Which came first: supermassive black holes or galaxies? Insights from JWST

Joseph Silk, Mitchell Begelman, Colin Norman, Adi Nusser, Rosemary Wyse

The Astrophysical Journal Letters, 961:L39 (8pp), 2024 February 1 https://doi.org/10.3847/2041-8213/ad1bf0

Which Came First: Supermassive Black Holes or Galaxies? Insights from JWST

Joseph Silk[1,2,3], Mitchell C. Begelman[4,5], Colin Norman[2], Adi Nusser[6], and Rosemary F. G. Wyse[2]
[1] Institut d'Astrophysique, UMR 7095 CNRS, Sorbonne Université, 98bis Bd Arago, F-75014 Paris, France; silk@iap.fr
[2] Department of Physics and Astronomy, The Johns Hopkins University, Baltimore, MD 21218, USA
[3] Beecroft Institute of Particle Astrophysics and Cosmology, Department of Physics, University of Oxford, Oxford OX1 3RH, UK
[4] JILA, University of Colorado and National Institute of Standards and Technology, 440 UCB, Boulder, CO 80309-0440, USA
[5] Department of Astrophysical and Planetary Sciences, 391 UCB, Boulder, CO 80309-0391, USA
[6] Department of Physics and Asher Space Research Institute, Israel Institute of Technology, Technion, Haifa 32000, Israel
Received 2023 November 16; revised 2023 December 25; accepted 2024 January 2; published 2024 January 30

Abstract

Insights from JWST observations suggest that active galactic nuclei feedback evolved from a short-lived, high-redshift phase in which radiatively cooled turbulence and/or momentum-conserving outflows stimulated vigorous early star formation ("positive" feedback), to late, energy-conserving outflows that depleted halo gas reservoirs and quenched star formation. The transition between these two regimes occurred at $z \sim 6$, independently of galaxy mass, for simple assumptions about the outflows and star formation process. Observational predictions provide circumstantial evidence for the prevalence of massive black holes at the highest redshifts hitherto observed, and we discuss their origins.

Unified Astronomy Thesaurus concepts: Active galactic nuclei (16); Quasars (1319); Black holes (162)

1. Introduction

The first year of JWST science is changing our perspective on high-redshift galaxy formation. There is an intimate connection with active galactic nuclei (AGN) that is not yet effectively incorporated into current simulations. Broad-line AGN seem ubiquitous at high redshift and their host galaxies are often ultracompact and dust-reddened (Greene et al. 2023).

The spectroscopically confirmed AGN have black hole masses of 10^7–$10^9\,M_\odot$ and are inferred to be 2 orders of magnitude more abundant than the faintest UV-selected quasars, yet they represent only 1% of the UV-selected (star-forming) galaxies at $z \gtrsim 5$ (Matthee et al. 2023). Their compactness is key: we argue that cooling times in the AGN-shocked gas are significantly reduced, leading to shock-boosted star formation by compression of ambient gas clouds via mechanical feedback from fast outflows.

At present no one data set for high-redshift galaxies constrains all the important parameters that underlie the unified theory we develop below. Different samples provide complementary insights, so for present purposes, we assume broader applicability of observed behavior. This will allow us to make predictions to test whether our assumptions are valid.

We note that galaxies at high redshift with sufficient cooling to be in the positive feedback stage all have their central AGN obscured. Since they are $\mathcal{O}(0.1$–$1)$ Compton thick, we will argue that new ways of finding them directly (e.g., Square Kilometre Array (SKA), future very sensitive X-ray missions) or indirectly (e.g., very high specific star formation rates (sSFR)) are needed.

Star formation rates are enhanced and exceptionally luminous galaxies (compared to current models) are found at $z \gtrsim 10$ (Harikane et al. 2024). Yet the stellar content of massive galaxy halos seems to be depleted when the black hole masses are compared to stellar and dynamical masses, respectively (Maiolino et al. 2023). We infer that most of the stars have yet to form, but that the normal amount of gas is present in standard halos. We further infer that AGN luminosities most likely dominate in the ultraluminous galaxies, at least in those offset from the $M_{\rm BH}$–M_* scaling relation at redshifts $z \gtrsim 5$.

Chemical signatures in the highest redshift galaxies (Bunker et al. 2023) mimic those found in quasar emission line regions (Dietrich et al. 2003) as well as in some massive star clusters. These observational results, while still based on sparse data and subject to large uncertainties (Wang et al. 2023), motivate a fresh look at the coevolution of galaxies and supermassive black holes (SMBHs). They suggest an intimate connection between black hole feedback and rapid (massive) star formation, indicative of what is often referred to as positive feedback, because the chemical timescales must be relatively short. This connection, resulting in the speed-up of chemical evolution, would be further enhanced by including the effects of stellar tidal disruptions as suggested to account for the relative-to-solar nitrogen-to-carbon abundance ratio enhancement and abundance pattern consistent with a single stellar tidal disruption event in a nearby AGN (Miller et al. 2023), or by the possible role of very massive or even supermassive stars (Marques-Chaves et al. 2024; Vink 2023).

Let us first discuss how feedback might be positive or negative. The idea is that in a star-forming galaxy that hosts an AGN:

1. The accumulation of gas in the central region initially enhances both accretion onto the central SMBH and star formation.
2. Black hole outflows are invigorated, driving shocks and turbulence into nearby gas.
3. Cooling of shocked gas is effective, leading to a dense, cool phase where star formation is boosted. Feedback is positive.
4. At late times, cooling is ineffective, energy-conserving winds drive gas outflows, and feedback on star formation is negative as the gas reservoir is depleted.

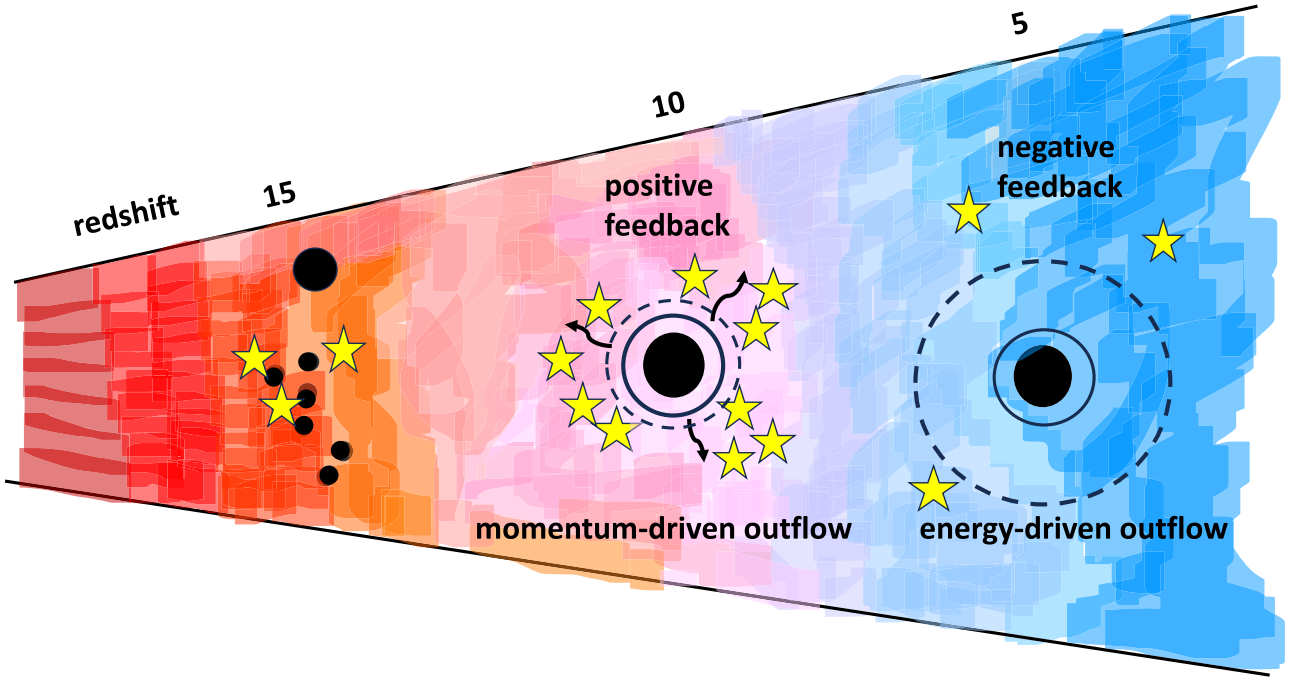

Figure 1. The transition in star formation rates and black hole growth as redshift decreases from regimes where positive feedback dominates to a later epoch when feedback is largely negative. (After Figure 2 of Costa et al. 2014.) The epoch of reionization is indicated by the change from red to blue.

We argue that the high-redshift SMBH-containing galaxy population is ultracompact (Section 2), leading to rapid cooling of outflow-shocked gas, and triggering star formation through positive feedback. This transitions to negative feedback at lower z, as cooling becomes inefficient and star formation is quenched by energetic outflows. In Section 3 we estimate the transition redshift and describe the resulting SMBH scaling laws and observational implications. The sequence of events we propose, and their approximate timing, are illustrated in Figure 1. We discuss possible early sources of massive black hole seeds in Section 4 and summarize our results in Section 5.

2. Compactness Drives Cooling and Triggering of Star Formation

The newly discovered population of ultracompact red galaxies—"little red dots"—at $10 \gtrsim z \gtrsim 5$ from JWST data exhibits effective radii of only around 150 pc yet are the natural candidates for massive galaxy precursors, with central stellar densities comparable to or exceeding those of nearby ellipticals (Baggen et al. 2023). These are lower-mass galaxies $\sim 10^9$–$10^{10} M_\odot$, whereas the resolution of stellar light in massive quasar host galaxies at $z > 6$ (Ding et al. 2023) suggests that these systems have radii of ~ 1 kpc, still significantly smaller than the several kiloparsec scales of nearby massive ellipticals.

The best size/mass measurement for a little red dot comes from Furtak et al. (2023) for the triply lensed A2744-QSO1 source, only 30 pc in size and with a quoted stellar mass of $\lesssim 2 \times 10^9 M_\odot$ and spectroscopically measured SMBH mass of $3 \times 10^7 M_\odot$ at $z \sim 7$. We caution that there are uncertainties due to contamination by AGN light, and that definitive stellar mass and central stellar density measurements await imminent spectroscopic confirmation by JWST. The tentative puzzle is to understand how such high stellar densities and large black hole masses, relative to stellar mass, are present at such high redshift. We address this in Section 4.

Many little red dots are dust-reddened AGN with black hole masses in the range 10^7–$10^8 M_\odot$ (Greene et al. 2023). We note that, if we plausibly conflate these with compact red galaxies, Baggen et al. (2023) state that $4/9$ of such objects could well host AGN, but this is indeed an open question. We predict they should contain AGN. The high number density, some 100 times higher than the faintest UV-selected quasars, and near-Eddington luminosities, suggest that they may play a key role in black hole evolution at high redshift, as we now argue. These estimated numbers of high-redshift AGN may be significantly underestimated because of dust obscuration, especially in ultracompact host galaxies (Andonie et al. 2024).

Mechanical feedback can be driven in both AGN modes: in radio mode, through powerful, fast jets (Cielo et al. 2018), and in quasar mode (Giustini & Proga 2019), through disk winds, which may be radiatively driven. Generic mechanical feedback arises from both modes and energizes the surrounding gas, which either cools or does not. When cooling is efficient, energy injection may take the form of widespread virial turbulence, balanced by local radiative losses, or—if the ambient medium is sufficiently diffuse—momentum-conserving radiative shocks. When cooling is inefficient, feedback leads to a fast, energy-conserving flow that escapes the central region of the galaxy.

The feedback can terminate black hole growth by accretion at sufficiently large black hole masses, and also regulate or quench growth of the galactic stellar mass, leading to a universal scaling relation between black hole mass and stellar velocity dispersion. Whether the black hole can be said to regulate galaxy growth, or vice versa, depends on initial conditions as well as the details of cooling efficiencies and outflow timescales. It simplifies matters to consider two limiting cases, depending on the role of cooling as the AGN outflow propagates through the surrounding interstellar medium (ISM). The first case features rapid cooling compared to the crossing time of turbulent motions in the ultracompact

protobulge, and characterizes early times; in the second case, valid at later times, cooling is ineffective in the feedback-expanded, now conventionally sized, gas-rich bulge. We assume that the ultracompactness at high redshift evolves, via merging and accretion, into relatively normal early-type galaxies at low redshift, perhaps terminating their growth by cosmic noon at $z \sim 2$.

Let us apply this comparison of dynamical and cooling timescales to the protobulge, which we assume is of order the median observed radius $r_{150} = r_{\rm eff}/150\,{\rm pc}$ for a galaxy mass normalized to $M_{10} = M/10^{10} M_{\odot}$. First, we summarize the various parameters:

1. The mean particle density is $\bar{n} \approx 3M/4\pi\mu m_p r^3 = 6 \times 10^4\, M_{10} r_{150}^{-3}\ {\rm cm}^{-3}$, corresponding to a column density $N \approx \bar{n} r = 3 \times 10^{25}\, M_{10} r_{150}^{-2}\ {\rm cm}^{-2}$.
2. The virial temperature is $T \approx GM\mu m_p/2kr = 10^7\, M_{10} r_{150}^{-1}\ {\rm K}$.
3. The dynamical timescale is $t_{\rm dyn} \approx (r^3/GM)^{1/2} = 3 \times 10^5\, r_{150}^{3/2} M_{10}^{-1/2}\ {\rm yr}$.
4. The bremsstrahlung cooling rate is $\lambda_{\rm ff} n^2 T^{1/2}\ {\rm erg\ cm^{-3} s^{-1}}$, where $\lambda_{\rm ff} = 1.4 \times 10^{-27}$ in centimeter–gram–second units and the corresponding cooling timescale is $t_{\rm cool} = 5kT^{1/2}/2\mu\lambda_{\rm ff} n \approx 300\, r_{150}^{5/2} M_{10}^{-1/2}\ {\rm yr}$.
5. The postshock column density of cooled gas behind shocks at close to the virial speed, $v_s \approx 600\, (M_{10}/r_{150})^{1/2}\ {\rm km\ s^{-1}}$, is at least $N_{\rm cool} = n t_{\rm cool} v_s \approx 2 \times 10^{22}\, M_{10} r_{150}^{-1}\ {\rm cm}^{-2}$.

Traversing this critical column density is key to our argument below on the transition between the momentum and energy-conserving regimes. Note that in the likely presence of clouds, the cooling timescale is reduced by the density contrast, which could further reduce the local cooling timescale by 1 order of magnitude or more. In general, we are in a cooling-dominated regime, especially for the ultracompact galaxies, arguing strongly for radiatively damped turbulence or momentum-driven flows in the gas-rich systems we are considering at high redshift.

3. Transitioning from High to Low z

Formation of ultracompact galaxies at high z necessarily occurs in a dense protogalactic environment, where the high density favors massive star formation. Compaction offers another pathway to forming so-called blue nuggets (Lapiner et al. 2023), but will inevitably be too rare and fine-tuned a process for the very large population of "little red dots," the ultracompact galaxies at $z \gtrsim 5$. So-called red nugget galaxies observed at low redshift are also compact and some have obese black holes (Cohn et al. 2023), that is, black holes that are overly massive for the $M_{\rm BH}$–σ_* scaling relation. Indeed, similarly obese black holes, with a much larger scatter relative to their host galaxies than in the nearby universe, have been found at $z = 4$–6 (Stone et al. 2023). We argue below that this evidence points to a scenario in which the formation of SMBHs, and indeed AGN, may even precede the formation of the dominant stellar components of galaxies.

In theoretical support of this picture, we note that at high z, bremsstrahlung cooling is very efficient within the central hundreds of parsecs in ultracompact galaxies. At low z, especially at reduced gas fraction in gas-rich AGN hosts of "typical" size, bremsstrahlung cooling is inefficient. This provides the motivation for our bimodal scenario of radiative turbulence and momentum-conserving outflows at high z and energy-conserving outflows at low z.

We acknowledge, however, that observational support remains ambiguous, due in part to degeneracies among the diagnostics used to measure high-redshift stellar masses, AGN luminosity, initial stellar mass function (IMF), burstiness, and reddening. Selection biases need to be more thoroughly understood before our inferences can be regarded as robust (Li et al. 2022).

3.1. Gas-rich Initial Conditions

Our model simplifies the observational situation by making the premise that the little red dots are 10 times smaller and 10 times more gas-rich, but with ~10% of the mass of their low-z counterparts that host SMBH. The body of current data generally supports this view but is far from conclusive. Most uncertain is our claim that such objects are, at least initially, gas-rich. We infer this from the measured extreme stellar densities. However, whether gas is in a compact bulge-like distribution or in a more extensive disk is unclear. The observations nevertheless reinforce our confidence in a significant cooling transition, from early strong cooling to late-epoch cooling suppression. Although complications of enrichment and molecular gas cooling ultimately need to be taken into account, it is unlikely that our general conclusions will change. In summary, for the observed and inferred column densities of $z \gtrsim 5$ galaxies, cooling occurs rapidly and over very short length scales for the dominant gas-rich and AGN-containing galaxy population.

Cooling of the outflows is generally inefficient at lower redshifts, where one might expect to observe generic energy-conserving flows (Wagner et al. 2012; Costa et al. 2014; King & Pounds 2015). These can reach large radii (Rupke et al. 2017), depleting the gas reservoir and quenching the star formation rate. Sporadic momentum-conserving episodes could still occur, triggering rarer cases of star formation in dense gas clumps.

Further evidence for the preponderance of gas rather than stars in massive galactic halos at high z can be inferred from the SMBH scaling relations. Under cooling-dominated conditions, SMBH scaling relations are embedded early and close-in. Maiolino et al. (2023) find a ratio of SMBH mass to stellar mass, $M_{\rm BH}/M_*$, enhanced by a large factor $\gtrsim$100, whereas the ratio of SMBH to dynamical mass, reflected in the $M_{\rm BH}$–σ_* correlation, follows the standard low-redshift behavior. This is consistent with SMBH growth at close to the Eddington rate, tightly coupled to feedback stirring up the circumnuclear gas and promoting nuclear star formation. Here, star formation is triggered by the compression of dense nuclear clouds, a phenomenon we may refer to as positive feedback. In the low-redshift universe, cases of positive feedback are rare. Other samples also show undermassive SMBH at high z in the mass range $M_{\rm BH} = 10^6$–$10^8 M_{\odot}$ relative to the local $M_{\rm BH}$–σ_* correlation (see Harikane et al. 2023), but nevertheless in general agreement with the preceding result when the large gas mass is taken into account. Note, however, that other efforts to determine similar scalings at high redshift do differ, with indications of overmassive SMBH at high z from dynamical constraints, see Übler et al. (2023) and from a recent Chandra X-ray detection at $z = 10.1$ (Natarajan et al. 2024). On balance, we cautiously argue from the evidence that the most massive halos at high redshift are plausibly star-poor, guaranteeing a gas reservoir for triggering star formation at early epochs.

3.2. Positive Feedback at High Redshift

Observational signatures of positive feedback include spatially extended star formation near radio jets (Duggal et al. 2023); star formation triggered at the edges of radio-jet-induced bubbles (Venturi et al. 2023); interactions of low-powered trapped jets with the ISM (Girdhar et al. 2022); enhanced nuclear star formation in quasar hosts (Molina et al. 2023); nuclear rings of enhanced star formation (Pak et al. 2023; Zhang & Ho 2023) created by the outflowing nuclear jet driving through and compressing the interstellar gas (Dugan et al. 2017); and enhanced star formation rates in galactic outflows (Gallagher et al. 2019). More generally, positive feedback can be monitored by emission line signatures in star-forming regions, enhanced nuclear star formation accompanying accelerated SMBH growth, as well as associated supernovae and chemical enrichment.

There are indications that the number of luminous galaxies at $z \gtrsim 10$ exceeds those of standard simulation predictions (Harikane et al. 2024). We note that the lensed sample of UV luminous galaxies at $z \gtrsim 9$ studied by Chemerynska et al. (2023) included a high fraction of AGN (Fujimoto et al. 2023).

We speculate that positive feedback is a plausible pathway to account for central bursts of star formation that generate the early high luminosities, especially with inevitably bursty star formation (Sun et al. 2023). Rival suggestions seem more contrived, including feedback-free star formation, which requires several metal-free cycles (Dekel et al. 2023), or the fine-tuning of a top-heavy IMF (Finkelstein et al. 2023; Yung et al. 2024). While such other interpretations abound, we argue that the observed coevolution of AGN and star formation merits a more fundamental explanation.

3.3. Transition

The transition redshift from radiative to energy-conserving feedback is hard to estimate a priori, and likely depends on a number of uncertain factors. As discussed in Section 3.4, in early radiative stages the feedback energy may be mixed thoroughly with the ISM, leading to virial turbulence that pervades the galaxy. As the transition approaches, this mixing becomes less complete and the situation begins to resemble the large-scale, cooled (i.e., momentum-conserving) shocks considered by Fabian (1999). Substantial mass loading in the outflow-driven shock, whether disk-wind-driven (Raouf et al. 2023) or radio-jet-driven (Venturi et al. 2023), sets favorable conditions for postshock cooling. The cooling column can be accumulated by some combination of shock compression and turbulent mixing (Begelman & Fabian 1990) facilitated by Rayleigh–Taylor and Kelvin–Helmholtz instabilities. Because of large density contrasts between the AGN wind or jet and the dense ISM, the speeds governing the shock or mixing process are likely to be smaller than observed outflow speeds (typically a few thousand km s^{-1}) but may exceed the virial speeds used to estimate the cooling column in Section 2. These considerations suggest that a typical cooled layer can have a column density several times larger than the estimate of Section 2, e.g., $N_{\rm cool} \sim 10^{23}$ $\rm cm^{-2}$.

The estimate above is for a single cooling layer. However, the total cooled column may be substantially larger than this due to the turbulent nature of the mixing process, which can create a two-phase medium characterized by a multiplicity of dense cooled layers interspersed with hot gas throughout the galaxy. If there are $\mathcal{N}$ such layers along a typical radius, then $N_{\rm cool}$ will increase by this factor.

The transition to energy-conserving feedback occurs when the actual column density in the galaxy drops below $N_{\rm cool}$. Empirical data on ISM column densities in massive and AGN host galaxies suggest that the gas column evolves with redshift approximately as $(1+z)^{3.3}$ (Gilli et al. 2022), and is some 100 times larger than local values at $z = 4$–6. The estimated normalization is $N_{\rm H} = 10^{21}(1+z)^{3.3}\rm cm^{-2}$. Then, if $\mathcal{N} \sim O(10)$, the total cooled column is $\sim 10^{24}$ $\rm cm^{-2}$. With this order-of-magnitude estimate, we infer a transition redshift of $\sim$6, consistent with observational constraints.

This estimated transition epoch separates momentum-conserving flows at high z and energy-conserving flows at low z. Beyond $z \sim 6$, Compton thick values may be reached as in Parlanti et al. (2023), explaining the likely absence of any X-ray detections of these AGN at such high redshift. Radio jet detections are predicted, but would be challenging, as discussed below.

3.4. Global Considerations

The evolutionary scenario discussed above can be understood in terms of the evolving thermal structure of the galactic gas. At early times, the cooling timescale of gas at the mean density is so short that the medium likely assumes a multiphase structure, dominated by a "hot" phase at roughly the virial temperature, filling most of the volume, and a cold, cloudy phase containing most of the mass. Under these highly inhomogeneous conditions, we argue that most of the feedback energy is trapped and does not escape as a large-scale outflow, but rather drives turbulence and localized shocks throughout the galaxy. Energy is lost both through the cooling of the hot phase and strong radiative shocks due to collisions between cold clouds.

A key controlling parameter is the "covering factor" C of the clouds, which could be large. Physically, C can be interpreted as the number of shocks encountered by a given parcel of cold gas per dynamical time. The injection of energy due to feedback regulates C by shredding the clouds until C is large enough to dissipate the feedback energy. A toy model illustrating these ideas is presented in the Appendix. When $C \gg 1$, the feedback energy is radiated away and this corresponds to the radiative regime. As the galaxy grows and gas is used up to form stars, C gradually decreases. When $C \gtrsim 1$, the galaxy approaches the transition discussed in Section 3.3, where cooling is still important but the feedback drives a momentum-conserving outflow. Finally, when $C \lesssim 1$, the feedback energy can sweep out the hot phase and entrain the cold gas in a wind, corresponding to the energy-conserving phase.

Generally, we would associate the radiative and momentum-conserving phases with enhanced star formation and the energy-conserving phase with negative feedback. But the multiphase character of the medium can also lead to negative feedback early in the radiative phase. When C is very large, the clouds can be shredded to the point where the size of a typical cloud is smaller than its Jeans length, so that star formation is suppressed (negative feedback). This three-stage evolution fits naturally with models in which an overmassive black hole forms first and a galaxy builds around it, with σ_* increasing with time. At first, star formation is inefficient (negative feedback due to clouds smaller than the Jeans length) but then

switches to rapid star formation as C decreases (positive feedback), depleting the gas fraction and leading to a second phase of negative feedback: when C becomes smaller than some threshold value, an energy-conserving wind sweeps through the galaxy. If the galaxy assembly timescale is longer than the growth time of the black hole (a few tens of Myr for Eddington-limited accretion), then the SMBH regulates the size of the galaxy during early radiative stages: for a given SMBH mass the velocity dispersion increases until it hits the M–σ_* line, $M_{\rm BH} \sim \sigma_*^5$.

3.5. Negative Feedback at Low Redshift

Once the AGN triggering of star formation fades, disk formation and instabilities are expected to limit star formation and to slow down any final collapse. Large-scale gravito-turbulence allows convergence to a global Schmidt–Kennicutt (SK) law (Nusser & Silk 2022), regulates the star formation efficiency, and controls star formation rates in the nearby universe.

The next phase is an energy-conserving outflow. Energy-driven feedback takes over at late times as the gas reservoir expands due to the feedback. Star formation efficiencies, as deduced from the correlation of integrated star formation with molecular gas masses from NOEMA observations, favor gas depletion times of 0.1–1 Gyr up to $z \sim 6$, typically preferring mild starbursts (Berta et al. 2023). JWST data confirm the presence of AGN-driven neutral outflows in massive galaxies at $z \sim 2$ (Davies et al. 2023), where star formation rates nevertheless follow an SK law (Schulze et al. 2019).

Observations of outflows driven by AGN accretion disk winds at $\sim 0.1c$ tend to support energy conservation (Tombesi et al. 2015) from subparsec to kiloparsec scales and beyond, where molecular outflows are observed. The more recent data on outflows has a large dispersion, but generally supports black hole and galaxy coevolution (Capelo et al. 2023).

3.6. Scaling Laws

In the momentum-conserving limit, the central black hole grows by accretion until inflow is halted and overpowered by dynamical feedback. The black hole mass growth saturates at $M_{\rm BH} = \dfrac{f_{\rm gas}\,\sigma_T}{\pi G^2 m_p}\sigma_*^4$, with reasonable assumptions about the wind speed and efficiency (Fabian 1999). This expression approximately fits the observed $M_{\rm BH} - \sigma_*$ correlation in slope and normalization, according to theory (Costa et al. 2014) and observations for elliptical galaxies (and classical bulges; Kormendy 2020) and AGN with $M_{\rm BH} = 10^7$–$10^9 M_\odot$ (Bennert et al. 2021).

In the energy-conserving limit, the feedback generates a late-time scaling law $M_{\rm BH} = \dfrac{11 f_{\rm gas}\,\sigma_T}{\epsilon \pi G^2 m_p c}\sigma_*^5$ (Silk & Rees 1998; Costa et al. 2014), where ϵ is the radiative efficiency of the outflow. Late contributions to SMBH growth in this regime suggest that the final scaling law should be intermediate between $M_{\rm BH} \propto \sigma_*^4$ and $M_{\rm BH} \propto \sigma_*^5$. Local observations indeed prefer $\alpha \approx 4.4$ (Kormendy & Ho 2013). Our model actually predicts a steepening of the scaling law with redshift from $z \gtrsim 5$ to $z \lesssim 5$. The steeper slope and lower normalization for the energy-driven limit, dictated by the factor $\sim \sigma_*/c$, suggest that SMBHs forming at low z will be slightly undermassive (Baldassare et al. 2020). Outflows control late epoch quenching of star formation. They might also be relevant for dwarf galaxies, where outflows are expected to be even more dominant at early gas-rich epochs because of the shallow potential wells. This contrasts with the obese (overmassive) mode of the $M_{\rm BH}$–σ_* correlation, inferred dynamically at high z, that we argue coexists with stimulated star formation and stellar mass growth.

3.7. Indications from Simulations

Early studies showed that AGN outflows can produce positive feedback on star formation by overpressurizing circumgalactic gas (Bieri et al. 2016). Radio jets are recognized to suppress star formation by stirring up ISM turbulence as well as producing both positive and negative feedback on star formation by driving the compression of ambient clouds (Mandal et al. 2021). At later epochs ($z \lesssim 5$), the consensus view is that global negative feedback drives massive outflows, exhausts the gas supply, quenches star formation, and terminates galaxy growth.

Positive feedback involving strong compression of off-center gas clouds by quasar winds is indeed found in the most recent simulations (Mercedes-Feliz et al. 2023), echoing the preceding analytic discussion and early 1D simulations (Silk & Nusser 2010).

3.8. Observational Probes

Early chemical enrichment provides a new dimension to our model. A remarkable [N/O] $\gtrsim +0.6$ enrichment (Cameron et al. 2023) is seen in GN-z11 (Bunker et al. 2023), with one of the highest confirmed spectroscopic redshifts for any galaxy. One possible source could be very massive stars with initial masses of hundreds of solar masses (Vink 2023), associated with a top-heavy IMF (Bekki & Tsujimoto 2023). An alternative could be supermassive stars (Marques-Chaves et al. 2024), with initial masses of thousands of solar masses, or direct-collapse black hole precursors with masses possibly reaching 10^5–$10^6 M_\odot$. Either option requires rapid early star formation in massive central star clusters (Belokurov & Kravtsov 2023), possibly in intermittent episodes (Kobayashi & Ferrara 2023). Similar episodes are predicted in our boosted nuclear star formation model.

Such N excesses are common to quasar emission line regions (Dietrich et al. 2003), further justifying the AGN–star formation connection. There is one more consequence of the positive feedback episodes driven by radiative feedback. In addition to triggering star formation, these can also lead to the compactification of gas clouds whose mergers provide sources of clumpy accretion in the nuclear environment. These may boost black hole growth via early super-Eddington accretion in massive halos (Bennett et al. 2024).

4. Massive Black Hole Seeds

The large population of ultracompact, dust-reddened red galaxies sets the scene for the preceding discussion of coevolution of star formation and AGN outflows. We have argued that this implies the existence of a large population of massive black holes at very early times. How did they form? We first focus on Population III. Massive seeds are one possible precursor to the observed high-z massive black hole population. Suppression of H_2 by UV in the Lyman–Werner bands is one ingredient that can promote the formation of

massive seeds. The near-dominance (and possible predominance) of AGN at the earliest epochs suggests that massive black hole seeds could coevolve with Population III if the latter provide a sufficient local UV background at $z = 10$–15. A plausible environment would be that of early-forming protoclusters that contain AGN forming coevally with the first galaxies.

Direct collapse black hole seeds can also be abetted at early times by magnetic pressure in the central gas disks that coexist with the first AGN. This could result from early dynamo activity, with analytic studies favoring magnetic levitation suppressing disk fragmentation to stars (Begelman & Silk 2017) and catalyzing black hole growth by reducing the density and speeding up the central inflow (Begelman & Silk 2023).

Magnetic boosting and/or turbulence driven by low-angular-momentum cold flows provide an alternative channel for suppressing premature fragmentation into stars. Recent MHD simulations suggest that these may generate strong toroidal fields that dominate the inner accretion disk and thereby suppress disk fragmentation (Hopkins et al. 2023).

The early formation of a large population of (ultracompact) galaxies is problematic in the standard ΛCDM model. A possible resolution could involve excess small-scale power in the primordial density fluctuation power spectrum on dwarf galaxy scales, as has been explored in earlier studies of ways of boosting early formation of galaxies (Chevallard et al. 2015) and AGN (Habouzit et al. 2016) by primordial galaxy-scale non-Gaussianities (Sabti et al. 2023) or by primordial spectral bumps (Tkachev et al. 2024). It is well known that the formation of SMBH, especially at high z, cannot be explained if accretion growth from stellar-mass black holes is Eddington accretion-limited.

This suggests that any solution should simultaneously resolve the issues of AGN host compactness at high z and the early formation of SMBHs. We consider several possible pathways:

1. The seed black holes may be generated at sufficiently early epochs by supermassive stars (Begelman et al. 2006; Begelman 2010) forming from excess small-scale power in the distribution of primordial density fluctuations (Tkachev et al. 2023), or by scale-dependent non-Gaussianity (Habouzit et al. 2016). The supermassive star ansatz might provide one way to account for the N excess in GN-z11 (Nagele & Umeda 2023), but the production of a large number of such seeds would be constrained by early heavy element production.
2. Alternatively, there could be a seeding population of primordial black holes that need only contribute $\sim 10^{-4}$ of the dark matter, specified by the seed abundance requirement and the observationally constrained 10^3–$10^6 M_\odot$ mass range, and restricted by microlensing, dynamical friction, and accretion limits (Yuan et al. 2023). The transformation from primordial black holes to ultracompact galaxies is accounted for by gas accretion with predicted growth since the last scattering by a factor ~ 100 (Carr & Silk 2018). The ultracompact galaxies are expected to contain obese central back holes, relative to the usual dynamical scaling relation, with the ratio of black hole to galaxy mass decreasing toward lower redshift.
3. We might also consider the super-Eddington growth of moderate-mass seeds inside quasi-stars (Begelman et al. 2008). These would resemble red giants (or Thorne–Żytkow objects) but powered by black hole accretion with gaseous envelopes that are orders of magnitude more massive than the central black hole. Accretion at the Eddington limit for the envelope mass can rapidly produce $\sim 10^5$–10^6 black hole seeds.

The primordial black hole scenario—which has support in diverse inflationary scenarios—is more direct, given the initial hypothesis that SMBHs ($\gtrsim 10^6 M_\odot$) form before significant star formation. The required mass range of primordial intermediate-mass black holes (IMBHs; 10^3–$10^6 M_\odot$) is allowed at a contribution of up to $\sim 10^{-4}$ of the dark matter density. The observed (comoving) number density of high-z AGN is $\sim 10^{-5}\,\mathrm{Mpc}^{-3}$, corresponding to a mass density $\sim 10(M/10^6 M_\odot) M_\odot\,\mathrm{Mpc}^{-3}$, whereas the dark matter density is $\sim 10^4\,M_\odot\,\mathrm{Mpc}^{-3}$.

Early fragmentation might even be a positive factor in the context of boosted small-scale primordial density fluctuations. For example, early-forming exceptionally dense star clusters could be factories for runaway astrophysical black hole mergers and early formation of IMBH seeds that would be enhanced in mass by gas accretion and dwarf galaxy mergers. They would produce distinct gravitational wave signals for future-generation experiments (Kritos et al. 2023). Nearby dwarf galaxies often have nuclear star clusters (Carlsten et al. 2022) and IMBHs (Reines et al. 2020).

5. Discussion

We find three phases of the early coevolution of SMBHs and their galaxy hosts:

a) $z \gtrsim 15$: black hole growth and the rise of AGN.

b) $5 \lesssim z \lesssim 15$: star formation bursts triggered by radiative turbulence and momentum-conserving AGN outflows into a dense clumpy ISM.

c) $z \lesssim 5$: star formation quenched and gas depleted by energy-conserving AGN outflows.

The case for SMBHs forming simultaneously with galaxies seems compelling. We focus here on the interplay between AGN activity and star formation and argue for a bimodal synergy. AGN feedback may have evolved from a short-lived but vigorous phase of positive feedback via radiative turbulence and momentum-driven outflows in ultracompact galaxy hosts. Radiative shocks, both from cloud collisions and outflows, triggered positive feedback and drove the first episodes of vigorous star formation at $z \gtrsim 10$. Energy-driven outflows in initially gas-rich galaxies then depleted halo gas reservoirs and stimulated negative feedback, transitioning at $z \lesssim 6$.

The coevolution of AGN outflows and star formation may be the missing link to the high-redshift Universe. Nearby examples abound of both negative and positive feedback associated with AGN outflows. Positive feedback may be a missing ingredient for the early boosting of star formation that is intimately linked to the earliest AGN activity. The rapid evolution of stellar mass, the compactness of the galaxy population, and the growth of SMBHs at early epochs are outstanding issues that remain to be adequately addressed in cosmological simulations (Byrne et al. 2023).

5.1. Obscuration

For massive galaxies, we have found that at high redshifts, cooling of QSO winds ($\sim$3000 km s^{-1}) occurs above $z \sim 6$ for host-galaxy gas column densities $\lesssim 10^{23}$ cm^{-2}. This range characterizes the transition to momentum-driven feedback as z increases. Such molecular column densities are inferred from theory and observations (Gilli et al. 2022) and indicate that it is not the "classical" torus obscuring the AGN, but the galaxy's ISM itself. The minimum obscuration along the line of sight for transitioning from momentum-driven to energy-driven outflows is

$$N_{\rm cool} \approx 10^{23}\ {\rm cm}^{-2}\left(\frac{v_s}{3000\ {\rm km\ s}^{-2}}\right)^2,$$

where v_s is the feedback outflow velocity. Even Compton-thick obscuring host galaxies with no X-ray detections may well exceed this cooling column by more than 1 order of magnitude. All of these high-redshift objects should have cooled, momentum-driven winds and outflows. For the relevant range of cooling columns $N_{\rm cool} \approx 10^{23}$–$10^{24}$ cm^{-2}, with Compton depths 0.1–1.0, the AGN will be obscured. While this simple estimate assumes a spherical outflow, geometrical arguments suggest that in the more general case of directed outflows, high-z AGN are always buried or obscured along the cooling momentum-driven outflow. In this case, viewed from the side, one would not see the full outflow velocity. Observable effects might, however, include turbulence, narrow line emission, and dust.

5.2. Observational Predictions

Simple radiative or momentum-driven feedback for massive high-redshift galaxies has such a high requirement on the column density that the central AGN driving the feedback is hidden and will not generate detectable emission lines or X-rays. It is hidden by the host galaxy ISM, and not by a Type I/Type II AGN torus. Our best bet to detect hidden momentum-driven AGN feedback may be via searching for very high specific star formation rates due to the cooling shock overpressure. In principle, it should be much easier to find a positive feedback signature than has hitherto been the case in attempts to find direct evidence of the negative signature associated with the quenching of star formation. The former effect is of order $10(v_s/3000\ {\rm km\ s}^{-1})/(v_{\rm cir}/300\ {\rm km\ s}^{-1})$ faster, where $v_{\rm cir}$ is the galaxy potential-defined circular velocity of the massive host galaxy.

Other observational predictions include redshift dependence in the SMBH mass-scaling relations with stellar and halo masses, AGN-correlated star formation rates, and chemical abundance signatures. Improved statistics, imaging, and spectroscopic follow-up are needed to justify the systematic global interpretation that is being proposed here with admittedly simplistic but plausible arguments. Finally, we note that specific star formation rates are not yet available for the JWST galaxies in the $z = 5$–15 range. These would be an important indicator of positive feedback, especially if correlated with the presence of an AGN or various multiwavelength luminosities, including in the radio and/or X-ray bands, as well as optical/IR emission lines and infrared dust continuum. The star formation and AGN duty cycles may be out of phase and dust-obscured, so detection of any such correlations would be challenging. With current X-ray missions, there is little hope of detection unless there is an object in the center of an existing deep field such as the CDFS 7 Ms deep field. JWST might help with the detection of faint broad wings of IR lines but the Compton-thick obscuration is a formidable obstacle.

One option currently being explored is the use of SKA radio observations to detect these deeply buried high-redshift QSOs (G. Mazzolari, private communication), but the expected detection rates are small. Finally, we note that at lower redshifts than $\sim$6, there is no relation in our model between AGN obscuration due to the "torus" and our cooling-driven wind/outflow feedback mechanism. The latter is dependent on the host galaxy ISM. Only at the higher redshifts considered here can we relate obscuration and ISM shock cooling. This regime may eventually be revealed by deep Atacama Large Millimeter/submillimeter Array observations.

Acknowledgments

We thank Roberto Gilli, Luis Ho, Erini Lambrides, John Silverman, Alex Wagner, and Feng Yuan for inspiring comments and the referee for an encouraging report. A.N. is supported by the Israel Science Foundation (ISF) grant 893/22 and the Asher Space Research Institute. R.F.G.W. acknowledges support through the generosity of Eric and Wendy Schmidt, by recommendation of the Schmidt Futures program.

Appendix
Toy Model for Radiative Feedback

In the radiative feedback limit, we assume a two-phase medium with cold clouds at $10^3 T_3$ K and a diffuse hot phase at $T_{\rm vir} = 10^7 M_{10} r_{150}^{-1}$ K, where we use the same notation as in Section 2. For simplicity, we assume that both phases absorb similar amounts of feedback power, roughly 10% of the Eddington luminosity for a $10^8 m_8 M_\odot$ SMBH. If the hot phase radiates away this energy via bremsstrahlung, its density is given by $n_h \approx 600 m_8^{1/2} M_{10}^{-1/4} r_{150}^{-5/4}$ cm^{-3}. For characteristic parameters, we see that this is much smaller than the mean density $\bar{n}$ from Section 2. In pressure equilibrium, the cold phase density is $n_c \approx 6 \times 10^6 m_8^{1/2} M_{10}^{3/4} r_{150}^{-9/4} T_3^{-1}$ cm^{-3} and the volume filling factor is small, $f_c = f_{\rm gas}\bar{n}/n_c \approx 0.01 f_{\rm gas} m_8^{-1/2} M_{10}^{1/4} r_{150}^{-3/4} T_3$.

We now introduce the cloud covering factor, C, discussed in Section 3.4, which is, in effect, an optical depth for cloud–cloud collisions. If we regard the clouds as spheres of radius r_c, we have $r_c \approx C^{-1} f_c\, r \approx C^{-1} f_{\rm gas} m_8^{-1/2} M_{10}^{1/4} r_{150}^{1/4} T_3$ pc. The Jeans length for these clouds is given by $r_J \approx 0.1 m_8^{-1/4} M_{10}^{-3/8} r_{150}^{9/8} T_3$ pc. We thus find that $r_c < r_J$ when $C > 10 f_{\rm gas} m_8^{3/4} M_{10}^{5/8} r_{150}^{-7/8}$, implying that the clouds are too small to gravitationally fragment and form stars when C is sufficiently large.

We propose that the clouds are stirred up to virial random velocities by the feedback, then radiate away this energy at a rate $\sim C f_{\rm gas} \sigma_*^5/G$ via cloud–cloud collisions. Equating this dissipation rate to the feedback power supply determines C. Adopting our earlier assumption about the magnitude of the feedback power supply, we can also express this relation as $m_8 \sim 0.4 C f_{\rm gas} (\sigma_*/200\ {\rm km\ s}^{-1})^5$, which is very close to the modern-day M–σ_* relation when $C \sim$ a few. This relation also suggests that star formation can be suppressed by high values of C at early times if SMBH seeds are initially overmassive with respect to the σ_* of the protogalaxies forming around them.

ORCID iDs

Joseph Silk https://orcid.org/0000-0002-1566-8148
Mitchell C. Begelman https://orcid.org/0000-0003-0936-8488
Colin Norman https://orcid.org/0000-0002-5222-5717
Adi Nusser https://orcid.org/0000-0002-8272-4779
Rosemary F. G. Wyse https://orcid.org/0000-0002-4013-1799

References

Andonie, C., Alexander, D. M., Greenwell, C., et al. 2024, MNRAS: Letters, 527, L144
Baggen, J. F. W., van Dokkum, P., Labbé, I., et al. 2023, ApJL, 955, L12
Baldassare, V. F., Dickey, C., Geha, M., & Reines, A. E. 2020, ApJL, 898, L3
Begelman, M. C. 2010, MNRAS, 402, 673
Begelman, M. C., & Fabian, A. C. 1990, MNRAS, 244, 26P
Begelman, M. C., Rossi, E. M., & Armitage, P. J. 2008, MNRAS, 387, 1649
Begelman, M. C., & Silk, J. 2017, MNRAS, 464, 2311
Begelman, M. C., & Silk, J. 2023, MNRAS, 526, L94
Begelman, M. C., Volonteri, M., & Rees, M. J. 2006, MNRAS, 370, 289
Bekki, K., & Tsujimoto, T. 2023, MNRAS, 526, L26
Belokurov, V., & Kravtsov, A. 2023, MNRAS, 525, 4456
Bennert, V. N., Treu, T., Ding, X., et al. 2021, ApJ, 921, 36
Bennett, J. S., Sijacki, D., Costa, T., Laporte, N., & Witten, C. 2024, MNRAS, 527, 1033
Berta, S., Stanley, F., Ismail, D., et al. 2023, A&A, 678, A28
Bieri, R., Dubois, Y., Silk, J., Mamon, G. A., & Gaibler, V. 2016, MNRAS, 455, 4166
Bunker, A. J., Saxena, A., Cameron, A. J., et al. 2023, A&A, 677, A88
Byrne, L., Faucher-Giguère, C.-A., Wellons, S., et al. 2023, arXiv:2310.16086
Cameron, A. J., Katz, H., Rey, M. P., & Saxena, A. 2023, MNRAS, 523, 3516
Capelo, P. R., Feruglio, C., Hickox, R. C., & Tombesi, F. 2023, in Handbook of X-ray and Gamma-ray Astrophysics, ed. C. Bambi & A. Santangelo (Berlin: Springer), 126
Carlsten, S. G., Greene, J. E., Beaton, R. L., & Greco, J. P. 2022, ApJ, 927, 44
Carr, B., & Silk, J. 2018, MNRAS, 478, 3756
Chemerynska, I., Atek, H., Furtak, L. J., et al. 2023, arXiv:2312.05030
Chevallard, J., Silk, J., Nishimichi, T., et al. 2015, MNRAS, 446, 3235
Cielo, S., Bieri, R., Volonteri, M., Wagner, A. Y., & Dubois, Y. 2018, MNRAS, 477, 1336
Cohn, J. H., Curliss, M., Walsh, J. L., et al. 2023, ApJ, 958, 186
Costa, T., Sijacki, D., & Haehnelt, M. G. 2014, MNRAS, 444, 2355
Davies, R. L., Belli, S., Park, M., et al. 2023, arXiv:2310.17939
Dekel, A., Sarkar, K. C., Birnboim, Y., Mandelker, N., & Li, Z. 2023, MNRAS, 523, 3201
Dietrich, M., Hamann, F., Shields, J. C., et al. 2003, ApJ, 589, 722
Ding, X., Onoue, M., Silverman, J. D., et al. 2023, Natur, 621, 51
Dugan, Z., Gaibler, V., & Silk, J. 2017, ApJ, 844, 37
Duggal, C., O'Dea, C. P., Baum, S. A., et al. 2023, arXiv:2309.00110
Fabian, A. C. 1999, MNRAS, 308, L39
Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., et al. 2023, ApJL, 946, L13
Fujimoto, S., Wang, B., Weaver, J., et al. 2023, arXiv:2308.11609
Furtak, L. J., Labbé, I., Zitrin, A., et al. 2023, arXiv:2308.05735
Gallagher, R., Maiolino, R., Belfiore, F., et al. 2019, MNRAS, 485, 3409
Gilli, R., Norman, C., Calura, F., et al. 2022, A&A, 666, A17
Girdhar, A., Harrison, C. M., Mainieri, V., et al. 2022, MNRAS, 512, 1608
Giustini, M., & Proga, D. 2019, A&A, 630, A94
Greene, J. E., Labbe, I., Goulding, A. D., et al. 2023, arXiv:2309.05714
Habouzit, M., Volonteri, M., Latif, M., et al. 2016, MNRAS, 456, 1901
Harikane, Y., Nakajima, K., Ouchi, M., et al. 2024, ApJ, 960, 56
Harikane, Y., Zhang, Y., Nakajima, K., et al. 2023, ApJ, 959, 39
Hopkins, P. F., Squire, J., Su, K.-Y., et al. 2023, arXiv:2310.04506
King, A., & Pounds, K. 2015, ARAA, 53, 115
Kobayashi, C., & Ferrara, A. 2023, arXiv:2308.15583
Kormendy, J. 2020, in IAU Symp. 53, Galactic Dynamics in the Era of Large Surveys, ed. M. Valluri & J. A. Sellwood (Cambridge: Cambridge Univ. Press), 186
Kormendy, J., & Ho, L. C. 2013, ARA&A, 51, 511
Kritos, K., Berti, E., & Silk, J. 2023, PhRvD, 108, 083012
Lapiner, S., Dekel, A., Freundlich, J., et al. 2023, MNRAS, 522, 4515
Li, J., Silverman, J. D., Izumi, T., et al. 2022, ApJL, 931, L11
Maiolino, R., Scholtz, J., Curtis-Lake, E., et al. 2023, arXiv:2308.01230
Mandal, A., Mukherjee, D., Federrath, C., et al. 2021, MNRAS, 508, 4738
Marques-Chaves, R., Schaerer, D., Kuruvanthodi, A., et al. 2024, A&A, 681, A30
Matthee, J., Naidu, R. P., Brammer, G., et al. 2023, arXiv:2306.05448
Mercedes-Feliz, J., Anglés-Alcázar, D., Hayward, C. C., et al. 2023, MNRAS, 524, 3446
Miller, J. M., Mockler, B., Ramirez-Ruiz, E., et al. 2023, ApJL, 953, L23
Molina, J., Ho, L. C., Wang, R., et al. 2023, ApJ, 944, 30
Nagele, C., & Umeda, H. 2023, ApJL, 949, L16
Natarajan, P., Pacucci, F., Ricarte, A., et al. 2024, ApJL, 960, L1
Nusser, A., & Silk, J. 2022, MNRAS, 509, 2979
Pak, M., Lee, J. H., Jeong, H., & Jeong, W.-S. 2023, AJ, 165, 109
Parlanti, E., Carniani, S., Übler, H., et al. 2023, arXiv:2309.05713
Raouf, M., Viti, S., García-Burillo, S., et al. 2023, MNRAS, 524, 786
Reines, A. E., Condon, J. J., Darling, J., & Greene, J. E. 2020, ApJ, 888, 36
Rupke, D. S. N., Gültekin, K., & Veilleux, S. 2017, ApJ, 850, 40
Sabti, N., Muñoz, J. B., & Kamionkowski, M. 2023, arXiv:2305.07049
Schulze, A., Silverman, J. D., Daddi, E., et al. 2019, MNRAS, 488, 1180
Silk, J., & Nusser, A. 2010, ApJ, 725, 556
Silk, J., & Rees, M. J. 1998, A&A, 331, L1
Stone, M. A., Lyu, J., Rieke, G. H., Alberts, S., & Hainline, K. N. 2023, arXiv:2310.18395
Sun, G., Faucher-Giguère, C.-A., Hayward, C. C., et al. 2023, ApJL, 955, L35
Tkachev, M. V., Pilipenko, S. V., Mikheeva, E. V., & Lukash, V. N. 2023, arXiv:2307.13774
Tkachev, M. V., Pilipenko, S. V., Mikheeva, E. V., & Lukash, V. N. 2024, MNRAS, 527, 1381
Tombesi, F., Meléndez, M., Veilleux, S., et al. 2015, Natur, 519, 436
Übler, H., Maiolino, R., Curtis-Lake, E., et al. 2023, A&A, 677, A145
Venturi, G., Treister, E., Finlez, C., et al. 2023, A&A, 678, A127
Vink, J. S. 2023, A&A, 679, L9
Wagner, A. Y., Bicknell, G. V., & Umemura, M. 2012, ApJ, 757, 136
Wang, B., Leja, J., Atek, H., et al. 2023, arXiv:2310.06781
Yuan, G.-W., Lei, L., Wang, Y.-Z., et al. 2023, arXiv:2303.09391
Yung, L. Y. A., Somerville, R. S., Finkelstein, S. L., Wilkins, S. M., & Gardner, J. P. 2024, MNRAS, 527, 5929
Zhang, L., & Ho, L. C. 2023, ApJL, 953, L9