THE NON-CAUSAL ORIGIN OF THE BLACK-HOLE–GALAXY SCALING RELATIONS

Knud Jahnke and Andrea V. Macciò
Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany; jahnke@mpia.de, macci@mpia.de

Received 2010 June 2; accepted 2011 March 31; published 2011 May 31

ABSTRACT

We show that the $M_{\text{BH}}-M_{\text{bulge}}$ scaling relations observed from the local to the high-$z$ universe can be largely or even entirely explained by a non-causal origin, i.e., they do not imply the need for any physically coupled growth of black hole (BH) and bulge mass, for example, through feedback by active galactic nuclei (AGNs). Provided some physics for the absolute normalization, the creation of the scaling relations can be fully explained by the hierarchical assembly of BH and stellar mass through galaxy merging, from an initially uncorrelated distribution of BH and stellar masses in the early universe. We show this with a suite of dark matter halo merger trees for which we make assumptions about (uncorrelated) BH and stellar mass values at early cosmic times. We then follow the halos in the presence of global star formation and BH accretion recipes that (1) work without any coupling of the two properties per individual galaxy and (2) correctly reproduce the observed star formation and BH accretion rate density in the universe. With disk-to-bulge conversion in mergers included, our simulations even create the observed slope of $\sim 1.1$ for the $M_{\text{BH}}-M_{\text{bulge}}$ relation at $z = 0$. This also implies that AGN feedback is not a required (though still a possible) ingredient in galaxy evolution. In light of this, other mechanisms that can be invoked to truncate star formation in massive galaxies are equally justified.

Key words: galaxies: bulges – galaxies: evolution – galaxies: fundamental parameters – galaxies: nuclei

Online-only material: color figures

1. INTRODUCTION

About a decade ago tight correlations between galaxy properties and those of central supermassive black holes (BHs) were empirically established: BH masses scale with the luminosity, the mass, and the velocity dispersion of their host galaxies’ bulges (Magorrian et al. 1998; Ferrarese & Merritt 2000; Gebhardt et al. 2000; Tremaine et al. 2002; McLure & Dunlop 2002; Marconi & Hunt 2003; Häring & Rix 2004; Gültekin et al. 2009; Jahnke et al. 2009; Merloni et al. 2010). These correlations were quickly interpreted to yield two important implications. (1) If these correlations were to exist for a random set of galaxies, then every galaxy should contain a supermassive BH (e.g., Kormendy & Richstone 1995). (2) If every galaxy contains a BH and given the observed scaling relations, the evolution of galactic bulges and central BHs should be coupled by a physical mechanism (the “co-evolution” picture).

The first statement appears to be true at least above some mass threshold, and introduced a new ingredient, the central BH, in galaxy evolution. The second statement is based on the vast energy available from accreting BHs, providing the easiest conceivable mechanism to physically couple BH and bulge properties despite the difference in linear scales. Coupling only a few percent of this energy in an “active galactic nucleus (AGN) feedback” (Silk & Rees 1998) to the gas within the galaxy would have vast implications on the temperature and structure of the surrounding interstellar medium. Ad hoc models (Granato et al. 2004; Di Matteo et al. 2005; Croton et al. 2006) were very successful in generating feedback loops that involve the energy from AGNs for quenching star formation (SF) and fueling of the AGNs themselves. Different incarnations of AGN feedback are in principle able to not only couple SF and BH accretion, but to simultaneously fix a number of existing problems in galaxy evolution, namely an overproduction of massive galaxies in semianalytic models as well as the inability to truncate SF fast enough to reproduce the observed color–magnitude bimodality of galaxies (Baldry et al. 2004). This motivated the inclusion of AGN feedback in a yet-to-be-determined form and physical description as the driving force behind the BH–bulge scaling relations.

Although the actual effectiveness and impact of at least “quasar mode” feedback models is still unclear, the interpretation of the scaling relations as a physically coupled evolution is largely assumed to be correct and continues to provide the basis for many studies. As an example, the evolution of the scaling relations (Treu et al. 2004, 2007; Peng et al. 2006a, 2006b; Schramm et al. 2008; Jahnke et al. 2009; Merloni et al. 2010; Bennert et al. 2010; Decarli et al. 2010) is investigated in order to constrain the physical drivers behind co-evolution and the growth mechanisms of BHs.

1.1. An Alternative Origin of the Scaling Relations

Peng (2007) demonstrated a thought experiment for the potential origin of an $M_{\text{BH}}-M_{\text{bulge}}$ relation without a physical coupling, but as the result of a statistical convergence process. In short, he showed that, in principle, arbitrary distributions of $M_{\text{BH}}/M_*$ ratios in the early universe converge toward a linear relation through the process of galaxy merging. In this central-limit-theorem view, a large number of mergers will average out the extreme values of $M_{\text{BH}}/M_*$ toward the ensemble average. What was deliberately left open in this experiment was whether there are enough major galaxy mergers in the history of an actual galaxy in order to drive this process far enough.

In this present study, we pick up this thought by following realistic ensembles of dark matter (DM) halos through cosmic time. Our immediate aim is to test whether the simple assembly of galaxies and their BHs according to a ΛCDM merger tree is able to produce BH–galaxy scaling relations from initial conditions at early times, where $M_{\text{BH}}$ and $M_{\text{bulge}}$ (or $M_*$) were completely uncorrelated per individual galaxy. We circumvent the inherent problems and degrees of freedom of a full semianalytic model by not trying to simultaneously solve the
problem of SF truncation or correct BH or stellar mass function, but restrict the question solely to the genesis of the scaling relations. As an input for SF and BH accretion, we use observed relations, but prevent any recipes that implicitly or explicitly couple BH and stellar mass growth per individual galaxy.

2. NUMERICAL SIMULATIONS AND MERGER TREES

In this paper, we use the Lagrangian code PINOCCHIO (Monaco et al. 2002) to construct high-resolution ΛCDM merger trees (for a comparison between N-body codes and PINOCCHIO see Li et al. 2007). The simulation has a box size of 100 Mpc and 1500³ particles; this ensures a very high mass resolution, \( m_p = 1.01 \times 10^7 M_\odot \). The construction of merger trees is straightforward with PINOCCHIO: the code outputs a halo mass every time a merger occurs, i.e., when a halo with more than 10 particles merges with another halo. From an initial 6.5 × 10⁶ halos, we receive a resulting sample of 10,932 halos with \( M > 10^{11} M_\odot \) at \( z = 0 \).

When two halos merge, the less massive one can either survive and continue to orbit within the potential well of the larger halo until \( z = 0 \) or merge with the central object. The averaging process described above will only apply to this second category of halos (the ones that actually merge). We then adopt the dynamical friction formula presented in Boylan-Kolchin et al. (2008) to compute the fate of a halo. The orbital parameters of the halo are extracted from suitable distributions that reproduce the results of N-body simulations as described in Monaco et al. (2007). If the dynamical friction time is less than the Hubble time at that redshift, we consider the halo to merge at a time \( t = t_{\text{dyn}} \); if it is longer the satellite halo is removed from our catalog. Each halo in our sample at \( z = 0 \) has formed by at least 200 mergers and the most massive ones have had more than \( 5 \times 10^3 \) encounters.

3. CREATING SCALING RELATIONS: AVERAGING AND MASS FUNCTION BUILD UP BY HIERARCHICAL MERGING

The main message of this work is to demonstrate what effect merging over cosmic time has on an ensemble of halos with initially uncorrelated \( M_{\text{BH}} \) and \( M_* \) values. These values change their distribution and converge toward a linear relation by \( z = 0 \)—in the absence of SF, BH accretion, disk-to-bulge conversion, and hence any physical connection between the two masses.

For this task, we follow DM halos through their assembly chain. We assign a stellar and a BH mass to each DM halo once its mass becomes larger than \( 10^8 M_\odot \), the corresponding redshift in the following is called \( z_2 \). We set our initial guesses for \( M_* \) and \( M_{\text{BH}} \) as a fixed fraction of the DM mass plus a (large) random scatter. We used \( M_*/M_{\text{DM}} = 10^{-3} \) and \( M_{\text{BH}}/M_{\text{DM}} = 10^{-7} \) for the initial ratio; the scatter is taken from a lognormally flat distribution of 3 dex for the two quantities (blue squares in Figures 1 and 2).

We have no knowledge of any realistic seed mass scatter, but take four orders of magnitude variations in the \( M_{\text{BH}}/M_* \) ratio as a proxy for “uncorrelated.” Empirical constraints on the possible parameter space for seed BH mass do not seem to support seeds more massive than \( \sim 10^5 M_\odot \) (Volonteri & Natarajan 2009); toward lower masses a few solar mass BHs are clearly being produced by stars. Whether this matches the true distribution of seed masses is not important, but by simply taking a large range of values that is currently not ruled out represents a rather conservative starting point for our demonstration. Halos are then propagated along the merger tree to \( z = 0 \) (the red points in Figures 1 and 2). When two halos merge according to our dynamical friction formula, we set the resulting stellar and BH masses equal to the sum of the individual masses before the merger (Volonteri et al. 2003). The final mass in \( M_{\text{BH}} \) and \( M_* \) as well as the corresponding normalization is determined simply by the sum of the individual halos contributing to a final halo.

Figure 1 shows that the hierarchical formation of galaxies provides a strong inherent driver from the uncorrelated initial distribution to a linear relation. This effect is independent of the chosen initial conditions as Figure 2 demonstrates, where completely different initial conditions result in a relation with the same slope and very similar scatter.

This experiment shows that the dominating structural parts of the observed \( M_{\text{BH}}-M_* \) scaling relation—i.e., (1) the existence of such a correlation, (2) that it extends over several orders of magnitude in mass, (3) the fact that the slope is near unity, and (4) an increasing scatter to lower masses—can be explained by this physics-, feedback-, and coupling-free process. A slope \( \sim 1 \) scaling relation does not need any physical interaction of galaxy and BH. In the next sections, we will show that this also holds when adding “second-order” effects like actual SF and BH accretion, as well as disk–bulge conversion.

---

1. We picked this mass since at lower masses halos likely did/do not form stars at all (Macciò et al. 2010). The most massive progenitors of \( z = 0 \) galaxies form according to this definition in the range \( z = 15–17 \), while low-mass satellites can form as late as \( z \sim 3 \).

2. The convergence is in fact too strong (see the next sections), as the scaling relation it produces by \( z = 0 \) is much tighter than the observed 0.3 dex scatter. In principle, the scatter has a \( \propto N \) dependency on the number \( N \) of merger generations, but the relation gets complicated by the different masses of the merging components and different merging times across the tree.
Figure 2. Scaling relations are produced independently of the initial conditions. Shown are four vastly different initial conditions—for a subset of our halos for better visibility—all leading to a slope $= 1$ relation at $z = 0$ with similar scatter. The $z = 0$ normalization comes out differently since the geometric mean of the initial masses is different (flat distribution in logarithm for both quantities). Symbols and the line have the same meaning as in Figure 1.

(A color version of this figure is available in the online journal.)

4. ADDING STAR FORMATION, BLACK HOLE ACCRETION, AND DISK-TO-BULGE CONVERSION

We so far demonstrated that merging alone is the basic mechanism to create an $M_{\text{BH}}-M_*$ scaling relation: However, we need to add a number of ingredients to our model. Placing all mass already at high redshift is not conservative, since all of stellar and BH mass is subject to the full merger averaging process. In the actual universe, we know that the majority of BH and stellar mass in the universe was created after $z \sim 6$ (Soltan 1982; Hopkins & Beacom 2006) and hence experiences less merger generations. Moreover, pure merging produces a monotonic relation between $M_{\text{DM}}$ and $M_*$ as shown in the upper panel of Figure 3, which is at odds with empirical results. Since SF and BH accretion density depend on redshift, we will add the approximately right amounts of SF and BH growth at the right cosmic times, and thus at the right place in time with respect to the merger cascade. The goal of this exercise is not to create a full semianalytic model of galaxy formation, but to test what effect realistic assumptions about mass growth have on the resulting $M_{\text{BH}}-M_*$ scaling relation.

To construct SF and BH accretion recipes, we will use three observed relations as input: (1) the halo occupation distribution (HOD), i.e., the relation between DM halo mass $M_{\text{DM}}$ and $M_*$ (Moster et al. 2010), (2) the Lilly–Madau relation for the evolution of SF rate (Hopkins & Beacom 2006), and (3) the evolution of the AGN bolometric luminosity (Hopkins et al. 2007).

We want to explicitly note that by using these relations, we do not force any coupling of $M_{\text{BH}}$ and $M_*$ or $M_{\text{bulge}}$ per individual galaxy, but all recipes relate only to ensemble averages. Any potential implicit couplings act only on the ensemble and not...
on an individual galaxy; hence they cannot induce a correlation in the (originally uncorrelated) data points.

4.1. Star Formation

Up to now, at redshift \( z = 0 \) the stellar mass of our Galaxy is simply given by the sum of all stellar masses of its \( j \)-progenitors \( M_{*}^{j}(i) = \sum M_{*}(j,i) \), masses that as explained were originally drawn from a random distribution. As shown in the upper panel of Figure 3, for a given halo mass the stellar mass obtained in this way is too low when compared to the empirical expectations from HOD models (e.g., Moster et al. 2010). This is because the total galaxy is more than the sum of its seeds and we have neglected SF so far. On the other hand, Figure 3 tells us exactly how much stellar mass each halo is missing. Hence, we take this constraint, the HOD results, to fix the stellar mass produced through SF:

\[
M_{*}^{SF}(i) = M_{*}(\text{dm}(i)) - M_{*}^{0}(i),
\]

where \( M_{*}(\text{dm}(i)) \) is the expected stellar mass for the \( i \)-th halo with DM mass \( M_{\text{dm}}(i) \) as predicted by the HOD model presented in Moster et al. (2010).

Now we need to distribute this stellar mass from SF along time, i.e., among all progenitors of galaxy \( i \) along the merger tree. We do that according to the following formula that gives the stellar mass produced through SF for the \( j \)-th halo in the merger tree of the final i halo:

\[
M_{*}^{SF}(i,j) = A \times M_{*}^{j}(j) \times \text{LT}(j) \times f(z_{f}(j),z_{m}(j)).
\]

The constant \( A \) is fixed by the requirement that \( M_{*}^{SF}(i) = \sum M_{*}^{SF} \) and is the same for all progenitors. The time LT is the lifetime of a halo, defined as the time between the formation redshift \( z_{f} \); when \( M_{\text{DM}} > 10^{8} \, M_{\odot} \) and the moment \( z_{m} \) when it merges with a more massive halo, which is not necessarily the main branch of the merger tree. With this definition, we assume that galaxies are able to actively form stars only when they are the central objects within their host halo.

The function \( f \) is used to give different weights to the lifetime LT at high and low redshift; in this way, for a given lifetime, a galaxy will produce more (fewer) stars at high (low) redshift, according to the Lilly–Madau plot.\(^3\) We define \( f(z_{1},z_{2}) \) as the integral between \( z_{1} \) and \( z_{2} \) of the assumed star formation rate (SFR):

\[
f(z_{1},z_{2}) = \int_{z_{1}}^{z_{2}} \text{SFR}(z) \, dz.
\]

In our reference model, we assume for the redshift evolution of the SFR the functional form suggested by Hopkins & Beacom (2006), namely the results listed in Table 1 for a modified Salpeter initial mass function (see Hopkins & Beacom 2006 for more details). Finally, the factor \( M_{*}^{j}(i) \) takes care of the observed mass dependence of specific SFRs and we fixed the exponent \( q = 0.8 \) (e.g., Daddi et al. 2007; Bouché et al. 2010). In Appendix A, we will present results for different choices for \( q \) and SFR(z).

Let us summarize one more time our parameterization for SF: When the \( j \)-th halo appears at \( z_{f}(j) \) it gets an initial stellar mass \( M_{*}(z_{f}(j)) \) from a random distribution as described in Section 3. Then it will “produce” its own stars \( M_{*}^{SF}(j) \) until it is accreted onto, and becomes part of, a more massive halo. During its lifetime it will also accrete stellar mass from merging (lower) mass halos. If halo \( j \) merges with the central galaxy it will add a fraction of its stellar mass to the bulge of the central galaxy as described below in Section 4.2. If halo \( j \) merges with a more massive halo \( k \) before merging with the central halo it will give all its stellar mass to halo \( k \) and cease to exist as a halo of its own.

4.2. Disk-to-Bulge Conversion

We assume that all stellar mass produced through SF will occur in the disk component of each halo and then apply a recipe to convert part of this disk mass to bulge mass as a consequence of mergers.

The amount of disk-to-bulge mass conversion depends on a multitude of parameters such as mass ratio, gas fraction, and orbital parameters of the merger, which are impossible to implement in our context. Instead, we follow a simpler recipe

\(^3\) We note that we do not include a different shape of this star formation history as a function of mass. As Appendix A shows, this will not have any effect on the results.
inspired by the numerical results of Hopkins et al. (2010),
depending solely on the stellar mass ratio of the two merging
partners: (1) bulge mass of main halo and satellite will just be
co-added, (2) the disk mass of the satellite goes fully into the
resulting bulge, and (3) a fraction of the main halo disk, directly
proportional to the mass ratio, also gets converted into bulge
mass. With this recipe we are able to approximately reproduce
the ratio of bulge to total mass observed in the local universe.

4.3. Black Hole Accretion

Since the mechanisms of BH accretion continue to be unclear,
we assume a simple recipe: the BH will double its mass in a
stochastic way on a characteristic timescale $\tau$. This will happen
for all $j$ progenitors of halo $i$. Similar to what we have done
for SF, we link the number of doublings of a given halo to its
time and we weight this time with a function $g$ similar to the
function $f$ in Section 4.1.

In practice, the number of mass doublings of the $j$th BH in
the $i$ merger tree is given by the expression:

$$N_{\text{doub},j} = LT(j) \times g(z_f(j), z_m(j))/\tau, \quad (4)$$

where LT has the same meaning as in Equation (2). For BH
accretion, we choose as the weighting function $g(z_1, z_2)$ the
integral between $z_1$ and $z_2$ of the AGN bolometric luminosity
$(\text{AGN}_{L} (z))$:

$$g(z_1, z_2) = \int_{z_1}^{z_2} \text{AGN}_{L}(z) \, dz. \quad (5)$$

In our reference model, the redshift evolution of the AGN
bolometric luminosity is modeled with a double power law
aimed to reproduce the results of Hopkins et al. (2007,
Figure 8):

$$\log \left( \frac{\text{AGN}_{L}(z)}{L_\odot \text{Mpc}^{-3}} \right) = \begin{cases} 2.02 \cdot \log(z) + 7.83 & \text{for } z < 1.7 \\ -2.09 \cdot \log(z) + 8.78 & \text{for } z \geq 1.7. \end{cases} \quad (6)$$

Similarly to the $f$ function, $g$ can be used to allow for higher
accretion rates at high redshift compared to low redshift for a
fixed halo lifetime.

The characteristic time $\tau$ is chosen in order to match the
normalization of the observed $M_{\text{BH}}$-$M_{\text{bulge}}$ scaling relation at
$z = 0$ for $\log(M_{\text{BH}}) = 7$ and we obtained $\tau = 1.9 \times 10^9$ Gyr.
As a stochastic element to BH growth, we cast for each of the
$N_{\text{doub},j}$ events a random number from a flat distributing in the range [0:1] and effectively double the BH mass only if
this random number is $>0.5$. For the main branch the number of
doublings is of the order of 6–8, while it is in the range 0–3 in the other branches of the tree. In Appendix A, we will
present results for different choices of the parameters involved
in Equation (4).

4.4. Resulting Scaling Relations

The $M_{\text{bulge}}$-$M_{\text{BH}}$ distribution at $z = 0$ with the above recipes
added is shown in Figure 4, with the observed values over-
plotted. Compared to the pure merger assembly in Figures 1
and 2, we note a substantially increased scatter—closer to the
observed—and a steeper than linear relation. Still, despite both
the shift of SF and BH growth to later times as well as the
random parts of SF and BH accretion, a clear correlation is
produced. The effects of SF, BH accretion, and disk-to-bulge
conversion do not destroy the correlations but only induce a
“second-order” modification of them. Over time a mass func-
tion is built up and the scatter decreases substantially from the
initial 4 dex, particularly for the $\log(M_{\text{bulge}}) > 10$ regime.
The simulated relation is very similar to the observed points: it reproduces the slope \( > 1 \) almost perfectly, even the fact that at the high-mass end the observed points lie above the mean slope—and this without adjustable parameters beyond normalization. Also, the higher scatter at low masses can be seen in both simulations and observations; the “cloud” above the \( z = 0 \) relation in our model data represents disk galaxies with small bulge masses as observed by Greene et al. (2008). As can be seen in Figure A.4, these objects would still lie on the local mass scaling relation, but with their total and not their bulge mass.

We want to stress that the results in Figure 4 do not depend on our parameterization of SF rate or BH accretion. For example, changing the functional form of \( f \) or \( g \) in Equations (2) and (4) only marginally affects the scatter of the simulated \( M_{\text{BH}}-M_{\text{bulge}} \) relation and leaves the slope unchanged. The same is true for the other parameters described in the previous sections; see Appendix A for details. There is a sole exception to this—the assumed initial seeding masses. The larger the initial \( M_{\text{BH}} \) and \( M_\ast \), the less mass has to be created by SF and BH accretion. In this way, less mass is entering the halos at later times and more mass is subject to the full cascade of mergers, which will lead to a smaller scatter in the scaling relations at \( z = 0 \).

5. DISCUSSION

We showed above that all basic properties of the BH–bulge mass scaling relation in the local universe—a relation between properties of individual galaxies—are produced naturally by the merger-driven assembly of bulge and BH mass, and without any coupling of SF and BH mass growth per individual galaxy.

The convergence power of galaxy merging is very strong for a realistic halo merging history, even with a correct placement of SF and BH accretion along cosmic time. This means that the mechanism Peng (2007) sketched in his thought experiment works also in a realistic universe—there is enough merging occurring in the universe and hence (most of) the scaling relations can be entirely explained without any physical mechanisms that directly couple \( M_{\text{bulge}} \) and \( M_{\text{BH}} \) growth for a given object.

5.1. Implicit Coupling of Black Hole and Galaxy?

The scaling relations are produced naturally in our toy model—but does this preclude any implicit coupling of \( M_{\text{BH}} \) and \( M_\ast \) by design? The most obvious features to be discussed in this respect are (a) the shape of the HOD and (b) the question of which mechanism determines the relative value of BH accretion to SF, and hence the absolute normalization of the \( M_{\text{BH}}-M_{\text{bulge}} \) relation.

1. **HOD shape.** The HOD was empirically inferred and shows that the ratio of stellar to DM mass is not constant but is a function of mass itself. Toward the massive end stellar mass does not increase in parallel to DM; SF appears suppressed. Its impact on the \( M_{\text{BH}}-M_{\text{bulge}} \) relation is the slight curvature in Figure 4 with the noticeable upturn at the massive end—consistent with the observations. The HOD shape represents a long known feature of galaxy formation and was “fixed” in models by the introduction of a quenching mechanism, suppressing SF above some mass, often by inclusion of AGN feedback recipes.

In our toy model, however, we do not make specific assumptions about which mechanism produces the HOD we use as a constraint. We do not tie BH accretion or a merger event to the suppression of SF—it can be suppressed by any mechanism, e.g., a modified supernova feedback recipe or gravitational heating by infalling clumps of matter. The latter mechanisms are completely independent of BH accretion and, while modifying the HOD, by nature cannot have an impact on the \( M_{\text{BH}}-M_{\text{bulge}} \) relation. Even if AGN feedback was the source for shaping the HOD, this would only be the cause of the second-order shape deviation from a linear slope, not the existence of the \( M_{\text{BH}}-M_{\text{bulge}} \) relation itself.

2. **Absolute normalization.** Our initial model just propagates the (rather ad hoc) seed masses in stars and BH to \( z = 0 \), i.e., the normalization of the \( M_{\text{BH}}-M_{\text{bulge}} \) relation at \( z = 0 \) is directly set by the ensemble mean ratio of seed masses. Since most of stellar and BH mass in reality is produced by SF and BH accretion later on, the high-\( z \) ratio is in fact unconnected to the \( z = 0 \) normalization.

In our simulation we set the normalization by requesting a match of our simulation results with the empirical \( M_{\text{BH}}-M_{\text{bulge}} \) relation at \( M_\ast = 10^{10} M_\odot \). Arguments have been brought forward that the actual normalization must be the result of a regulatory feedback loop involving BH accretion and SF. This is an attractive scenario, as it would explain both the creation of the scaling relations as well as their normalization. Models of this kind have been implemented in several semianalytic models of galaxy formation, in all cases with free parameters that actually control the absolute normalization of the resulting scaling relations, set to ad hoc values to again match this and other observations. Since we show in this paper that (most aspects of) the \( M_{\text{BH}}-M_{\text{bulge}} \) scaling relations are created automatically by hierarchical assembly, these feedback models actually seem to achieve too much—the creation of a certain \( M_{\text{BH}}/M_\ast \) ratio for each galaxy and at all times, which as a conspiracy would come on top of the formation path of the scaling relations demonstrated here.

This said, we want to sketch the outline of an alternative scenario that could well be responsible for the absolute normalization but has not yet been explored. The main ingredients for both SF and BH accretion are (1) gas and (2) a trigger to form stars or to bring down gas to the BH. For both stars and BHs, the amount of growth is basically a product of the two ingredients. At early times gas was ample and the number of both galaxy mergers and gas disk instabilities was high until the peak of activity around \( z = 2 \). The triggering mechanisms subsequently decreased with the decreasing number of mergers and reduced gas reservoirs, either in number or duration or both. If both SF and BH accretion were to be ruled by a set of random triggering mechanisms and the specific gas fraction in a galaxy, then very high-\( z \) galaxies might exhibit a strong variance in their \( M_{\text{BH}}/M_\ast \) distribution as well as in their (instantaneous and also time averaged) BH accretion over SF ratio, with a dependency on the actual total mass, morphology, gas mass, trigger type, environment, etc. However, as a cosmic mean there will be a global \( M_{\text{BH}}/M_\ast \) value, just as an average over the efficiency of the spectrum of random triggering mechanisms and realized conditions to produce new stars or BH mass—a number which could also change with time.

The hierarchical assembly of galaxies and its averaging mechanism now relieves us from having to search for a regulatory mechanism per galaxy, since with each galaxy merger the spectrum of \( M_{\text{BH}}/M_\ast \) values will increasingly average out. The fact that the slowly starting depletion of gas reservoirs in galaxies at \( z \sim 2 \) is accompanied with, and not preceded by, a slowly decreasing merger rate has the consequence that at higher
redshifts there were enough mergers to average out the extreme \( M_{\text{BH}}/M_* \) systems—at lower \( z \), with a decaying gas reservoir and merger rate and the transition to a “secular universe” (see, e.g., Cisternas et al. 2011), the prerequisites for producing new extreme values become less and less frequently fulfilled. The ensemble converges toward the observed \( M_{\text{BH}}-M_{\text{bulge}} \) relation at \( z = 0 \).

Recently, the first nested multi-scale simulations of gas inflow into the very centers of galaxies have been successful (Hopkins & Quataert 2010) and give a first impression of how in principle random instabilities, strongly depending on the actual conditions in the galaxy, can create gas inflow into the very center and the fueling of either BH or SF or both. Mechanisms of this kind will result in a certain mean value of \( M_{\text{BH}}/M_* \) averaged over all galaxies, while they can be vastly different for an individual galaxy. How different is so far unclear, and whether this mechanism for BH fueling precludes “runaway” growth of BHs to 100- or 1000-fold in a single instance, though unlikely, needs to be seen. All of this can in principle be realized without any AGN feedback—though it does not rule it out—and has the freedom to have a mass- or environmental-dependent efficiency component, as we also observe a non-unity slope of the \( z = 0 \) scaling relation.

Coming back to the initial question, we conclude that there is no necessity that our toy model makes any implicit assumptions about a physical \( M_{\text{BH}}-M_* \) coupling.

5.2. Further Consequences of This Mechanism

Our mechanism is also able to explain other observational results that are often used as evidence to support the picture of AGN feedback.

1. One of them is the potentially different \( M_{\text{BH}}-M_{\text{bulge}} \) relations in galaxies with classical and pseudo-bulges (Greene et al. 2008; Gadotti & Kauffmann 2009). Pseudo-bulges are thought to be formed through secular processes, rather than major merging (Kormendy & Kennicutt 2004). This has the immediate implication that the bulge has taken a different long- or mid-term assembly route compared to the BH; hence it is actually not expected that galaxies with pseudo-bulges obey the same \( M_{\text{BH}}-M_{\text{bulge}} \) relation as those with classical ones.

2. Hopkins et al. (2010) found that the stellar mass inside a very small radius near the BH \( M_*(< R) \), comparable to the BH’s sphere of influence, shows a much larger scatter than the scatter in \( M_{\text{BH}}/M_* \). They argue that this is an indication that gas was transported to near the BH, to form said stars, but that this had apparently no impact on the scatter in \( M_{\text{BH}} \); hence a self-regulating mechanism should have been at work.

With our model this also can be simply explained. The BH and bulge of a galaxy take part in the same long-term assembly and averaging cascade; hence their values correlate and the scatter in their ratio is small. The central density of stars, on the other hand, does not. It is governed by more short-term gas inflow, SF, and redistribution mechanisms; hence it is expected that it does not correlate well with the overall mass of the bulge or the BH.

3. Following this line of argument, we in principle expect a correlation with \( M_{\text{BH}} \) for any (mass) parameter that is subject to the same \( \Lambda \)CDM assembly chain. This includes the halo mass, to some extent the total mass of the galaxy (for both see Appendix B and Figure A.4), but also, e.g., the total mass of globular clusters in a galaxy, which recently has been found empirically by Burkert & Tremaine (2010).

4. On the other hand, our model is too basic to be able to reproduce higher-order effects. A number of studies (Aller & Richstone 2007; Hopkins et al. 2007; Feoli & Mancini 2009; Feoli et al. 2010) have suggested that the most fundamental relation with \( M_{\text{BH}} \) is neither \( M_{\text{bulge}} \) nor \( M_\Lambda \) but rather galaxy binding energy or potential well depth.

What these studies actually find are residual correlations in the scaling relations that are in some way related to the compactness of a bulge/spheroid at a given \( M_{\text{BH}} \). This can be effective radius or binding energy or any other quantity that is explicitly or implicitly a measure of compactness. Since bulge mass depends just linearly on the progenitor mass ratios, the exact size, compactness, and binding energy of a spheroid produced in a merger will depend nonlinearly on other parameters like gas fraction, merger orbit, etc. These parameters could easily be responsible for the residual correlation found in the \( M_{\text{BH}}-M_{\text{bulge}} \) relation as they are measures of the specific short- or mid-term merger history of each galaxy, while \( M_{\text{bulge}} \) is a long-term integral. Our deliberately simple toy model itself cannot make any statements on this issue as it does not trace, e.g., galaxy scale radii.

Interpreting the above points with respect to the relevance of AGN feedback, we find no strong argument for AGN feedback as a necessary mechanism at work. However, this does not mean that AGN feedback does not exist. It only means that AGN feedback is still a possible mechanism involved in parts of galaxy evolution, which might be important, e.g., for subclasses of the galaxy population.4 This implies that other proposed means of energy (or momentum) injection that have been proposed to quench SF in massive galaxies (e.g., Dekel & Birnboim 2006; Khochfar & Ostriker 2008; Lo Faro et al. 2009) are viable options. In other words, all other mechanisms that can be invoked to truncate SF in massive galaxies appear to be equally justified.

5.3. The Comparison with Previous Work

Our result is different to all previous studies, and after more than a decade of semianalytic models including BHs, one of the obvious questions is why have others not found this before? The closest studies to ours are a short proceedings contribution by Gaskell (2010), which is basically paraphrasing Peng (2007), and the work by Hirschmann et al. (2010). Their study is a very systematic assessment of how the scaling relation scatter evolves under the influence of galaxy merging. However, their initial setup in all cases was an existing correlation of \( M_{\text{BH}} \) and \( M_* \), and due to their specific goal they deliberately ignored the influence from SF and BH accretion.

Other recent studies investigating this subject either explicitly include AGN feedback (e.g., Croton et al. 2006; Robertson et al. 2006; Booth & Schaye 2009; Johansson et al. 2009; Shankar et al. 2009) or couple the merging scenario with a self-regulation prescription for BH growth (e.g., Kauffmann & Haehnelt 2000; Volonteri & Natarajan 2009). In retrospect it is quite obvious why they did not find our results earlier, as most of these models were developed for galaxy evolution in general. Once BHs entered the equations, they added terms of (regulated) BH growth and, when evaluating the \( M_{\text{BH}}-M_{\text{bulge}} \) relations

4. Undoubtedly, at least “radio mode” feedback (Croton et al. 2006) is actually observed in some massive clusters (e.g., Fabian et al. 2006; Best et al. 2006) and possibly on the group level (Giodini et al. 2010).
generated, noted that their prescription managed to produce a decent match to the empirical relation. The interpretation that this meant the AGN feedback or coupling prescription were correct now requires re-evaluation from our new point of view—it was the simple result of hierarchical assembly at work.

5.4. Impact on Evolution Studies

In the last decade, a substantial number of attempts were made to measure local and higher redshift scaling relations of \(M_{\text{BH}}-M_{\text{bulge}}\), \(M_{\text{BH}}-\sigma_{\text{bulge}}\), or \(M_{\text{BH}}-L_{\text{bulge}}\) (Treu et al. 2004, 2007; Peng et al. 2006a, 2006b; Woo et al. 2006; Schramm et al. 2008; Jahnke et al. 2009; Merloni et al. 2010; Bennert et al. 2010; Decarli et al. 2010) and to interpret them with respect to the mechanisms that couple BH growth and their impact on galaxy formation. What implications does the non-causal origin of the scaling relations have for these results? If the mean cosmic \(M_*\) and \(M_{\text{BH}}\) actually evolved similarly, our results explain at least a part of the bulge scaling relation evolution for galaxies with substantial disk components: it is the simple conversion of disk-to-bulge mass in galaxy mergers. What still remains interesting and needs to be substantiated is how at high redshifts the relation between \(M_{\text{BH}}\) and total \(M_*\) (or even \(M_{\text{bulge}}\) for bulge-dominated galaxies) evolves (e.g., Walter et al. 2004). This would continue to predict a substantial early BH growth—with corresponding implications for BH feeding models.

One other aspect that could serve as a diagnostic is the evolution of the scaling relation scatter. When coupled with predictions of BH and stellar mass assembly from a proper model, the scatter can be used to study, e.g., the distribution of seed \(M_{\text{BH}}\) at early times. We will follow up on this issue in a future publication.

K.J. is funded through the Emmy Noether Programme of the German Science Foundation (DFG). The authors thank F. Fontanot for his help in creating the \texttt{pinocchio} merger trees, H.-W. Rix, E. F. Bell, F. Walter, R. Decarli and C. Y. Peng for valuable feedback, and L. Mancini for access to his data. Numerical simulations were performed on the PIA and PanStarrs2 clusters of the MPIA at the MPG Rechenzentrum in Garching. We thank the anonymous referee for a very thorough job and very helpful comments and suggestions.

APPENDIX A

TESTING THE MODEL: ARE THE PARAMETER CHOICES SPECIAL?

Few (free) parameters enter in our parameterization of SF, BH accretion, and bulge-to-disk conversion. In this section, we want to explore different choices with respect to our reference model and test their impact on the final results. The different models are listed in Table 1.

Figure A.1 shows the effect of varying our “weighting” functions. In model B, we set SFR\((z)\) = 1.0, i.e., we assume a constant SFR as a function of redshift; in model C we also set AGN\(_L\)(\(z\)) = 1.0. The resulting \(M_{\text{BH}}-M_{\text{bulge}}\) relations are indistinguishable from the original (A) model. The bottom right panel of Figure A.1 (model D) shows an even tighter correlation between \(M_{\text{BH}}-M_{\text{bulge}}\) with respect to our reference model. This is because in model D we remove the stochasticity in the BH mass doubling, forcing all BHs to double their mass every \(\tau\) Gyr. This increases the number of doublings in the merger tree branches with short lifetime, increasing the fraction of BH mass that is accreted through mergers and hence is subject to the central limit theorem.
In Figure A.2, we check the effect of our parameterization of dynamical friction (model E, where the dynamical friction time is multiplied by five) and of our disk-to-bulge conversion, assuming that a fraction of the disk mass proportional to the...
square root (model F) or proportional to the square of the merger ratio is promoted into the disk (model G).

Finally, models H–L (Figure A.3) test the importance of a dependence of SFR on stellar mass. In these models, we vary the $q$ exponent in Equation (2): models H and I do not show any appreciable variation when compared with model A. Model L, which has an absolutely unrealistic value for $q$, is the only model where we were able to break the $M_{\text{bulge}}$–$M_{\text{BH}}$ correlation. This result can be easily understood in the following way: if SF is too strong a function of stellar mass, then the vast majority of stars will be formed in the main branch of the tree that hosts (by definition) the most massive progenitor of our Galaxy. This implies that the bulk of stellar mass will not be subject to any averaging process and hence the central limit theorem does not apply. Moreover, given the artificially high fraction of stars produced within the central galaxy, there will be no major merger; this explains why in model L we do not get any bulge more massive than $4 \times 10^{10} M_\odot$.

All other models are not distinguishable from the reference model A: this underlines one more time the convergence power of galaxy mergers and shows that the actual implementation of SF, dynamical friction, BH accretion, and bulge formation produce only secondary effects.

**APPENDIX B**

**SCALING RELATIONS WITH STELLAR MASS AND HALO MASS**

Hierarchical assembly produces correlations between any two parameters that take part in this cascade. The main focus of this paper lies on the $M_{\text{BH}}$–$M_{\text{bulge}}$ relation, but $M_{\text{BH}}$ also correlates with total stellar mass and DM halo mass. This is shown in the two panels of Figure A.4, which include the recipes for BH accretion. The clearest and simplest correlation is the one with $M_{\text{DM}}$ (right side), which in the simple framework of this toy model is very close to slope $= 1$, with some scatter and no bending.

$M_{\text{BH}}$ versus total stellar mass (including SF recipe), as shown on the left side of Figure A.4, is at the massive end largely identical to the $M_{\text{BH}}$–$M_{\text{bulge}}$ view, because most galaxies there will have had sufficient numbers of minor and major mergers in order to make them bulge dominated. At the low-mass end, the scatter is still larger than at high masses, due to the smaller number of (averaging) past mergers for each halo, but it is somewhat smaller than for bulge mass, since the extra random element from disk-to-bulge conversion is absent. This also explains the missing plume of seemingly high-$M_{\text{BH}}$ systems at low masses visible in Figure 4, which in this way can be explained to actually be disk galaxies with small bulges and normal sized BHs for the amount of total stellar mass.

**REFERENCES**

Aller, M. C., & Richstone, D. O. 2007, ApJ, 665, 120
Baldry, I. K., Glazebrook, K., Brinkmann, J., Ivezić, Ž., Lupton, R. H., Nichol, R. C., & Szalay, A. S. 2004, ApJ, 600, 681
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., Kaiser, C. R., Heckman, T. M., & Kauffmann, G. 2006, MNRAS, 368, L67
Booth, C. M., & Schaye, J. 2009, MNRAS, 398, 53
Bouché, N., et al. 2010, ApJ, 718, 1001
Boylan-Kolchin, M., Ma, C., & Quataert, E. 2008, MNRAS, 383, 93
Burbker, A., & Tremaine, S. 2010, ApJ, 720, 516
Cisternas, M., et al. 2011, ApJ, 726, 57
Croton, D. J., et al. 2006, MNRAS, 365, 11
Daddi, E., et al. 2007, ApJ, 670, 156
Decarli, R., Falomo, R., Treves, A., Labita, M., Kotilainen, J. K., & Scarpa, R. 2010, MNRAS, 402, 2453
Dekel, A., & Birnboim, Y. 2006, MNRAS, 368, 2
Di Matteo, T., Springel, V., & Hernquist, L. 2005, Nature, 433, 604
Fabian, A. C., Sanders, J. S., Taylor, G. B., Allen, S. W., Crawford, C. S., Johnstone, R. M., & Iwasawa, K. 2006, MNRAS, 366, 417
Feoli, A., & Mancini, L. 2009, ApJ, 703, 1502
