 2009, ApJ, 701, 1765
Marconi A., Risaliti G., Gilli R., Hunt L. K., Maiolino R., Salvati M., 2004, MNRAS, 351, 169
Martini P., 2004, in Ho L. C., ed., Coevolution of Black Holes and Galaxies. Cambridge Univ. Press, Cambridge, p. 169
Melbourne J. et al., 2008, AJ, 136, 1110
Merloni A., Heinz S., 2008, MNRAS, 388, 1011
Mihos J. C., Hernquist L., 1994, ApJ, 437, L47
Mullaney J. R. et al., 2012, MNRAS, 419, 95
Myers A. D. et al., 2006, ApJ, 638, 622
Myers A. D., Brunner R. J., Nichol R. C., Richards G. T., Schneider D. P., Bahcall N. A., 2007a, ApJ, 658, 85
Myers A. D., Brunner R. J., Richards G. T., Nichol R. C., Schneider D. P., Bahcall N. A., 2007b, ApJ, 658, 99
Naab T., Jesseit R., Burkert A., 2006, MNRAS, 372, 839
Nandra K. et al., 2007, ApJ, 660, L11

Peng C. Y., Impey C. D., Rix H.-W., Kochanek C. S., Keeton C. R., Falco E. E., Lehár J., McLeod B. A., 2006, ApJ, 649, 616
Pérez-González P. G. et al., 2008, ApJ, 675, 234
Porciani C., Norberg P., 2006, MNRAS, 371, 1824
Puech M. et al., 2008, A&A, 484, 173
Puech M., Hammer F., Hopkins P. F., Athanassoula E., Flores H., Rodrigues M., Wang J. L., Yang Y. B., 2012, ApJ, 753, 128
Richards G. T. et al., 2005, MNRAS, 360, 839
Richards G. T. et al., 2006, AJ, 131, 2766
Robertson B., Cox T. J., Hernquist L., Franx M., Hopkins P. F., Martini P., Springel V., 2006a, ApJ, 641, 21
Robertson B., Bullock J. S., Cox T. J., Di Matteo T., Hernquist L., Springel V., Yoshida N., 2006b, ApJ, 645, 986
Rosario D. J. et al., 2013, ApJ, 763, 59
Salucci P., Szuszkiewicz E., Monaco P., Danese L., 1999, MNRAS, 307, 637
Salviander S., Shields G. A., Gebhardt K., Bonning E. W., 2007, ApJ, 662, 131
Sanders D. B., Mirabel I. F., 1996, ARA&A, 34, 749
Santini P. et al., 2012, A&A, 540, A109
Sargent M. T., Béthermin M., Daddi E., Elbaz D., 2012, ApJ, 747, L31
Satyapal S., Ellison S. L., McAlpine W., Hickox R. C., Patton D. R., Mendel J. T., 2014, MNRAS, 441, 1297
Schawinski K., Treister E., Urry C. M., Cardamone C. N., Simmons B., Yi S. K., 2011, ApJ, 727, L31
Schweizer F., 1998, in Kennicutt R. C., Jr, Schweizer F., Barnes J. E., Friedli D., Martinet L., Pfenniger D., eds, Saas-Fee Advanced Course 26: Galaxies: Interactions and Induced Star Formation. Springer-Verlag, Berlin, p. 105
Schweizer F., Seitzer P., 2007, AJ, 133, 2132
Serber W., Bahcall N., Ménard B., Richards G., 2006, ApJ, 643, 68
Shankar F., Crocce M., Miralda-Escude J., Fosalba P., Weinberg D. H., 2010, ApJ, 718, 231
Shankar F., Weinberg D. H., Miralda-Escudé J., 2009, ApJ, 690, 20
Shapley A. E., Coil A. L., Ma C.-P., Bundy K., 2005, ApJ, 635, 1006
Shen Y. et al., 2009, ApJ, 697, 1656
Shen Y. et al., 2013, ApJ, 778, 98
Shen Y. et al., 2010, ApJ, 719, 1693
Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1
Shi Y. et al., 2007, ApJ, 669, 841
Shields G. A., Menezes K. L., Massart C. A., Vanden Bout P., 2006, ApJ, 641, 683
Siana B. et al., 2008, ApJ, 675, 49
Silverman J. D. et al., 2008, ApJ, 675, 1025
Snyder G. F., Hopkins P. F., Hernquist L., 2011a, ApJ, 728, L24
Snyder G. F., Cox T. J., Hayward C. C., Hernquist L., Jonsson P., 2011b, ApJ, 741, 77
Snyder G. F., Hayward C. C., Sajina A., Jonsson P., Cox T. J., Hernquist L., Hopkins P. F., Yan L., 2013, ApJ, 768, 168
Soltan A., 1982, MNRAS, 200, 115
Somerville R. S., 2009, MNRAS, 399, 1988
Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, MNRAS, 391, 481
Springel V., Hernquist L., 2005, ApJ, 622, L9
Stewart K. R., Bullock J. S., Wechsler R. H., Maller A. H., 2009a, ApJ, 702, 307
Stewart K. R., Bullock J. S., Barton E. J., Wechsler R. H., 2009b, ApJ, 702, 1005
Tacconi L. J. et al., 2008, ApJ, 680, 246
Tacconi L. J. et al., 2010, Nature, 463, 781
Treister E., Schawinski K., Urry C. M., Simmons B. D., 2012, ApJ, 758, L39
Treu T., Woo J.-H., Malkan M. A., Blandford R. D., 2007, ApJ, 667, 117
Trump J. R., 2011, MNRAS, preprint (arXiv:1112.3970)
Trump J. R. et al., 2009, ApJ, 700, 49
Ueda Y., Akiyama M., Ohta K., Miyaji T., 2003, ApJ, 598, 886
Vanden Berk D. E. et al., 2006, AJ, 131, 84
Veilleux S. et al., 2009, ApJ, 701, 587
Wang R. et al., 2008, ApJ, 687, 848
White M. et al., 2012, MNRAS, 424, 933
Woods D. F., Geller M. J., 2007, AJ, 134, 527
Woods D. F., Geller M. J., Barton E. J., 2006, AJ, 132, 197
Woo J.-H., Treu T., Malkan M. A., Blandford R. D., 2006, ApJ, 645, 900
Yencho B., Barger A. J., Trouille L., Winter L. M., 2009, ApJ, 698, 380
Younger J. D., Hopkins P. F., 2011, MNRAS, 410, 2180
Younger J. D., Hopkins P. F., Cox T. J., Hernquist L., 2008, ApJ, 686, 815
Younger J. D., Hayward C. C., Narayanan D., Cox T. J., Hernquist L., Jonsson P., 2009a, MNRAS, 396, L66
Younger J. D. et al., 2009b, ApJ, 704, 803
Yu Q., Lu Y., 2004, ApJ, 602, 603
Yu Q., Tremaine S., 2002, MNRAS, 335, 965
Yu Q., Lu Y., Kauffmann G., 2005, ApJ, 634, 901
Zakamska N. L. et al., 2006, AJ, 132, 1496
Zakamska N. L., Gómez L., Strauss M. A., Krolik J. H., 2008, AJ, 136, 1607
Zamojski M., Yan L., Dasyra K., Sajina A., Surace J., Heckman T., Helou G., 2011, ApJ, 730, 125
Zheng X. Z. et al., 2009, ApJ, 707, 1566

This paper has been typeset from a TEX/LATEX file prepared by the author.