Physics of Coevolution of Galaxies and Supermassive Black Holes

Renyue Cen
Princeton University Observatory, Princeton, NJ 08544, USA; cen@astro.princeton.edu

Received 2011 February 24; accepted 2012 May 24; published 2012 July 24

Abstract

A new physically based model for coevolution of galaxies and supermassive black holes (SMBHs) is presented. The evolutionary track starts with an event that triggers a significant starburst in the central region of a galaxy. In this model, the main SMBH growth takes place in the post-starburst phase, fueled by recycled gas from inner bulge stars in a self-regulated fashion on a timescale that is substantially longer than 100 Myr and at a diminishing Eddington ratio with time. We argue that the SMBH cannot gorge itself during the starburst phase, despite the abundant supply of cold gas, because star formation (SF) is a preferred mode of gas consumption over accretion to the central SMBH in such an environment. We also show that feedback from SF is at least as strong as that from an active galactic nucleus (AGN); thus, if SF is in need of being quenched, AGN feedback generally does not play the primary role. The predicted relation between SMBH mass and bulge mass/velocity dispersion is consistent with observations. A clear prediction is that early-type galaxy hosts of high-Eddington-rate AGNs are expected to be light blue to green in optical color, gradually evolving to the red sequences with decreasing AGN luminosity. A suite of falsifiable predictions and implications with respect to relationships between various types of galaxies, AGNs, and others are made. For those where comparisons to extant observations are possible, the model appears to be in good standing.

Key words: accretion, accretion disks – black hole physics – galaxies: active – galaxies: star formation – ISM: kinematics and dynamics – supernovae: general

1. Introduction

The tight correlation between galactic center supermassive black hole (SMBH) mass ($M_{\text{BH}}$) and bulge mass ($M_{\text{B,GC}}$) or velocity dispersion ($\sigma$) in the nearby universe (e.g., Richstone et al. 1998; Ferrarese & Merritt 2000; Tremaine et al. 2002) strongly suggests coevolution of the two classes, at least over the Hubble time. In many semi-analytic calculations one of the most adopted assumptions, to put it simply, is that active galactic nucleus (AGN) feedback is able to prevent most of the gas from accreting onto the SMBHs and at the same time is able to fix most of the “defects” of galaxy formation models, such as the shape of the galaxy luminosity function and star formation (SF) history (e.g., Kauffmann & Haehnelt 2000; Croton et al. 2006; Somerville et al. 2008) with the underlying feedback physics parameterized. The substantial success in explaining a variety of observations enjoyed by these semi-analytic models is indicative of the relevance of AGN feedback. Calculations of the coupled evolution of SMBHs and galaxies using three-dimensional hydrodynamic simulations deploy thermal energy feedback in regions significantly outside the Bondi radius of the putative SMBH that effectively couples to the surroundings to regulate the SF and eventually drive the gas away. These pioneering detailed simulations have provided much physical insight and appear to be remarkably successful in accounting for many intricate observables, including AGN light curves, Eddington ratio distributions, and the SMBH–bulge relation and its scatter, for certain chosen values of feedback energy strength (e.g., Di Matteo et al. 2005; Hopkins et al. 2006). What is hitherto left open in these calculations is the physical origin of the adopted energy feedback. One concern is that the derived SMBH–bulge relation depends very sensitively on the adopted energy feedback parameter due to the strong radiative cooling (e.g., Silk & Nusser 2010; Choi et al. 2012). Thus, it is prudent to seek underlying physical origins for these successful models and, before that is achieved, continue to explore alternative models.

This paper synthesizes an alternative physical model largely based on known physics. Before describing our overall model, we shall first, in Section 2, examine the plausibility of the fundamental claim that AGN feedback is primarily responsible for regulating not only SMBH growth but also SF. We argue that scenarios invoking AGNs as the primary “blowing machine” during the intense starburst phase may logically require significant fine-tuning. We then describe the evolutionary path from a starburst to an elliptical galaxy, including the coupled evolution of SF and SMBH growth in the ensuing two sections.

In Section 3, we show that growth of an SMBH during the starburst phase is limited and constitutes a small fraction of overall SMBH consumption. The physical reason is that this phase is over-supplied with gas such that only a very small central disk is gravitationally stable (Toomre parameter $Q > 1$) for gas accretion onto the SMBH, while all other regions are unstable and more conducive to SF. Since the SF timescale is much shorter than the Salpeter accretion timescale, most of the gas forms into stars. The accreted mass during this phase is probably limited to a few percent of the final SMBH mass. In Section 4, we point out that energy or momentum feedback from SF is at least as competitive as that from the AGN during the starburst phase. Therefore, SF is largely responsible for blowing most of the last patch of gas away to end the starburst phase. In short, during the starburst phase, the SMBH does not grow significantly and does not play the leading role in quenching the SF.

In Section 5, we show that most of the growth of the SMBH occurs in the ensuing post-starburst period, when the bulge/elliptical galaxy is largely in place and SF enters “passive” evolution. The fuel for this primary growth phase is
provided by the gas recycled back into the interstellar medium (ISM) from aging bulge stars, proposed earlier by Norman & Scoville (1988) in the context of a central stellar cluster and stressed recently by Ciotti & Ostriker (2007) in the context of elliptical galaxies. It provides a relatively “diffuse” (compared to the starburst phase) but steady gas supply that, as we show, is ideal for feeding an SMBH via an accretion disk. Meanwhile, SF is the dominant mode for gas consumption in the outer region because the accretion is unstable to fragmentation there even in this phase. Self-regulation is at work for the growth of the SMBH during this period and is provided by much more robust (compared to energy feedback) radiation pressure-induced momentum. The amplitude and slope of the resultant SMBH–bulge relation with this self-regulation is consistent with observations.

In this model, the entire evolution from the onset of the starburst, due to a gas-rich merger or some significant event that drives a large amount of gas into the central region within a short period of time, to becoming a quiescent elliptical galaxy (or a bulge of a future spiral galaxy) consists of three distinct periods, as summarized in Section 6.1 and in Figure 2. (1) “Starburst Period”: a merger of two gas-rich spiral galaxies or some other significant event induces a starburst that lasts about $10^7–10^8$ yr. The SMBH grows modestly during this period. The feedback energy/momentum from the starburst, i.e., supernova, drives the last patch of gas away and helps shut down SF. (2) “SMBH Prime Period”: several hundred million years after the end of the starburst, aging low- to intermediate-mass stars, now in the form of red giants and other post-main-sequence states, start to return a substantial fraction of their stellar mass to the ISM. The SMBH accretion is mostly supply-limited in most of this period, except during the first several hundred million years or so, and lasts for the order of a gigayear. Because the rate of gas return from stars diminishes with time, the Eddington ratio of the SMBH decreases with time and the SMBH spends most of the time during this period at low Eddington ratio ($\lesssim 10^{-3}$). The SMBH growth is nearly synchronous with SF from recycled gas during this period. The accompanying star formation rate (SFR) is quite substantial, roughly $\sim$($5–10$)($M_\odot$/yr$^{11}$ $M_\odot$/yr$^{-1}$)($t$/1 Gyr)$^{-1.3}$ $M_\odot$ yr$^{-1}$, where $t$ is time in Gyr and $M_\odot$ is the stellar mass of the elliptical galaxy formed during the starburst (at $t = 0$). The duration of this phase depends sensitively on the lower cutoff mass of the initial mass function (IMF). (3) “Quiescent Elliptical Galaxy”: several gigayears after the end of the starburst the elliptical galaxy is now truly red and dead—the gas return rate is now negligible, so both accretion to the central SMBH and residual SF have ceased. It is possible, at least for an elliptical galaxy that is not too massive (i.e., $M_{\text{tot}} \lesssim 10^{12} M_\odot$), that it may grow a disk. The feeding of the central SMBH in the bulge of the spiral galaxy during this period is no longer contributed by aging stars, but rather by occasional objects (molecular clouds, stars, etc.) that happen to be on some plunging orbits due to secular or random events.

We present some predictions and implications of this model in Sections 6.2–6.9, followed by conclusions in Section 7. Where comparisons can be made between the predictions of the model and observations, they appear to be in good agreement. Some additional predictions could provide further tests of the model.

2. AGNs CANNOT REGULATE STAR FORMATION DURING A STARBURST

While the subsequent sections of quantitative physical analysis are independent of statements made in this section, we shall argue for the assertion in the title of this section with logic, in the hope of being able to provide some conceptual clarity to the role of AGN feedback on SF during the starburst phase. The starting point of the evolutionary sequence is a starburst. It may be triggered by a major merger of two gas-rich galaxies or by other significant events that channel a large amount of gas into the central region in a short period of time. Consider that an event causes a large amount of gas of mass $M_{\text{gas}}$ to land in the central region. Physical processes then operate on the gas to produce a starburst, accompanied by some growth of the central SMBH, along with some associated feedback from both. Extreme events of this kind may be identified with observed ultraluminous infrared galaxies (ULIRGs, e.g., Sanders et al. 1988) or submillimeter galaxies (SMGs, e.g., Chapman et al. 2005). Theoretical models (e.g., Silk & Rees 1998; Hopkins et al. 2006) have proposed that feedback from AGNs is responsible for the regulation of SF and SMBH growth so as to produce the observed Magorrian et al. (1998) relation, where the ratio of the final SMBH to bulge stellar mass is $M_{\text{BH}}/M_{\text{BG}} \sim 2:1000$. We shall now re-examine this case.

Consider how the inflamed gas may be partitioned. Mass conservation requires $M_{\text{BH}} + M_{\text{BG}} + M_{\text{out}} = M_{\text{gas}}$, where $M_{\text{out}}$ is the amount of gas that is blown away from the bulge. Clearly, only a very small fraction of the initial gas $M_{\text{gas}}$ can possibly end up in the central SMBH, i.e., $f_{\text{BH}} = M_{\text{BH}}/M_{\text{gas}} \ll 1$. Let us assume that the reason for a very small $f_{\text{BH}}$ is because the feedback from the central SMBH prevented its own further growth during this phase. Since SMBH masses are observed to span a very wide range, it must be that this purported SMBH feedback process that regulates its own growth is galaxy specific, i.e., dependent on at least some physical variables characterizing the galaxy. A usual and reasonable assumption (which we are not advocating at the moment) for this is that either the gravitational potential well of the bulge or of the total halo determines the final SMBH mass, in coordination with its feedback.

Does SMBH feedback dominate that of the starburst in terms of regulating both SMBH growth and the starburst? While we will show later (in Section 3) that the answer is largely no to regulating the starburst at least, we assume that the answer is yes to both for the sake of continuing the present thought experiment. The simplified sequence of events then plays out more or less as follows: the central SMBH accretes gas and builds up its feedback strength until its mass has reached the observed value, then blows away all the remaining gas and both SMBH accretion and SF stop abruptly. What might have happened to SF during all this time before the gas was blown away? There are three possible scenarios. Scenario 1: the SMBH accretion is so competitive and quick that most of the gas is blown away by the SMBH feedback before much SF has occurred. Of course, that cannot have happened, because that would be inconsistent with the observed $M_{\text{BH}}$–$M_{\text{BG}}$ relation.

Scenario 2: SF proceeds at a pace that is in concert with the SMBH feedback such that by the time that $M_{\text{BH}} = 0.002$ ($M_{\text{gas}} - M_{\text{out}}$), the amount of stars formed is equal to $M_{\text{BG}} = 0.998$ ($M_{\text{gas}} - M_{\text{out}}$); the rest of the gas of mass $M_{\text{out}}$ was blown away by the feedback from the SMBH. This scenario is designed to match the observed $M_{\text{BH}}$–$M_{\text{BG}}$ relation. What remains undetermined is how large $f_{\text{out}} = M_{\text{out}}/M_{\text{gas}}$ is. Is it close to 1 or 0? In the case where $f_{\text{out}} \sim 0$, because (1 ~ $f_{\text{out}}$) is a small number, there is no particular preferred value for it. The potential well created by the eventual bulge stars would be much shallower than the original one already created by the residing gas. In other words, the SMBH only knew the potential well of
the original gas, and the number of stars the SMBH decides to allow the bulge to have would be rather arbitrary. If one argues that it is the potential well of the total halo mass that matters, the SMBH still did not know how to let SF take place at such a rate that we have the very tight observed $M_{\text{BH}} - M_{\text{BG}}$ relation for the bulge region. Thus, this case also appears to require much fine-tuning. Besides, if $(1 - f_{\text{out}})$ is too small, the bulge will be too small compared to what is observed.

The opposite case with $f_{\text{out}} \ll 1$ is at least substantially more stable since a large fraction of the original gas formed into stars before the remainder of the gas was blown away. In this case, the SMBH would “know better” the gravitational potential well eventually sustained by bulge stars because it is not too far from that created by the initial gas. In this case, how did the SMBH know when to blow away the remaining gas left over from SF and SMBH accretion? Should the SMBH blow away the gas when $f_{\text{out}} = 0.90$ (an arbitrarily picked number for illustration purpose) or should it wait a bit longer to finally blow away the gas when $f_{\text{out}} = 0.10$? It may require more energy or momentum in the former than in the latter, but that can readily be accommodated by a proportionally increased amount of gas accreted in the vein of feedback from an SMBH providing the required feedback energy or momentum. Since the amount of gas available before $f_{\text{out}} = 0.90$ blown away in this hypothetical case is capable of growing the SMBH to be 900 times more massive than observed and the amount of time available (cosmological scale) is much longer than the Salpeter time, there is no obvious reason why the SMBH cannot grow 10 times (or whatever factor) larger to blow away the gas when $f_{\text{out}} = 0.90$ instead of when $f_{\text{out}} = 0.10$. How the SMBH communicated with the bulge to ration the gas consumption would be a mystery. Thus, even in this case with $f_{\text{out}} \ll 1$, taking it as a given that the SMBH always stands ready to provide the necessary feedback, having SMBH feedback to regulate the overall SF in the bulge such that the ratio of the two matches the observation again requires a substantial amount of fine-tuning. Nevertheless, since it is reasonable to expect that the dependence of the outcome, such as the $M_{\text{BH}} - M_{\text{BG}}$ relation, on any proposed feedback processes (including those based on thermal energy deposition near the galaxy center) is a monotonic function of the adopted feedback strength, it should be expected that a solution be found such that the observed $M_{\text{BH}} - M_{\text{BG}}$ relation is obtained, for some chosen value of feedback strength, at least for some narrow range in $M_{\text{BG}}$. But, until there is a clear physical reason or direct observational evidence to support the chosen value of the feedback parameter on which the solution sensitively depends, such an approach remains to be refined. We will provide an alternative, significantly less contrived, quantitative physical mechanism to circumvent this concern of fine-tuning.

3. STARBURST PHASE: MODEST SMBH GROWTH AND SF SHUTDOWN BY STARS

We argued in the previous section that AGN feedback cannot logically play the leading role in regulating SF, in the sense that while some feedback from the SMBH can certainly affect its surrounding gas, there is no particular reason why this could provide a precise (within a factor of a few) rationing mechanism during the starburst phase so as to produce the observed relation between the two. We shall now argue for Scenario 3: during the starburst phase the SF is self-regulated and self-limited, while SMBH growth is modest, does not need regulation, and does not provide significant feedback to SF.

We now give a physical reason for why, even though there is a very large supply of gas in the bulge region during the starburst phase, the SMBH growth is modest. We will make three simplifying assumptions to present a trackable illustration without loss of generality. We (1) assume for the regions of interest a geometrically thin Keplerian disk dominated by the SMBH gravity (at least at the radii of interest here) is in a steady state, meaning the accretion rate (Frank et al. 1992),

$$M = 3\pi \upsilon_{g} \Sigma_{g} (1 - (r_{\text{in}} / r)^{1/2})^{-1} \sim 3\pi \upsilon_{g} \Sigma_{g}, \quad (1)$$

is constant in radius $r$ and time, where $\Sigma_{g}$ is the gas mass surface density and $\upsilon$ is the viscosity; the last equality is valid because the radii of interest here are much larger than the radius of the inner disk $r_{\text{in}}$; note that it is inevitable to form a disk in the central region given the rapid cooling and finite angular momentum. We (2) adopt the $\alpha$-disk viscosity (Shakura & Sunyaev 1973):

$$\upsilon = \alpha c_{g}^{2} \Omega^{-1}, \quad (2)$$

where $\alpha$ is a dimensionless viscosity constant for which the magnetorotational instability process (Balbus & Hawley 1991) provides a physical and magnitude-wise relevant value; $c_{g}$ is the sound speed and $\Omega$ is the angular velocity (equal to epicyclic frequency for a Keplerian disk). The Toomre $Q$ parameter of the gas disk can be obtained from Equations (1) and (2):

$$Q \equiv \frac{c_{g} \Omega}{\pi G \Sigma_{g}} = \frac{1}{3^{1/2} \pi^{3/2} \alpha^{1/2}} \left( \frac{M}{M_{\text{BH}}} \right)^{1/2} \left( \frac{G^{-1/4} M_{\text{BH}}^{5/4}}{\Sigma_{g}^{3/2} c_{g}^{9/4}} \right), \quad (3)$$

where $G$ is a gravitational constant. The slope of the surface brightness profiles of the inner region of the observed power-law elliptical galaxies, which are assumed to be the product of the starbursts resulting from the gas-rich galaxy mergers, has a value concentrated in the range $-0.8 \leq -0.5$ (e.g., Faber et al. 1997; Kormendy et al. 2009), reproduced in merger simulations (e.g., Hopkins et al. 2009). Presumably, the initial gas density profile is similar to the final observed stellar density profile in the inner regions. For ease of algebraic manipulations, we (3) assume the de Vaucouleurs mass surface density profile (with a half-mass radius $r_{c}$) but with the inner region at $r \leq r_{\text{p}} \equiv 0.07r_{c}$ modified to be a Mestel disk as

$$\Sigma_{g}(r) = \Sigma_{0} \left( \frac{r}{r_{0}} \right)^{-1} \quad \text{for} \quad r \leq r_{\text{p}}, \quad (4)$$

where $\Sigma_{0}$ is the normalizing surface density at some radius $r_{0}$. We will only be dealing with region $r \leq r_{\text{p}}$; the notional nuclear velocity dispersion of the system without the central SMBH at $r \leq r_{\text{p}}$ is related to $\Sigma_{0}$ and $r_{0}$ by

$$\sigma_{n}^{2} = \pi G \Sigma_{0} r_{0}. \quad (5)$$

Subsequent results do not sensitively depend on the exact slope. The total mass of such a hybrid profile is equivalent to a truncated isothermal sphere with a truncation radius of $2r_{c}$ and velocity dispersion on galactic scales of $\sigma_{g}$ such that

$$\sigma_{n} = 1.55 \sigma_{g}. \quad (6)$$

Since the dynamic time, say at 1 kpc for a $200 \text{ km s}^{-1}$ bulge being only $5 \times 10^{6}$ yr, is much shorter than the Salpeter time, it
is appropriate to assume that the gas disk is assembled instantaneously with respect to accretion to the SMBH when infalling gas lands on the disk. Combining Equations (3), (4), (5), and (6), we rewrite $Q$ as

$$Q = 0.32 \alpha_{0.01}^{-1/2} \epsilon_{0.1} \sigma_200^3 \sigma_200^3 \sigma_200^3, \tag{7}$$

where $\alpha_{0.01} = m_0.01$, $\epsilon = 0.1 \epsilon_{0.1}$ is the SMBH radiative efficiency, $I_E$ is the Eddington ratio, $M_8 = M_8/10^8 M_\odot$, $\sigma_200 = \sigma_200/200 \text{ km s}^{-1}$, and $r_{pc} = r/1 \text{ pc}$. The value of $\alpha$ is quite uncertain, possibly ranging from $10^{-4}$ to 1 (e.g., Hawley et al. 1995; Brandenburg et al. 1995; Stone et al. 1996; Armitage 1998; Gammie 2001; Fleming & Stone 2003; Fromang & Papaloizou 2007). Setting $Q$ in Equation (7) to unity defines the disk stability radius

$$r_Q = 0.22 \alpha_{0.01}^{-2/3} \epsilon_{0.1} \sigma_200^3 \sigma_200^3 \sigma_200^3. \tag{8}$$

within which $Q > 1$ and the disk is stable to gravitational fragmentation, and outside which $Q < 1$ and the disk is subject to gravitational fragmentation to form stars, supported by both simulations (e.g., Gammie 2001; Rice et al. 2003) and circumstantial observational evidence of the existence of a stellar disk at small Galactic radius ($\sim 0.1$ pc) (e.g., Levin & Beloborodov 2003; Paumard et al. 2006). The demarcation value of $Q$ between stability and fragmentation does not appear to be qualitatively different even if the disk is under strong illumination (e.g., Johnson & Gammie 2003), as might happen to a nuclear gas disk in the starburst phase. The disk mass within $r_Q$ is

$$M_0 = 9.8 \times 10^4 \alpha_{0.01}^{-2/3} \epsilon_{0.1} \sigma_200^3 \sigma_200^3 \sigma_200^3 M_\odot. \tag{9}$$

This is the accreted mass out of the entire bulge region (note that some of the outer regions are more randomly motion-supported). This conclusion is in good agreement with Goodman (2003), who employs somewhat different assumptions than in this study in that he assumes local energy balance, while we impose the observationally inferred inner density profile to be self-consistent; the good agreement suggests that this result is quite robust, insensitive to assumptions made. Taking a cue from our own Galaxy, if we assume that the initial SMBH mass of each of the two merging spiral galaxies of mass $\sim 10^{12} M_\odot$ is $2.5 \times 10^6 M_\odot$ for a spiral galaxy of velocity dispersion of $200 \text{ km s}^{-1}$ we see that the amount of mass that could be readily accreted according to Equation (9) using $M_8 = 0.05$ is $6.7 \times 10^4 \alpha_{0.01}^{-2/3} \epsilon_{0.1} \sigma_200^3 \sigma_200^3 \sigma_200^3 M_\odot$. Note that the final SMBH mass for such a system is $\sim 1.3 \times 10^8 M_\odot$ (Tremaine et al. 2002) if we were to match the observations.

It is possible that the mass accreted to the SMBH may be larger than that indicated by Equation (9) due to replenishment. Replenishment of low angular momentum gas during the starburst phase may be possible in two ways: (1) through orbital decay of outer disk gas or (2) through direct infall of low-$J$ gas from outer regions. We will show below that process (1) does not significantly increase the accretable mass. Process (2) is probably unavoidable to some extent but is unlikely to be frequent enough to be significant for the following reasons: All the low angular momentum infalling gas falls into the inner regions initially according to its respective specific angular momentum driven by the torque of the trigger event (e.g., a merger or some other significant torquing event). To replenish low angular momentum gas directly to the central region some frequent and significant torquing events during the starburst phase are needed. It seems unlikely that such events will be frequent enough to be able to reach the observed final SMBH mass: about $\sim 100–1000$ replenishments will be required. One might approximately equate the number of replenishments (i.e., significant disturbance) to the number of generations of stars formed during the starburst phase (by assuming that each generation of SF manages to redistribute the angular momentum of a significant fraction of the gas), which is unlikely to be close to $\sim 100–1000$. In summary, taking into account possible additional accretion due to some replenishment and giving the benefit of the possibility of $\alpha_{0.01} < 1$, it seems improbable that the SMBH is able to acquire a mass during the starburst phase that would be as much more than 10% of the final value.

At $r \geq r_Q$, the disk is unstable to SF. For SF under the conditions relevant here, both the dynamic and the cooling time are short and do not constitute a significant bottleneck; if they were the only timescale bottleneck, SF would be too efficient. A possible bottleneck for SF is the timescale to rid the cloud of the magnetic flux (assuming the SF clouds are initially magnetically subcritical). The main ionization source in the depth of the molecular cloud cores is cosmic rays (CRs). While the exact ionization rate by CRs is unknown for other cosmic systems, we have some estimate of that for our own Galaxy, $\zeta_{CR,gal} = (2.6 \pm 1.8) \times 10^{-17} \text{ s}^{-1}$ (e.g., van der Tak & van Dishoeck 2000). If one assumes that the CR ionization rate in the starburst is 100 times (modeling a typical ULIRG in this case) that of the Galactic value, considering that the SF rate in ULIRGs is 100–1000 times the Galactic value occurring in a more compact region and that the CR in ULIRGs may be advected out via fast galactic winds (versus slow diffusion in the Galaxy), one may roughly estimate that the ambipolar diffusion time is $7 \times 10^8 \text{ yr}$ at a density of $n \sim 10^9 \text{ cm}^{-3}$ using standard formulae for recombination (e.g., McKee & Ostriker 2007). This estimate is, however, uncertain. We will again look to direct observations for a better gauge. Gao & Solomon (2004) show, from HCN observations, that ULIRGs and LIRGs convert molecular gas at $n \geq 3 \times 10^9 \text{ cm}^{-3}$ at an $\epsilon$-folding timescale of $t_{\phi} \sim 2 \text{ Myr}$, consistent with the above rough estimate. It is clear that the SF timescale is much shorter than the Salpeter time of $4.5 \times 10^8 \text{ yr}$; in other words, when gas is dense and unstable, SF competes favorably with the SMBH accretion with respect to gas consumption. Therefore, most of the gas at $r \geq r_Q$ will be depleted by SF. When the density profile of the disk at $r \geq r_Q$ steepens to be $\Sigma(r) = \Sigma_0 (r/r_Q)^{-5/2}$, where $\Sigma_0$ is the gas surface density at $\sim r_Q$, the disk at $r \geq r_Q$ may become stable again. While continued accretion supplied by gas on the outer disk is likely, albeit at a much lower level, the mass integral is convergent and most of the mass of this outer disk is at $r_Q$ given the density slope even if the entire outer disk at this time is accreted.

Thus, it appears that the amount of gas that is actually accreted by the SMBH during the starburst phase is rather limited. This new and perhaps somewhat counterintuitive conclusion is strongly supported by available observations of ULIRGs. This conclusion is also opposite to most models that rely on SMBH to provide the necessary feedback to regulate SF (e.g., Silk & Rees 1998; Hopkins et al. 2006). Observational evidence is that the SMBHs in ULIRGs and SMGs appear to be significantly smaller (an order of magnitude or more) than what the $M_{BH}$–$M_{NG}$ relation would suggest (e.g., Genzel et al. 1998; Ivison et al. 2000; Ptak et al. 2003; Ivison et al. 2004; Alexander et al. 2005a, 2005b; Kawakatu et al. 2006; Alexander et al.
2008). Nonetheless, it is expected that the AGN contribution in ULIRGs should become relatively more important for larger, more luminous galaxies (see Equation (9)), consistent with observations (e.g., Lutz et al. 1998). Starbursts occurring on rotating nuclear disks/rings in ULIRGs are also supported by circumstantial observational evidence (e.g., Downes & Solomon 1998).

The overall conclusion that SMBH feedback has little effect on the number of stars formed is in agreement with that of DeBuhr et al. (2010), who investigated the radiation-pressure-regulated SMBH feedback in the starburst phase of the merger simulations utilizing a subgrid model for SMBH accretion. One specific common outcome between our calculation and their simulation is that most of the gas formed into stars, regardless of the feedback strength. A notable difference between our calculation and theirs is that their calculation resolution, a gravitational softening length of 47 pc, is significantly larger than \( r_\text{Q} \) (Equation (8)). As a result, it is possible that their simulations do not resolve the small scale that separates stable accretion from an unstable, fragmenting disk, which is crucial to our quantitative conclusion (note that they use the viscosity parameter \( \alpha = 0.05-0.15 \), which is larger than our fiducial value of 0.01, which would yield a still smaller \( r_\text{Q} \); see Equation (9)). Thanks to that difference, we were able to conclude that, even without considering any feedback from the central SMBH, the SMBH during the starburst phase does not grow anywhere to close the observed final mass, because SF can more favorably deplete the gas that may otherwise accrete to the SMBH, whereas they find SMBH masses to be too large even with substantial feedback (note that they use 10 times \( L/c \) radiation pressure force assuming multiple scatterings of each converted FIR photon). It seems likely that their different conclusion may be due to a much higher accretion rate at their resolution scale, which we argue does not reflect the actual accretion onto the SMBH, but rather that the disk is unstable at that scale and mostly forms stars. As we noted in the previous paragraph, observations indicate that the SMBH masses in the starburst phase appear to be smaller than the final values seen in quiescent elliptical galaxies by an order of magnitude, consistent with our conclusion. Substantially higher resolution (a factor of \( \sim 100 \)) simulations may be necessary in order to realistically and more accurately simulate the intricate competition between accretion and SF.

4. A COMPARISON OF FEEDBACK ENERGETICS BETWEEN STAR FORMATION AND SMBH

Having shown the unlikeliness of a substantially growing SMBH during the starburst phase, we now turn to a comparison of the energetics of SMBH and SF to show that feedback from the starburst itself should play the leading role in shutting down or quenching SF, i.e., promptly sweeping away the final portion of the gas where needed.

To avoid any apparent bias against SMBH or a possibly circular argument by the assertion that most of the SMBH growth takes place in the post-starburst phase (as we will show in Section 4), we shall for the moment generously assume that all SMBH growth occurs during the starburst phase to maximize the energy output from the SMBH, when comparing the energetics from the SMBH and the starburst. In Table 1, under the assumption that \( M_{\text{BH}}: M_{\text{BH}} = 2:1000 \), a Salpeter IMF for stars and a radiative efficiency of SMBH accretion of 10%, and energy output from both SF and SMBH in various forms are listed: (1) total radiation energy, (2) ionizing radiation, (3) X-ray radiation in the 2–10 keV band, (4) mechanical energy, and (5) radio jets.

### Table 1

| No. | Form                           | SF                        | SMBH                      |
|-----|--------------------------------|---------------------------|---------------------------|
| (1) | Total radiation                | \( \epsilon_r(\text{rad}) = 7 \times 10^{-4} \) | \( \epsilon_{\text{BH}}(\text{rad}) = 2 \times 10^{-4} \) |
| (2) | Ionizing radiation (\( \geq 13.6 \text{ eV} \)) | \( \epsilon_r(\text{LL}) = 1.4 \times 10^{-4} \) | \( \epsilon_{\text{BH}}(\text{LL}) = 3 \times 10^{-5} \) |
| (3) | X-ray (2–10 keV)               | \( \epsilon_r(2-10 \text{ keV}) = 9 \times 10^{-6} \) | \( \epsilon_{\text{BH}}(2-10 \text{ keV}) = 5 \times 10^{-6} \) |
| (4) | Mechanical                     | \( \epsilon_r(\text{SN}) = 1 \times 10^{-5} \) | \( \epsilon_{\text{BH}}(\text{BAL}) = (0.2-2.8) \times 10^{-5} \) |
| (5) | Radio jets                     | \( \epsilon_r(\text{jet}) = 0 \) | \( \epsilon_{\text{BH}}(\text{jet}) = 4 \times 10^{-5} \) |

**Notes.** Under the assumption that \( M_{\text{BH}}: M_{\text{BH}} = 2:1000 \), a Salpeter IMF for stars and a radiative efficiency of SMBH accretion of 10% (Yu & Tremaine 2002), and energy output from both SF and SMBH in various forms are listed: (1) total radiation energy, (2) ionizing radiation, (3) X-ray radiation in the 2–10 keV band, (4) mechanical energy, and (5) radio jets.

It is evident that aside from the energy in the form of radio jets and hard X-rays, SF is at least competitive compared to SMBH. Heating due to hard X-rays from SMBH via metal line or Compton heating affects only the very central region surrounding the SMBH, not over the entire galaxy (Ciotti & Ostriker 2007). Within the physical framework outlined here, most of the SMBH growth occurs post-starburst and radio jets occur at a still later stage in core elliptical galaxies; energy output (or momentum output derived from it) from SF in all relevant forms should dominate that of SMBH. Our argument that radio jets occur at a later stage in galaxy evolution is not at present based on a physical model, but on empirical evidence. Observationally, it appears that all significant radio jets are launched in elliptical galaxies that have flat cores (Balmaverde & Capetti 2006), with very few exceptions that originated in disk galaxies (e.g., Evans et al. 1999; Ledlow et al. 2001) or S0’s (e.g., Véron-Cetty & Véron 2001). None has been associated with elliptical galaxies with an inner power-law brightness profile slope. It has been plausibly argued that power-law elliptical galaxies are produced by gas-rich mergers (we adopt the scenario in which a power-law elliptical galaxy
thick gas disk) subsequently formed from returned stellar gas. Because stars in the inner regions are already mostly rotationally supported, the gas that was shed rains almost “straight down” to land at a location that their specific angular momentum allows, to form a disk. Obviously, going out radially, the rotational support lessens and SF may occur in a three-dimensional fashion, but that does not alter our argument about what happened at small radii. The orientation of the disk is approximately the same as the previous disk out of which stars in the inner regions were formed, since the overall angular momentum distribution of stars has not changed much in the absence of any subsequent intrusions. The most important difference in this new accretion disk, compared to the disk formed during the starburst phase, is that this new disk starts with almost no material, and surface density increases with time gradually on the timescale of hundreds of megayears to gigayears.

To have a better gauge as to how the results obtained depend on the assumed inner density slope, instead of assuming a Mestel disk as is done in Section 3, here we present a more general case assuming the inner density profile of the form

\[ \Sigma_0(r) = \Sigma_0 \left(\frac{r}{r_0}\right)^{-n}, \]

where \( n \approx 0.5\) to 1 (e.g., Faber et al. 1997; Kormendy et al. 2009). For this case Equation (8) is modified, taking into account the gradual change of the gas disk surface density with time, to be

\[ r_Q = \left( \frac{1}{(\pi/3 \sqrt{2})^{4/3} r_Q^{2/3} \Sigma_0(3/2n-2)} \right)^{3/2} \left(\frac{M}{M_{BH}}\right)^{1/3} \left(\frac{f_{rec} f_g}{\alpha}\right)^{2/3} (3/2n-2)^{1/3} \left(\frac{\Sigma_0}{\Sigma_0(3/2n-2)}\right)^{2/3} \left(\frac{G}{M^{1/3}}\right)^{2/3n}, \]

where \( f_{rec} \) is the total fractional stellar mass that recycles back to ISM and \( f_g(t) \) is the fraction of recycled gas that has returned by time \( t \) (out of the fraction \( f_{rec} \)). The process of SMBH accretion in this case proceeds as follows: the SMBH will accrete all the gas within its Bondi radius \( r_B \) over some period of time, as long as \( r_Q \geq r_B \), where \( r_B \) is defined as

\[ r_B = \frac{G M_{BH}}{\sigma_n^2}, \]

with \( \sigma_n \) being the velocity dispersion of the inner region of the bulge (\( r \leq 20 \) pc or so for \( M_{BH} = 10^8 M_\odot \)). For the moment, we ignore any feedback effect from the SMBH. Since \( r_B \) grows with time and \( r_Q \) decreases with time with increasing \( f_g \) for \( r > r_B \) which has been accumulating gas, the condition \( r_Q \geq r_B \) may be violated at some time \( t \), at which point the SMBH is cut off from the gas supply at its Bondi radius. The SMBH will subsequently grow by consuming the final patch of gas on the disk within its Bondi radius. Before the condition \( r_Q \geq r_B \) is reached, the recycled gas that has landed outside (time varying) \( r_B \) continues to accumulate (some of the accumulated gas possibly forms stars). Using Equations (11) and (12) we find the turning point \( r_Q = r_B \) is reached when

\[ f_{rec} f_g = \frac{2 - n}{2} \]

with the disk mass within \( r_Q = r_B \), i.e., SMBH mass, being

\[ M_F = \frac{3(2-n)^3}{8} \left( \frac{M}{M_{BH}} \right)^{-1} \alpha \sigma_n^2 \frac{G}{M}. \]
From Equation (13) we see that \((2 - n)/(2f_{\text{rec}}) > 1\) for \(n = [0.5-1]\). Thus, we simply correct Equation (14) by a factor of \(2f_{\text{rec}}/(2 - n)\) to finally arrive at

\[
M_{\text{BH}} = \frac{3(2-n)^2}{4}\left(\frac{M}{M_\odot}\right)^{-1}\frac{f_{\text{rec}}\alpha\sigma_n^3}{G} = 1.9 \times 10^6 (2-n)^2\alpha_{0.01}l_E^{-1}\epsilon_{0.1}\left(\frac{\sigma_n}{200 \text{ km s}^{-1}}\right)^3 M_\odot, \tag{15}
\]

with the radius when \(r_Q = r_B = r_{BQ}\) being

\[
r_{BQ} = 34(2-n)^3\alpha_{0.01}l_E^{-1}\epsilon_{0.1}\left(\frac{\sigma_n}{200 \text{ km s}^{-1}}\right) \text{pc.} \tag{16}
\]

Equations (15) and (16) suggest that the SMBH accreted the recycled gas at \(r \leq 20(M_{\text{BH}}/10^8 M_\odot)\) pc or so for \(l_E \sim 1\); it could be substantially larger for smaller \(l_E\). The reason that the accretible mass is so much larger during this period than during the starburst phase is because the accretion disk in this period is replenished continuously at a moderate rate such that it is stable within a much larger radius than in the case of the starburst phase with a much thicker (surface density-wise) disk. Equation (15) resembles the observed \(M_{\text{BH}}-\sigma\) relation (Tremaine et al. 2002). We argue that the resemblance is deceptive, in a general sense, because it hinges on a value of \(\alpha \sim 0.1\) and \(l_E \sim 1\). As we mentioned earlier, the currently allowed value of \(\alpha\) could range from \(10^{-4}\) to 1, and at the moment we do not know at what value nature has picked to grow her SMBHs. In light of this situation, using Equation (15) to declare victory is premature. However, Equation (15) does suggest that there is enough material and time to grow the SMBH to the observed value during the post-starburst phase. This is in stark contrast with the starburst phase when there is not enough accretable matter even if one pushes the viscosity value to the limit (see Equation (9)).

A scenario where the allowed range of viscosity value is limited to one side, i.e., \(\alpha\) is allowed to have values greater than say 0.01, is much less fine-tuned. In this case, some self-regulation for the SMBH growth will be necessary. This self-regulation for the SMBH growth is indeed achievable during the post-starburst phase, as we will now describe. The total amount of radial momentum that radiation pressure of the SMBH may exert on the surrounding gas is \(\epsilon c M_{\text{BH}}\) (this is likely a lower boundary, neglecting the possibility of multiple scatterings of photons in the optically thick regime). Equating \(\epsilon\beta c M_{\text{BH}}\) to \(f_{\text{rec}} M_{\text{BG}}(1 - f_s)v_{\text{esc}}\) (that is, the momentum of the driven-away gas escaping the galaxy) gives

\[
\frac{M_{\text{BH}}}{M_{\text{BG}}} = \frac{f_{\text{rec}} v_{\text{esc}}}{\epsilon\beta c}(1 - f_s) = A \frac{2}{1000}\sigma_{200}, \tag{17}
\]

where \(A\) is

\[
A = \frac{(f_{\text{esc}}/0.15)(1 - f_s)}{(1 + (f_{\text{esc}}/0.15)f_s)\beta\eta_{0.1}/2\sigma}, \tag{18}
\]

where \(f_s\) is the fraction of recycled gas that subsequently reformed into stars and \(v_{\text{esc}}\) is the escape velocity (for an isothermal sphere truncated at virial radius \(r_v\), \(v_{\text{esc}}(r)/2\sigma = (1 + \ln(1 + r_s/r))/2\) at radius \(r_s\), \(\beta\) is the fractional solid angle that absorbs the radiation from the SMBH, and the term \((1 + f_{\text{esc}}f_s)\) takes into account additional stars added to the bulge stellar mass formed from the recycled gas. Of the parameters in

Equation (18), \(f_{\text{esc}} = 0.15\) is reasonable taking into account that about the half of mass return occurring at early times by Type II supernovae can escape without additional aid, the radiative efficiency of \(\epsilon = 0.1\) is consistent with observations (Yu & Tremaine 2002; Marconi et al. 2004), some fraction of the recycled gas forming into stars is probably unavoidable since some gas with column density greater than Compton column will slip through radiation pressure (see the discussion below), \(f_s\) also includes the (possibly very large) amount of gas at large radii that would not have accreted onto the SMBH in the first place even in the absence of any feedback (e.g., molecular clouds on the Galactic disk are not being fed to the Galactic center SMBH in a consistent fashion), and the factor \(\eta\) (greater than one) takes into account additional stars that are not formed from the postburst event. Overall, considering all these balancing factors, a value of \(A\) of order unity seems quite plausible. Figure 1 plots the relation between \(M_{\text{BH}}\) and \(\sigma_{\text{BG}}\) predicted by Equation (17) using \(A = 1\). It is clear that this provides a very good fit to the observed data. A similar scaling relation to Equation (17) was derived based on a radio jet feedback mechanism (Soker & Meiron 2011).

A similar scenario of linear momentum feedback from AGN radiation pressure was considered by Silk & Nusser (2010) to possibly produce the observed \(M_{\text{BH}}-\sigma_{\text{BG}}\) relation during the starburst phase, but they conclude that the radiation pressure is insufficient by an order of magnitude to be able to blow away the unwanted gas. The magnitude of the radiation pressure and escape velocity requirement considered here are the same as theirs. The difference is that here the amount of gas that needs to be regulated in the post-starburst phase is nearly a factor of 10 lower, and further allowance for SF from the recycled gas makes it possible that the radiation pressure from the central AGN may be adequate to self-regulate the SMBH growth so as not to overgrow it.

We note that Equation (17) would work without much variation if the gas that is blown away is uniformly distributed. The recycled gas is expected to be nonuniform. Even if it were initially uniform, thermal instabilities likely make the distribution nonuniform. Given that, we elaborate further on Equations (17) and (18) and the physical processes of radiation-pressure-driven
winds. Some distinction may be made between about 1/3 of the total solid angle where UV and other photons are directly seen from AGNs and the other $\beta \sim 2/3$ of the solid angle that has a nearly Compton thick or thicker obscuring screen, most of which probably stems from the so-called molecular torus (e.g., Risaliti et al. 1999). For every $\Delta M_{\text{acc}}$ of mass accreted, roughly $\epsilon_{\gamma}/t_{\text{esc}} \Delta M_{\text{acc}} = 1000 \times (t_{\text{esc}}/300 \text{ km s}^{-1})^{-1} \Delta M_{\text{acc}}$ of mass rained down by aging stars could be driven away by the radiation momentum from the AGN. In the 1/3 opening solid angle some portion of the radiation-pressure-driven winds will be accelerated to high velocities, perhaps in a fashion similar to what is seen in simulations (e.g., Kurosawa & Proga 2009), observationally manifested as broad emission or absorption lines as well as outflows seen in narrow lines (e.g., Crenshaw et al. 2003; Greene et al. 2011).

A significant fraction of the material may be accumulated in the remaining $\beta \sim 2/3$ of the solid angle (i.e., Type II AGNs), including the recycled gas that comes from the other 1/3 solid angle that is too heavy to be accelerated away “on the fly” by the radiation pressure. In this 2/3 of the solid angle, high-velocity winds radially exterior to the molecular torus are unlikely given the heaviness (i.e., low opacity) of the molecular torus. We discuss some of the physics here. To gain a more quantitative understanding, a look at some observed properties of the torus is instructive. Jaffe et al. (2004) measured the radius and height of the molecular torus of NGC 1068 to be 1.7 pc and 2.1 pc, respectively. The mass of the SMBH in NGC 1068 is $(8.3 \pm 0.3) \times 10^8 M_\odot$ (e.g., Marconi & Hunt 2003). If we extrapolate to a $10^9 M_\odot$ SMBH assuming that the location and height of the molecular torus is proportional to the SMBH mass, we have a surface area of the torus equal to 3200 pc$^2$ at an SMBH-centric radius of 20 pc. If we assume that the column density of the molecular torus is $10^{24}$ cm$^{-2}$ (e.g., Risaliti et al. 1999), its total mass is then $2 \times 10^7 M_\odot$. The dynamic time at 20 pc is $10^7$ yr. An SMBH of mass $10^9 M_\odot$ accreting at the Eddington rate would grow a mass of $10^5 M_\odot$ in $10^5$ yr, while the overall rate of gas return would be $\sim 2 \times 10^5 M_\odot$ over the entire bulge during that period. Thus, the abundant gas supply rate suggests that the necessary (not sufficient) condition for a near “steady” state is met such that the molecular torus may be kept roughly invariant with time, with the rate of driven-away gas by radiation pressure plus that of gas forming into stars equal to the rate of gas return from aging stars. Given the short SF timescale of the very dense gas in the molecular torus, it would be unavoidable that SF would occur there (as well as some regions exterior to it). This “lightens up” the torus to the extent that it may be pushed away by the radiation pressure, when the condition that the deposited radiation momentum divided by the accumulated mass exceeds the escape velocity (assuming, in the absence of radiation pressure, the torus would just be in a bound circular orbit). In this sense the radiation momentum from the SMBH serves to retard gas supply to accretion from the torus to let SF take over and have it mostly depleted. In combination with the analysis in the preceding paragraph, it seems physically plausible that radiation pressure and depletion of gas by SF is able to jointly reduce and regulate the amount of gas that feeds the central SMBH. Given that the overall margin is on average quite thin (i.e., $A \sim 1$ in Equation (18)), it is likely that there are significant variations in $A$, perhaps up to a factor of a few.

In the one-dimensional simulations of Ciotti & Ostriker (2007) for an elliptical galaxy, the SMBH growth appears to be intermittent. The intermittency in their simulations was caused by a hot X-ray-heated bubble that prevents continued gas accretion until it bursts, which is then followed by another accretion episode, and so on. We suggest that the Rayleigh–Taylor instability on the shell enclosing the X-ray bubble may prevent the X-ray bubble from inflating, as suggested by recent two-dimensional simulations by Novak et al. (2011). It is reasonable to assume that shell fragmentation in three dimensions is still more pronounced to allow continued deflation of a notional X-ray bubble. Observationally, the lack of significant X-ray emission from the circumnuclear region in power-law elliptical host galaxies of AGNs, which we argue are the post-starburst galaxies we consider here, supports the picture that the hot bubble is not robust (e.g., Pellegrini 2005). In the absence of a hot X-ray bubble guarding the SMBH, we suggest that the recycled gas from aging stars is able to reach the disk and the accretion, with self-regulation argued above, is quasi-steady without the major flares of magnitude seen in one-dimensional simulations. As we will show later, a steady declining accretion rate proportional to the gas return rate provides a much better match to at least two observations. (1) The observed early-type host galaxies of AGNs are mostly in the green valley of the galaxy color–luminosity diagram with a small fraction in the red sequence (Section 6.2) (e.g., Salim et al. 2007; Silverman et al. 2008; Hickox et al. 2009; Schawinski et al. 2010), but very few in the blue cloud, which would have been the case if AGN flares were accompanied by starbursts (Ciotti & Ostriker 2007). (2) The observed AGN accretion rate for early-type galaxies in the local universe displays a power-law distribution with an amplitude and decay rate (Kauffmann & Heckman 2009) that are expected from the nonflare scenario proposed here. This indicates that bursty AGN accretion, while quite possible and sometimes perhaps unavoidable, is probably not the dominant mode. It is currently a challenge but will be of great value to carry out three-dimensional high-resolution simulations to more accurately quantify this outcome.

6. MODEL PREDICTIONS AND DISCUSSION

We have presented a physically motivated picture for the coevolution of galaxies and SMBH starting with a triggered starburst. Let us now summarize the entire evolution in Section 6.1 and then give an incomplete list of implications and predictions in Sections 6.2–6.9 to be qualitatively compared/verified with observations.

6.1. Three Distinct Periods of Coevolution of Galaxies and SMBHs

From the onset of a significant central starburst to becoming a quiescent bulge there are three distinct periods, as summarized in Figure 2 for an example merger of two gas-rich spirals each of mass $\sim 10^{12} M_\odot$ that eventually become a power-law elliptical galaxy with a velocity dispersion of $200 \text{ km s}^{-1}$. We stress that the trigger event is not limited to major mergers. This three-stage scenario is not new, and its successes with respect to many observations have been discussed previously (e.g., Granato et al. 2004, 2006; Cirasuolo et al. 2005; Lapi et al. 2006; Lamastra et al. 2010). The new theoretical element here is the primary growth of an SMBH in the post-starburst phase, which is reflected in the color and other properties of AGN hosts and, as we will show, is in remarkable accord with the latest observations, in contrast with the conventional scenario where SMBH growth primarily occurs during the starburst phase. The time boundaries between different consecutive phases (three
SMGs being a result of major mergers of massive gas-rich gas. However, irrespective of the size of the starburst event, the timescales involved, being largely due to the physics of the stellar interior and accretion timescale, remain the same.

1. **Starburst Period.** This phase is triggered by some event. The SMBH grows modestly during this period to possibly attain a mass that is up to the order of 10% of its final mass. This phase lasts about $10^7$–$10^8$ yr for typical starbursts, and the host galaxies during this phase are in the blue cloud in the luminosity–color diagram. The feedback energy/momentum from the starburst, i.e., supernova, drives the last patch of gas away and shuts down SF, if needed. In other words, the starburst is self-regulated rather than regulated by the central AGN during this period.

2. **SMBH Prime Period.** Several hundred million years after the end of the starburst, aging low- to intermediate-mass stars, now in their post-main-sequence phases, start to return a substantial fraction of their stellar mass to the ISM. The SMBH accretion is fueled by this recycled gas lasting for the order of a gigayear. The growth of the SMBH is self-regulated, readily provided by the radiation pressure from the AGN. The host galaxies during this period start out light blue or in the “green valley” and migrate to the “red sequence.” Because the rate of gas return from stars diminishes and SMBH mass grows with time, the Eddington ratio of the SMBH decreases with time. The SMBH growth is synchronous with SF from recycled gas during this period. The accompanying SFR may also be substantial but typically does not constitute a starburst during this period. The entire duration of this phase depends sensitively on the lower cutoff mass of the IMF—a sensitive and powerful prediction of this model.

3. **Quiescent Bulge.** Several gigayears after the end of the starburst the bulge is now truly red and dead. The gas return rate is now negligible so both accretion to the central SMBH and residual SF have ceased. It is possible that a disk is grown later around the bulge. The feeding of the central SMBH in the bulge of the spiral galaxy during this period is no longer by overhead material from aging stars, but is by occasional objects that happen to be on some plunging orbits that are disrupted by the SMBH and form a short-lived accretion disk. Candidate objects may include molecular clouds, some tidally disruptive stars, or gas streams. Significant disturbances or torques, such as minor mergers and galactic bars, could provide the necessary drivers for some more consistent accretion events. How is a red and dead bulge with a hot atmosphere able to remain SF free? This is a major topic in its own right and is beyond the scope of the current paper but will be addressed in a future paper.

6.2. Some “Obvious” Implications of the Model

There are some unambiguous discriminating signatures of this model that are shown in Figure 2. We highlight several here.

1. Starburst and AGN growth are not coeval in this model. The AGN does not regulate the starburst, which is consistent with observations (e.g., Schawinski et al. 2009; Kaviraj 2009). AGN activities are expected to outline the starburst in agreement with observations (e.g., Georgakakis et al. 2008). These predictions are opposite to those of models that invoke AGN feedback as the primary regulating agent.
2. The apparent requirement of a rapid migration of early-type galaxies from the blue cloud to the red sequence, in order to produce a bimodal distribution in color (e.g., Blanton et al. 2003), is primarily due to the prompt shutdown of SF by stars (i.e., supernovae) at the end of the starburst phase; there is no need to invoke other ingredients, consistent with observations (e.g., Kauffmann et al. 2011). Observationally, there is no evidence that the presence of an AGN is related to the quenching of SF or the color transformation of galaxies (e.g., Aird et al. 2012). This prediction is different from that of models that invoke AGN feedback to quench SF.

3. AGN activities in ongoing starburst galaxies, i.e., buried AGN activities, are not expected to be dominant in this model, in agreement with observations (e.g., Genzel et al. 1998; Ivison et al. 2000; Ptak et al. 2003; Ivison et al. 2004; Alexander et al. 2005a, 2005b; Schweitzer et al. 2006; Kawakatu et al. 2006; Alexander et al. 2008; Veilleux et al. 2009). Note that the above statement is not inconsistent with AGNs/QSOs being associated with galaxies in the process of merging, which may enhance accretion activities in the involved (yet to merge) galaxies (e.g., Bahcall et al. 1997; Hennawi et al. 2010; Smith et al. 2010).

4. The most luminous quasars that accrete with high Eddington ratios occur an order of 100 Myr after the end of the starburst. They may contain substantially more merger signatures as indicated by observations (e.g., Bennert et al. 2008). If one were to identify a population between ULIRGs and more regular QSO hosts in terms of spectral properties, they should show some more signs of tidal interactions that have yet to fully settle in the period after the starburst, also consistent with observations (e.g., Canalizo & Stockton 2001).

5. Low Eddington ratio AGNs that are expected to last an order of a gigayear are not expected to show a close link to major disturbances that trigger the starburst (e.g., mergers), since possible signatures of the merger trigger event have largely been erased over time, consistent with observations (e.g., Grogin et al. 2005; Cisternas et al. 2011). Thus, one does not expect to see merger signatures associated with moderate luminosity AGNs, which is in contrast with AGN feedback-based models where most of the moderate luminosity AGNs are expected to coincide with the starburst.

6. While the green valley morphologically early-type galaxies that host AGNs are the evolutionary link between starburst galaxies (in the blue cloud) and the red elliptical galaxies (on the red sequence), it is useful to distinguish between them and the other class of green galaxies that simply continuously form a modest number of stars (such as our own Galaxy). The former are chronologically immediate successors to starburst galaxies and should be in early-type galaxies, strongly supported by observations (e.g., Salim et al. 2007; Silverman et al. 2008; Hickox et al. 2009; Schawinski et al. 2010), whereas the latter are not a chronologically intermediate class between the blue cloud and the red sequence. The total green galaxy population will be the sum of these two different morphological types, with some obvious implications, such as green galaxies having mixed morphological types with limited merger signatures, consistent with observations (e.g., Mendez et al. 2011). This prediction is in contrast with AGN-feedback-based models where most AGN hosts are expected to coincide with the starburst and a small fraction, mostly the more luminous AGNs (occurring near the end of the starburst phase), are expected to have matured early-type morphologies.

7. While the early-type AGN host galaxies may have similar morphologies and will eventually evolve into inactive elliptical galaxies, the former should have much bluer colors than the latter, consistent with observations (e.g., Sánchez et al. 2004). The basic morphological properties of the host galaxies of the most luminous quasars, corresponding to the most massive SMBHs in the prime growth phase should resemble those of giant elliptical galaxies, consistent with observations (e.g., Dunlop et al. 2003).

8. Because of the expected rate of gas return ($\propto t^{-1.3}$ on gigayear scales) to which both SMBH accretion and SF are proportional and because more powerful AGN accretion occurs closer in time to the preceding starburst, it is expected that more powerful AGNs are hosted by early-type galaxies with younger mean stellar ages, consistent with observations (e.g., Kauffmann et al. 2003; Jahnke et al. 2004).

9. The accompanying SFR of elliptical galaxies may be quite substantial, on the order of $(5 - 10)(M_\odot/10^{11} M_\odot)/t(1 / 1\ Gyr)^{-1.3} M_\odot\ yr^{-1}$. Thus, while most AGN host galaxies have left the blue cloud, a significant fraction of them, especially those hosting luminous AGNs, should still have a substantial SFR, consistent with observations (e.g., Silverman et al. 2009; Shi et al. 2009). It is expected that the incidence of SF signatures (e.g., dust) in the nuclear region should correlate positively with AGN activities for elliptical galaxies, because the strengths of both are proportional to the gas return rate, consistent with observations (e.g., Simões Lopes et al. 2007). These predictions are opposite to AGN-feedback-based models, where SF is expected to be completely quenched after AGN feedback clears out the gas.

6.3. Origin of Two AGN Accretion Regimes

Kauffmann & Heckman (2009) presented an insightful observational result of two distinct regimes of black hole growth in nearby galaxies, along with the apparent implications. They found that star-forming galaxies display a lognormal distribution of Eddington ratios; their interpretation is that in this regime, accretion onto the SMBH is not limited by the supply of gas but by feedback processes that are intrinsic to the SMBH itself. Our model provides the following alternative interpretation for this phenomenon: this lognormal distribution merely reflects two random processes at work. (1) The amount of gas that landed on the stable accretion disk to provide accretion to the SMBH during the starburst phase depends on many “random” variables of the triggering event (in the case of a merger, variables such as merging orbit inclination, velocity, spin alignment, etc.). (2) Observations catch a random moment during the accretion of this gas. The central theorem should then give rise to a lognormal distribution. Another class of possible triggering events for SMBH accretion in star-forming galaxies (e.g., dormant SMBH in the bulge of disk galaxies) is stochastic feeding due to random events, which should also follow a lognormal distribution.

Separately, they find that galaxies with old stellar populations are characterized by a power-law distribution function of Edington ratios and the AGN accretion rate is about 0.3%–1% of the gas return from recycling. In our model, the expected accretion rate is $M_{BH}/(\dot{M}_{gas}) = 1.3 \times 10^{-2} A_{\Delta / \sigma_{000}}$. This expected relation between the SMBH accretion rate and the gas return rate is remarkably close to their observed value. As
Kauffmann & Heckman (2009) pointed out, the power-law distribution is consistent with the recycling gas return rate $\propto t^{-1.3}$ (Ciotti et al. 1991) capped at the Eddington rate with a radiative efficiency of $\epsilon = 0.1$, starting 200 Myr after the end of the starburst. Also indicated along each track are the times in Gyr elapsed since the start of the accretion. Top right panel: the case for a low-mass cutoff for the IMF of 1 $M_\odot$ that has a turnoff lifetime of 2 Gyr. Bottom panels: tracks for the cases in the top panels but in the SMBH mass–luminosity plane.

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

**Figure 3.** Top left panel: evolutionary growth tracks in the SMBH-mass–Eddington-ratio plane of an example SMBH of final mass $10^9 M_\odot$ with two cases of seed black masses of $10^7$ and $10^8 M_\odot$, respectively. A low-mass cutoff for the IMF of 0.92 $M_\odot$ has a turnoff lifetime of 10 Gyr is assumed. We assume that the SMBH accretion rate is proportional to the recycled gas return rate of the form $\propto t^{-1.3}$ (Ciotti et al. 1991) capped at the Eddington rate with a radiative efficiency of $\epsilon = 0.1$, starting 200 Myr after the end of the starburst. Also indicated along each track are the times in Gyr elapsed since the start of the accretion. Top right panel: the case for a low-mass cutoff for the IMF of 1 $M_\odot$ that has a turnoff lifetime of 2 Gyr. Bottom panels: tracks for the cases in the top panels but in the SMBH mass–luminosity plane.

(To be continued...
etc., in order to have a more encompassing analysis. We shall carry out a more detailed analysis with additional parameters in a future study, especially when measurements of both SMBH masses and accretion rates become significantly more precise for a large sample of active galaxies.

6.5. Super solar Metallicity of Accreting Gas

One clear implication is that the accretion gas being shed from aging stars should be very metal rich with supersolar metallicity, in agreement with observations (e.g., Hamann & Ferland 1993), especially to explain the supersolar N/He ratio (e.g., Hamann & Ferland 1999). This is because nitrogen is believed to be secondary in nature, where its abundance scales quadratically with metallicity. The recycled gas that is feeding the SMBH in our model fits the bill most naturally. In addition, the metallicity of accretion gas is not expected to depend on redshift, being intrinsic to stellar evolution, consistent with all accreting gas being very metal-rich at all redshifts, including the highest redshift Sloan Digital Sky Survey quasars (e.g., Fan et al. 2006).

6.6. Relative Cosmic Evolution between Starburst Galaxies and AGNs

Given the modest amount of time delay (several 100 Myr) between the starburst phase and the SMBH prime growth phase, it is unsurprising that one should expect to see nearly synchronous evolution between the starburst and SMBH growth on longer, cosmic timescales, consistent with observations (e.g., Boyle et al. 1988; Nandra et al. 2005).

In the context of the observed cosmic downsizing phenomenon, the downsizing of galaxies (e.g., Cowie et al. 1996; Treu et al. 2005) should thus be closely followed by the downsizing of AGNs (e.g., Barger et al. 2005; Hasinger et al. 2005). There is, however, a very important difference between the two classes in post-peak activities predicted in this model. For starbursts, the shutdown timescale is expected to be about ~100 Myr, whereas for moderate-luminosity AGNs (i.e., Eddington ratio ~10^{-3}) the decay timescale is of the order of ~1 Gyr. With a deep AGN survey that is capable of subdividing early-type galaxies in terms of their masses, one should be able to differentiate between the downturn time of starburst galaxies and that of AGNs hosted by elliptical galaxies at a fixed mass. This prediction would be a strong differentiator between this model and AGN-based feedback models.

6.7. AGN Broad Emission and Absorption Lines

Some of the overhead material raining down onto the SMBH accretion disk from recycled gas from aging low- to intermediate-mass stars provides the material observed as broad emission lines (BELs) and BALs. When some of this gas, probably in the form of some discrete clouds, reaches the inner region of the SMBH (at r < 10^2 r_s, where r_s is the Schwarzschild radius), the clouds will be accelerated by radiation pressure, likely through some absorption lines, to velocities up to 0.1c. These clouds will be the observed BEL and BAL. The fact that only 15%–20% of Type I AGNs have BALs may be indicative of the discrete nature of the clouds, which is not unexpected from discrete stellar remnants or from cooling instabilities.

An advantage of this overhead material is that it naturally provides gas clouds that are presumed to be some ~50° off the equatorial plane, in order not to be obscured by the molecular torus (there are of course BEL and BAL gas clouds at smaller angles, but they are not seen directly). In this model we do not need any additional pressure force to lift the gas off the accretion disk—some of the gas clouds raining down from aging stars will be launched outward before they reach the disk, the physics of which is well known (e.g., Murray et al. 1995).

6.8. Evolution of SMBH Mass Relative to Bulge Mass

Massive elliptical galaxies appear to have increased their masses by 30%–100% in the last 7 Gyr (e.g., Brown et al. 2008). The growth of the elliptical mass is not always expected to be accompanied by corresponding growth in the mass of the central SMBH. For example, the merger of a spiral galaxy without a significant SMBH and an elliptical galaxy would make the final SMBH appear less massive. Given the dependence of $M_{BH}/M_{BG} \propto \sigma \propto (1+z)^{1/2}$ predicted in this model, we predict that the $M_{BH}/M_{BG}$ relation should evolve with a redshift greater than $(1+z)^{1/2}$ for quiescent elliptical galaxies.

6.9. On the Relation between SMBHs and Pseudo-bulges

It is useful to add a note on the difference between classic bulges and pseudo-bulges (Kormendy & Kennicutt 2004) with respect to the central SMBHs in this model. The relation, derived in Equations (17) and (18) and which matches the observed $M_{BH}/M_{BG}$ relation, is dependent on the abundant supply of recycled gas in the inner region. Given the sufficient gas supply from recycled gas, the feedback from the SMBH can then regulate its own growth. This essential ingredient of a sufficient gas supply is consistent with the observed inner slope of classic bulges (e.g., Faber et al. 1997; Kormendy et al. 2009) as we have shown.

The situation would be very different if SF is not as centrally concentrated as it is in classic bulges, for example, in rings (Kormendy & Kennicutt 2004, and references therein) of high angular momentum with a hollow core. In this case, the amount of recycled gas raining down from the innermost region may depend on other unknown factors. For instance, if secular processes act promptly enough, compared to the timescales of stellar gas recycling (~0.1–1 Gyr), to be able to substantially fill the central region with stars initially formed in outer regions, the SMBH may follow the track we described. If, on the other hand, secular processes evolve on longer timescales, the recycled stellar gas would predominantly land in outer regions that do not efficiently accrete to the SMBH, which would in turn not grow substantially. It would seem likely that there may be two trends for pseudo-bulges: (1) there will be large variations in the $M_{BH}/M_{BG}$ relation and (2) SMBH masses may lie below that of the $M_{BH}/M_{BG}$ relation derived from inactive classic elliptical galaxies/bulges, both of which are consistent with independent considerations in the context of a hierarchical structure formation model (e.g., Shankar et al. 2012). Observations, while very challenging, may have already provided some hints of both (Greene et al. 2008).

Moreover, we do not expect any discernible correlation between the SMBH and the galaxy disk or dark matter halo simply because the stars in disks do not affect SMBH growth and the overall dark matter halo, while it indirectly affects the escape velocity that enters Equation (18), does not control the amount of gas that feeds the SMBH. This prediction is consistent with observations (e.g., Kormendy & Bender 2011). In addition, some stellar populations in the outskirts (either on a disk or just at large radii of an elliptical galaxy) of AGN hosts may be unrelated to the preceding starburst and could be substantially different from bulge stars (e.g., Nolan et al. 2001).
7. CONCLUSIONS

We have presented an alternative physical model that has the following characteristics for the coevolution of galaxies and SMBHs. From the onset of a starburst to becoming a quiescent bulge (in the absence of any subsequent significant burst event) there are three distinct periods.

1. Starburst Period. Some significant event induces a starburst that probably lasts about $10^7$–$10^8 \text{yr}$. The SMBH grows modestly during this period to possibly attain a mass that is up to the order of 10% of its final mass. The feedback energy/momentum from the starburst drives the last patch of gas away and shuts down SF.

2. SMBH Prime Period. Several hundred million years after the end of the starburst, aging low- to intermediate-mass stars, now in their post-main-sequence phases, start to return a substantial fraction of their stellar mass to the ISM. Because the rate of gas return from stars diminishes with time, the Eddington ratio of the SMBH decreases with time roughly as $r^{-1/3}$. The SMBH growth is synchronous with SF from recycled gas during this period. The accompanying SFR may also be substantial. The duration of this phase depends sensitively on the lower cutoff mass of the IMF.

3. Quiescent Bulge. On the order of a gigayear after the end of the starburst, the elliptical galaxy is now truly red and dead—the gas return rate is now negligible, so both accretion to the central SMBH and residual SF have ceased. It is possible that a disk may grow around the bulge later. The feeding of the central SMBH in the bulge of the spiral galaxy during this period is not by overhead material from aging stars, but is by occasional objects that happen to be on some plunging orbits to be disrupted by the SMBH and form a short-lived accretion disk. Candidate objects may include molecular clouds or tidally disrupted stars.

In this model, the end of the starburst precedes the onset of prime SMBH growth by an order of 100 Myr. The starburst is responsible for shutting down its own activities; the AGN has returned a substantial fraction of their stellar mass to the ISM. The feeding of the central SMBH in the bulge of the spiral galaxy during this period is not by overhead material from aging stars, but is by occasional objects that happen to be on some plunging orbits to be disrupted by the SMBH and form a short-lived accretion disk. Candidate objects may include molecular clouds or tidally disrupted stars.

For critical and constructive reports. This work is supported in part by grants NNX08AH31G and NNX11AI23G.

REFERENCES

Adelberger, K. L., Steidel, C. C., Pettini, M., et al. 2005, ApJ, 619, 697
Aird, J., Coil, A. L., Moustakas, J., et al. 2012, ApJ, 746, 90
Alexander, D. M., Bauer, F. E., Chapman, S. C., et al. 2008a, ApJ, 632, 736
Alexander, D. M., Brandt, W. N., Smail, I., et al. 2008, ApJ, 683, 1068
Alexander, D. M., Smail, I., Bauer, F. E., et al. 2005b, Nature, 434, 738
Alexander, D. M., Swinbank, A. M., Smail, I., McDermid, R., & Nesvadba, N. P. H. 2010, MNRAS, 402, 2211
Allen, S. W., Dunn, R. J. H., Fabian, A. C., Taylor, G. B., & Reynolds, C. S. 2006, MNRAS, 372, 21
Armitage, P. J. 1998, ApJ, 501, L189
Bahcall, J. N., Kirhakos, S., Saxe, D. H., & Schneider, D. P. 1997, ApJ, 479, 642
Balbus, S. A., & Hawley, J. F. 1991, ApJ, 376, 214
Balmaverde, B., & Capetti, A. 2006, A&A, 447, 97
Barger, A. J., Cowie, L. L., Mushotzky, R. F., et al. 2005, AJ, 129, 578
Bender, R., Kormendy, J., Bower, G., et al. 2005, ApJ, 631, 280
Bennert, N., Canalizo, G., Jungwiert, B., et al. 2008, ApJ, 677, 846
Best, P. N., Kauffmann, G., Heckman, T. M., et al. 2005, MNRAS, 362, 25
Blanton, M. R., Hogg, D. W., Bahcall, N. A., et al. 2003, ApJ, 594, 186
Boyle, B. J., Shanks, T., & Peterson, B. A. 1988, MNRAS, 235, 935
Brandenburg, A., Nordlund, A., Stein, R. F., & Torkelsson, U. 1995, ApJ, 446, 741
Brown, M. J. I., Zheng, Z., White, M., et al. 2008, ApJ, 682, 937
Canalizo, G., & Stockton, A. 2001, ApJ, 555, 719
Chapman, S. C., Blain, A. W., Smail, I., & Ivison, R. J. 2005, ApJ, 622, 772
Choi, E., Ostriker, J. P., Naab, T., & Johansson, P. H. 2012, arXiv:1205.2082
Ciotti, L., D’Ercole, A., Pellegrini, S., & Renzini, A. 1991, ApJ, 376, 380
Ciotti, L., & Ostriker, J. P. 2007, ApJ, 665, 1038
Cirasuolo, M., Shankar, F., Granato, G. L., De Zotti, G., & Danese, L. 2005, ApJ, 629, 816
Cisternas, M., Jahne, K., Inskip, K. J., et al. 2011, ApJ, 726, 57
Cowie, L. L., Songaila, A., Hu, E. M., & Cohen, J. G. 1996, AJ, 112, 839
Crenshaw, D. M., Kraemer, S. B., & George, I. M. 2003, ARA&A, 41, 117
Croton, D. J., Springel, V., White, S. D. M., et al. 2006, MNRAS, 365, 11
DeBuhr, J., Quataert, E., & Ma, C. 2010, MNRAS, 406, L55
Di Matteo, T., Springel, V., & Hernquist, L. 2005, Nature, 433, 604
Dowdes, D., & Solomon, P. M. 1998, ApJ, 507, 615
Dunlop, J. S., McLure, R. J., Kukula, M. J., et al. 2003, MNRAS, 340, 1095
Dunn, J. P., Bautista, M., Arav, N., et al. 2010, ApJ, 709, 611
Elvis, M., Wilkes, B. J., McDowell, J. C., et al. 1994, ApJS, 95, 1
Emonts, B. H. C., Morganti, R., Tadhunter, C. N., et al. 2006, A&A, 454, 125
Evans, A. S., Sanders, D. B., Surace, J. A., & Mazzarella, J. M. 1999, ApJ, 511, 730
Faber, S. M., Tremaine, S., Ajhar, E. A., et al. 1997, AJ, 114, 1771
Fan, X., Strauss, M. A., Richards, G. T., et al. 2006, AJ, 131, 1203
Ferrarese, L., & Merritt, D. 2000, ApJ, 539, L9
Fleming, T., & Stone, J. M. 2003, ApJ, 585, 908
Frank, J., King, A., & Raine, D. (ed.) 1992, Accretion Power in Astrophysics (Cambridge: Cambridge Univ. Press)
Franx, M., & Hubble, E. 2001, ApJ, 554, 11
Frigione, D., & Mignani, R. 2006, MNRAS, 365, 11
Georgakakis, A., Nandra, K., Yan, R., et al. 2008, MNRAS, 385, 2049
Goodman, J. 2003, MNRAS, 339, 937
Granato, G. L., De Zotti, G., Silva, L., Bressan, A., & Danese, L. 2004, ApJ, 600, 580
Granato, G. L., Silva, L., Lapi, A., et al. 2006, MNRAS, 368, L72
Greene, J. E., Ho, L. C., & Barth, A. J. 2008, ApJ, 688, 159
Greene, J. E., Zakamska, N. L., Ho, L. C., & Barth, A. J. 2011, ApJ, 732, 9
Grogin, N. A., Conselice, C. J., Chaitzchistrou, E., et al. 2005, ApJ, 627, L97
Hamann, F., & Ferland, G. 1993, ApJ, 418, 11
Hamann, F., & Ferland, G. 1999, ARA&A, 37, 487
Hasinger, G., Miyaji, T., & Hasinger, G. 2005, A&A, 441, 417
Hawley, J. F., Gammie, C. F., & Balbus, S. A. 1995, ApJ, 440, 742
Heckman, T. M. 2001, in ASP Conf. Ser. 240, Gas and Galaxy Evolution, ed. J. E. Hibbard, M. Rupen, & J. H. van Gorkom (San Francisco, CA: ASP), 345
Hennawi, J. F., Myers, A. D., Shen, Y., et al. 2010, ApJ, 719, 1672
Hickox, R. C., Jones, C., Forman, W. R., et al. 2009, ApJ, 696, 891
Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2006, ApJS, 163, 1
