

Little Red Dots as a Transient Phase of Self-Interacting Dark Matter Assisted Black Hole Growth

Yu Rong,^{1,2,*} Shuang-Nan Zhang,^{3,4} Jian-Min Wang,^{3,4} Junxian Wang,^{1,2} and Zhicheng He^{1,2}

¹*Department of Astronomy, University of Science and Technology of China, Hefei, Anhui 230026, China*

²*School of Astronomy and Space Sciences, University of Science and Technology of China, Hefei, Anhui 230026, China*

³*Key Laboratory for Particle Astrophysics, Institute of High Energy*

Physics, Chinese Academy of Sciences, Yuquan Road, Beijing 100049, China

⁴*School of Astronomy and Space Science, University of Chinese Academy of Sciences, Yuquan Road, Beijing 100049, China*

(Dated: September 17, 2026)

The discovery of Little Red Dots (LRDs) with JWST has revealed a population of compact, red galaxies hosting rapidly growing black holes at early cosmic times. Their compact morphologies, broad emission lines, weak X-ray emission, and distinctive V-shaped spectral energy distributions indicate a short-lived phase of black-hole growth within a dense nuclear environment. Here we propose that LRDs arise from a transient episode of self-interacting dark matter (SIDM)-assisted black-hole growth during galaxy assembly. In this scenario, gas inflows first establish a **compact nuclear thick disk**, which modifies the central SIDM distribution and provides the obscuring structure around the accreting black hole. Once the SIDM density near the seed black hole becomes sufficiently enhanced, **rapid SIDM accretion drives a major increase in black-hole mass**, initiating the transient LRD phase. The resulting obscured growth phase naturally suppresses direct short-wavelength emission and redistributes the radiation field, producing the **red continuum and V-shaped spectral signatures** of LRDs. This framework links gas inflow, SIDM dynamics, and early black-hole assembly, predicting that LRDs preferentially occur in galaxies undergoing strong nuclear inflow and evolve into ordinary AGN after this transient phase. Unlike models that require globally rare halo histories, long super-Eddington gas growth, or purely phenomenological obscuration, our model ties the overmassive black hole, X-ray weakness, spectral shape, and finite duty cycle to **one local gas-triggered SIDM event**.

INTRODUCTION

Little Red Dots (LRDs) are among the most surprising compact populations revealed by JWST. They are selected as small, red rest-optical sources at $z \sim 4-9$, often with a blue or less-reddened ultraviolet component and with number densities that are high compared with ordinary luminous quasars [1–4]. Spectroscopy has sharpened the puzzle. Many point-like V-shaped sources show broad permitted lines, strong Balmer-limit curvature, Balmer absorption or dense-gas signatures, and continua that cannot be described by a single standard stellar population [5–9]. The inferred black-hole masses can be large compared with the stellar or dynamical masses of their hosts [10, 11]. At the same time, MIRI detections, weak X-ray emission, weak or absent standard hot-dust signatures in some systems, uncertain variability, and AGN-contaminated host-galaxy masses complicate a simple unobscured-quasar interpretation [8, 12–14]. Thus the emerging observational consensus is not that LRDs are one familiar object seen at high redshift, but that they are compact, common, frequently AGN-like, and embedded in dense reprocessing gas.

This combination of facts has led to several classes of models. Obscured or dust-reddened AGN models provide the most direct explanation for broad permitted lines and high inferred black-hole masses [2, 13, 14]. In this spirit, Madau and Maiolino [15] proposed a super-Eddington orientation model in which LRDs are obscured views of little blue dots, with a thick accretion flow, a polar funnel, an equatorial broad-line region, and a dusty screen. Such models explain the LRD/LBD connection and orientation dependence elegantly, but they still put much of the burden on gas accretion: the gas must grow an overmassive black hole, suppress escaping X-rays, generate a red continuum without always producing an ordinary hot-dust torus, and remain in place despite radiative feedback. Dense-gas or “black-hole star” models instead focus on the Balmer-limit curvature and absorption, interpreting these features as dense gas around a central ionizing source rather than as an old stellar break [8, 16]. This insight is important, but it leaves a dynamical question: what creates the dense gas, why is it maintained near the black hole, and why should the phenomenon be short-lived and concentrated at high redshift? Compact stellar, low-angular-momentum, low-spin halo, and forming-globular-cluster scenarios emphasize the small sizes, high surface densities, and possible stellar contribution of LRDs [17–20]. These models capture the compactness, but stellar compactness by itself does not naturally explain broad lines, weak X-rays, large black-hole-to-host ratios, and a transient AGN-like phase.

A second family of explanations places the origin of LRDs in special seeds or special halos. Direct-collapse black-hole and quasi-star scenarios connect LRDs to rapid early seed formation, dense inflow, and buried accretion [21–25]. Their strength is that they naturally contain compact, optically thick gas; their weakness is that the required conditions are

restrictive, involving rapid inflow, suppressed fragmentation, unusual thermodynamic histories, and often a limited host-halo population. Primordial-black-hole, collapsed-dark-star, and related exotic channels can place massive compact objects at very early times [26–28], but they still must explain the observed LRD spectra, X-ray weakness, Balmer curvature, and duty cycle. Low-spin compact-galaxy models face a similar demographic issue: low angular momentum can make compact systems, but it also selects a special subset of halos [19]. Recent clustering and assembly-bias studies make this point quantitative. Several channels, including direct collapse, SIDM core collapse, low-spin compact galaxies, and primordial-black-hole-related models, can predict distinctive or strong clustering because they select rare halo histories [29–31]. Current data do not yet require all LRDs to live in the rarest large-scale peaks. A successful model should therefore explain the spectra and black-hole masses without automatically making every LRD a globally extreme halo.

Self-interacting dark matter (SIDM) provides a way to address the black-hole mass problem through a dark, weakly radiating growth channel. SIDM was originally proposed to alleviate small-scale structure tensions in collisionless cold dark matter [32]; velocity-dependent SIDM can modify dwarf and galaxy-scale halos while satisfying larger-scale constraints [33, 34]. Baryons can strongly reshape SIDM centers: compact baryonic potentials can contract SIDM and generate dense cusps, whereas feedback-dominated systems can form lower-density cores [35, 36]. SIDM has been used to interpret the diversity of galaxy rotation curves and the clustering of diffuse dwarfs [33, 34, 37]. For LRDs, Jiang *et al.* [38] proposed that gravothermal collapse of SIDM halos can directly form massive black holes, elegantly avoiding a long phase of luminous super-Eddington gas accretion.

The pure SIDM gravothermal-collapse channel is attractive, but it also illustrates the problem we want to avoid. The collapse time is highly sensitive to halo concentration, formation time, and assembly history; the very conditions that make early collapse possible can imply strong assembly bias and special environments [30]. It also leaves a redshift question. If the visible LRD phase simply traced halos that had enough time to collapse gravothermally, lower-redshift halos should have had more time to produce analogous systems, whereas observed LRD demographics appear to decline rapidly toward later times [3, 4, 31]. These issues do not rule out SIDM. Instead, they suggest that SIDM should be coupled to baryonic compaction: SIDM may provide the dark mass reservoir, while gas inflow, obscuration, and viewing geometry determine when the system is visible as an LRD.

Here we develop this coupled picture by revisiting black-hole growth through accretion of collisional dark matter. Hu *et al.* [39] showed that a compact seed embedded in collisional dark matter can grow rapidly and saturate when the SIDM mean free path becomes comparable to the accretion radius; related calculations have been extended by Das and Kalita [40] and Sabarish *et al.* [41]. In our model the black hole is not born from collapse of the whole SIDM halo. Instead, a pre-existing central black hole must be large enough that, after baryonic compression, the SIDM density inside its influence radius satisfies the mean-free-path ignition condition. The seed mass therefore has a lower bound set by the finite compressive power of the gas disk. Once this condition is met, the subsequent saturated black-hole mass is set mainly by the SIDM sound speed and self-interaction cross-section. A gas-rich merger or compaction event forms a compact, turbulent, optically thick nuclear disk like those found in merger-driven direct-collapse simulations [21, 22]. This disk deepens the central baryonic potential, compresses the local SIDM, and triggers rapid SIDM-fed black-hole growth. The black-hole mass jump then further deepens the potential and helps maintain gas pile-up and high column density around the nucleus. In this way SIDM supplies the dark mass growth, while the nuclear thick disk supplies the obscuration, Balmer-active reprocessing layer, X-ray weakness, orientation dependence, and finite lifetime.

The goal of this paper is to construct a semi-analytic version of this gas-triggered SIDM-accretion model. We treat the obscurer as a compact nuclear thick disk rather than as a separate spherical cocoon, derive its column density from the gas mass and disk size, compute the SIDM ignition and saturation conditions, and generate absolute-flux mock spectra. We compare the resulting spectra with public LRD spectroscopy from Wang *et al.* [5] and de Graaff *et al.* [42]. The chronology is central: baryonic compaction turns on SIDM accretion; SIDM accretion creates the overmassive black hole; the same compact disk hides and reprocesses the subsequent AGN radiation; and the LRD phase ends when the high-covering, Balmer-active disk atmosphere is consumed, diluted, or expelled. Throughout this paper, we use G for Newton’s constant, m_p for the proton mass, σ_T for the Thomson cross-section.

SEMI-ANALYTIC MODEL FOR THE FORMATION OF AN LRD

Host Halo and Disk Compactness

We use RUBIES-EGS-55604 [5] as the fiducial case for developing the model. We adopt a SIDM halo with mass

$$M_h \simeq 5.0 \times 10^{10} M_\odot \quad (1)$$

at $z = 6.982$, the spectroscopic redshift of RUBIES-EGS-55604 [5]. This halo mass is compatible with a massive but not uniquely rare high-redshift host and with AGN-contaminated stellar mass estimates [13, 43]. The high-redshift virial scaling of Barkana and Loeb [44] gives a virial velocity V_{vir} and effective one-dimensional SIDM sound speed C_s of

$$V_{\text{vir}} = 120 \text{ km s}^{-1}, \quad C_s = \frac{V_{\text{vir}}}{\sqrt{2}} = 85 \text{ km s}^{-1}. \quad (2)$$

The baryon mass in this halo is approximately $M_B = f_B M_h = 7.8 \times 10^9 M_\odot$, where $f_B = 0.156$ [45]. There is a small central seed BH with typical mass of $M_{\text{BH,seed}} \simeq 10^5 M_\odot$ in this halo, along with the initial stellar mass of approximately $M_\star \sim 10^7 - 10^9 M_\odot$ with $M_{\text{BH,seed}}/M_\star \sim 10^{-4} - 10^{-2}$.

Motivated by the simulations of Mayer *et al.* [21, 22], a merger with this halo at high-redshift can trigger a gas inflow, forming a compact nuclear thick disk in the halo center. We adopt the thick-disk gas mass of

$$M_{\text{gas,d}} = 10^8 M_\odot, \quad (3)$$

approximately 1% of the baryonic budget. The compact disk radius is then inferred from the characteristic internal circular speed $v_{\text{c,d}}$ of such nuclear disks,

$$v_{\text{c,d}} = \chi C_s, \quad (4)$$

where the compactness parameter χ characterizes how deep the local baryonic disk potential is relative to the SIDM velocity scale. The compact nuclear disks in the simulations of Mayer *et al.* [21, 22] can be substantially more compact than our fiducial choice when their quoted disk masses and radii are translated into $v_{\text{c,d}}$ and hence χ . We deliberately adopt a conservative sub-extreme value, $\chi \simeq 1-3$, so that the fiducial calculation does not rely on the most compact simulated disks. A larger χ corresponds to a deeper local baryonic potential and, for fixed disk mass, a smaller characteristic disk radius and larger surface density. Such a state is expected to have a larger column density and a higher ionization/heating rate per unit area, and may correspond to an earlier, more compact phase of the nuclear thick disk. As the disk evolves, gas consumption, accretion, outflows, and turbulent expansion can lower the surface density and effective optical depth, while the escaping dense-gas continuum can shift from a hotter photoionized color temperature toward the cooler 3000–6000 K range seen in compact-disk simulations Mayer *et al.* [21, 22]. The object ceases to appear as an LRD once the disk no longer maintains the high-column, Balmer-active reprocessing layer. We estimate this column-clearing timescale in Section 2.4.

For RUBIES-EGS-55604, we adopt $\chi \sim 2$. The characteristic radius of the nuclear thick disk is then

$$R_d = \frac{GM_{\text{gas,d}}}{v_{\text{c,d}}^2} \simeq 14 \text{ pc}. \quad (5)$$

For such a turbulent or rotation-supported thick disk in vertical dynamical equilibrium, the scale height satisfies

$$\frac{H}{R_d} \simeq \frac{\sigma_g}{v_{\text{c,d}}}, \quad (6)$$

where H is the vertical scale height, and σ_g is the gas turbulent or vertical velocity dispersion [46]. We adopt $\sigma_g = 100 \text{ km s}^{-1}$. Thus it gives

$$\frac{H}{R_d} \sim 0.6, \quad H = 8.0 \text{ pc}. \quad (7)$$

The disk surface density is

$$\Sigma_g = \frac{M_{\text{gas,d}}}{\pi R_d^2} \simeq 1.6 \times 10^5 M_\odot \text{ pc}^{-2}. \quad (8)$$

The vertical hydrogen-equivalent column is

$$N_{\text{H},\perp} = \frac{\Sigma_g}{1.4 m_p} \simeq 1.5 \times 10^{25} \text{ cm}^{-2}, \quad (9)$$

where the factor $1.4m_p$ approximately includes helium. The corresponding Thomson depth is $\tau_{T,\perp} = N_{H,\perp}\sigma_T \sim 9.6$, so even the vertical column is Compton thick [47]. A line of sight through the disk body has a larger characteristic column,

$$N_{H,\text{edge}} \simeq \frac{N_{H,\perp}}{H/R_d} \sim 2.5 \times 10^{25} \text{cm}^{-2}, \quad (10)$$

with the corresponding Thomson depth of $\tau_{T,\text{edge}} = N_{H,\text{edge}}\sigma_T \sim 16.6$. The midplane density is

$$\rho_{\text{mid}} \simeq \frac{M_{\text{gas,d}}}{2\pi R_d^2 H} \simeq 1.0 \times 10^4 M_\odot \text{pc}^{-3}. \quad (11)$$

These values imply that the nuclear thick-disk is an obscurer with large optical depth. The highest columns occur through the disk midplane and the inner rim, while a turbulent disk atmosphere or disk-wind base can add high-covering, lower-temperature material above the midplane. The vertical column is already Compton thick, but a face-on observer can still receive a much bluer spectrum if the polar direction has a lower effective covering factor than the disk body.

SIDM Compression triggers BH rapid growth

The thick disk can trigger SIDM compression. Baryonic contraction of dark matter is a standard response to condensation of baryons [48, 49], and SIDM simulations with baryons show that central baryonic potentials can make SIDM profiles denser or cuspier than dark-matter-only SIDM halos [35, 36].

For approximately isothermal SIDM in hydrostatic balance,

$$\frac{d(\rho_\chi C_s^2)}{dr} = -\rho_\chi \frac{d\Phi}{dr}, \quad (12)$$

where ρ_χ is SIDM density and Φ is gravitational potential. The density amplification from R_d to the black-hole influence region is

$$\mathcal{C}_\chi \equiv \frac{\rho_\chi(r_B)}{\rho_\chi(R_d)} = \exp\left(\frac{\Delta\Phi}{C_s^2}\right) = \exp(\eta_\Phi \chi^2), \quad (13)$$

where $r_B = GM_{\text{BH}}/C_s^2$ is the black-hole influence radius and η_Φ parametrizes the disk potential shape. $\rho_\chi(r_B)$ and $\rho_\chi(R_d)$ denote the SIDM density at r_B and R_d , respectively. A uniform spherical potential would give $\eta_\Phi = 1/2$, while a centrally concentrated disk or cusp gives a larger potential drop and motivates $\eta_\Phi \sim 1$ as an upper illustrative value [46]. For $\chi \sim 2$, Equation 13 gives $\mathcal{C}_\chi \simeq 70$, suggesting that the SIDM density in the central BH influence radius is approximately amplified 70 times by the nuclear thick-disk.

The SIDM mean-free-path ignition condition is

$$\lambda_{\text{mfp}}(r_B) = \frac{1}{\rho_\chi(r_B)\sigma/m} \lesssim r_B, \quad (14)$$

where σ/m is the SIDM self-interaction cross-section per unit mass [39–41]. If equation (14) is satisfied, the central BH is able to efficiently accrete SIDM. For $\sigma/m \sim 1 \text{ cm}^2 \text{g}^{-1}$ [33, 34, 37] and $C_s \sim 85 \text{ km s}^{-1}$, the density required around a $10^5 M_\odot$ black hole is

$$\rho_{\text{ign}} \simeq 8 \times 10^4 M_\odot \text{pc}^{-3} \left(\frac{\sigma/m}{1 \text{ cm}^2 \text{g}^{-1}}\right)^{-1} \left(\frac{C_s}{85 \text{ km s}^{-1}}\right)^2 \left(\frac{M_{\text{BH}}}{10^5 M_\odot}\right)^{-1}. \quad (15)$$

Thus the pre-compression density needed at R_d is only

$$\rho_\chi(R_d) = \frac{\rho_{\text{ign}}}{\mathcal{C}_\chi} \simeq 1.2 \times 10^3 M_\odot \text{pc}^{-3}, \quad (16)$$

or 12% of ρ_{mid} , implying that the nuclear thick-disk can compress the inner SIDM density, and triggers central BH accreting SIDM.

Equations (15) and (16) also imply a lower limit on the BH seed mass. If the required pre-compression density satisfies $\rho_\chi(R_d) \gg \rho_{\text{mid}}$, the nuclear thick disk would not be able to compress the inner SIDM density efficiently

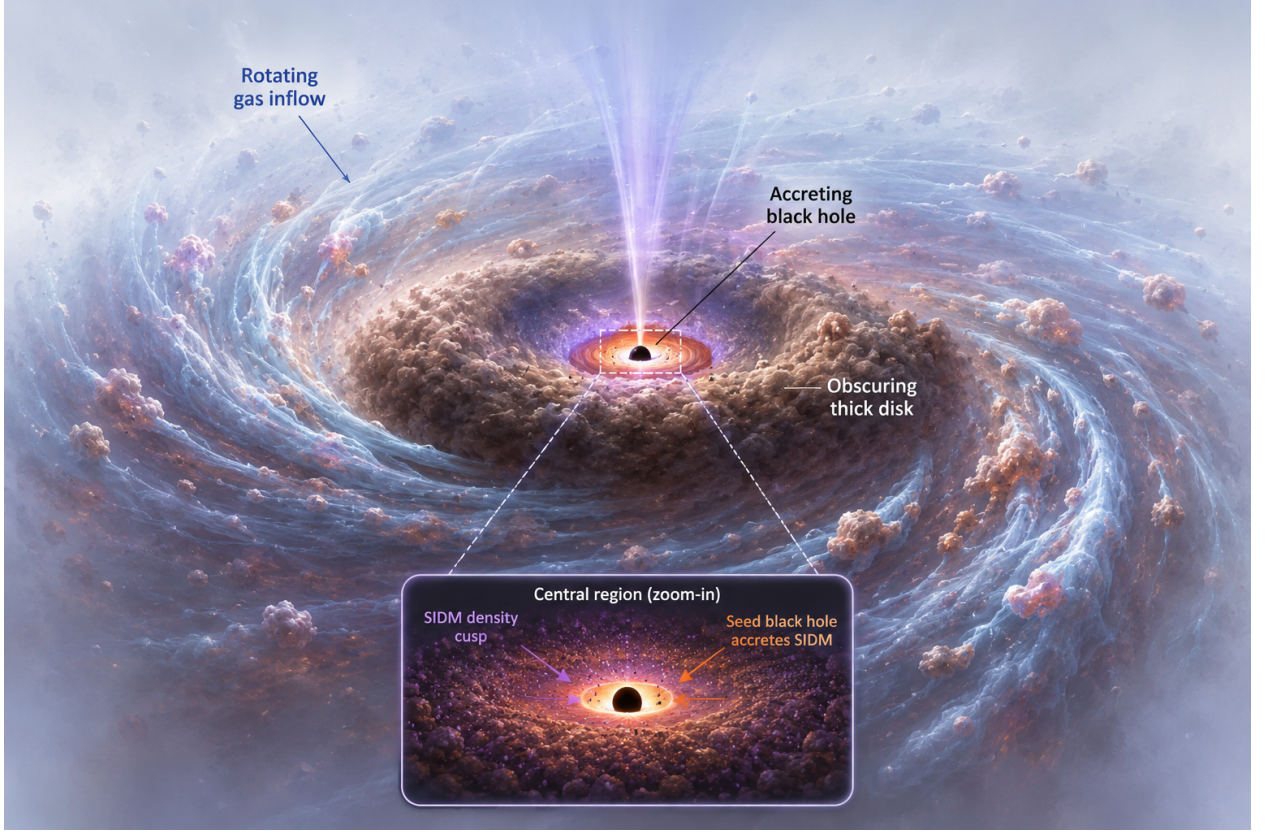


FIG. 1. Schematic picture of the model. A gas-rich inflow builds a compact optically thick nuclear disk. The disk potential compresses SIDM and triggers rapid SIDM-fed black-hole growth. Edge-on sight lines pass through high-column disk gas and appear as red, X-ray-weak LRDs.

enough to ignite SIDM accretion. Therefore, the lower limit of the BH seed mass can be estimated by requiring $\rho_\chi(R_d) \simeq \rho_{\text{mid}}$, and this limit strongly depends on the density of the nuclear thick disk. For a fiducial nuclear thick disk with $M_{\text{gas,d}} \sim 10^8 M_\odot$ and $R_d \sim 10$ pc, the lower limit of the BH seed mass is approximately $10^4 M_\odot$. For the most extreme nuclear thick disks in the simulations of Mayer *et al.* [21, 22], with $M_{\text{gas,d}} \sim 10^9 M_\odot$ and $R_d \sim 5$ pc, the lower limit could decrease to $\sim 100 M_\odot$, comparable to a Pop III remnant black hole [50–52].

For collisionless cold dark matter, black-hole growth by dark-matter capture is inefficient because the dark component lacks dissipation and only particles entering the loss cone can be captured. SIDM changes this because self-interactions can make the dark component behave as an effective collisional fluid on relevant scales.

This SIDM accretion process is very fast. The rapid SIDM accretion stage saturates when the mean free path becomes comparable to the accretion radius. Hu *et al.* [39] give the BH mass after accretion

$$M_{\text{BH,SIDM}} = 1.4 \times 10^6 M_\odot \left[\frac{\sigma/m}{1 \text{ cm}^2 \text{ g}^{-1}} \right] \left(\frac{C_s}{30 \text{ km s}^{-1}} \right)^4, \quad (17)$$

with characteristic time

$$t_{\text{SIDM}} = \frac{(\sigma/m)C_s}{4\pi G}. \quad (18)$$

For $\sigma/m \sim 1 \text{ cm}^2 \text{ g}^{-1}$ and $C_s \sim 85 \text{ km s}^{-1}$, equations 17–18 give

$$M_{\text{BH,SIDM}} \simeq 9 \times 10^7 M_\odot, \quad t_{\text{SIDM}} \simeq 0.32 \text{ Myr}. \quad (19)$$

Since the timescale of BH accretion is too fast, the halo and stellar components almost do not evolve during this phase.

SIDM accretion is electromagnetically dark. During the SIDM accretion period, the central BH should also accrete the surrounding gas. However, the accretion rate is limited by Eddington accretion rate and is much lower than the rate of accreting SIDM, and thus can be neglected.

The LRD mode

In our model, the LRD mode begins only when three conditions are simultaneously satisfied: the black hole has reached the rapid-SIDM saturation mass, baryonic gas is still present in an optically thick compact disk, and an edge-on or high-covering line of sight intersects the disk body or atmosphere. This makes the onset a potential-transformation event rather than simply the presence of a gas disk or the presence of an AGN, as shown in Fig. 1. The process is also a feedback loop. The inflow-built disk first compresses SIDM and initiates rapid dark accretion. The resulting black-hole mass jump then deepens the central potential on a timescale shorter than the disk replenishment or clearing time, which promotes further gas inflow, raises the central column, and helps maintain the obscuring disk atmosphere during the luminous phase.

After SIDM saturation, the black-hole accreting gas becomes more dominant, which begins to emit as an AGN. The surrounding nuclear thick-disk is retained because the disk remains self-gravitating during the luminous phase. The buried AGN luminosity is then set by baryonic accretion onto the SIDM-enlarged black hole,

$$L_{\text{bol}} = \lambda_{\text{eff}} L_{\text{Edd}} = 1.26 \times 10^{38} \lambda_{\text{eff}} \left(\frac{M_{\text{BH,SIDM}}}{M_{\odot}} \right) \text{erg s}^{-1}, \quad (20)$$

where λ_{eff} is the effective Eddington ratio and L_{Edd} is the Eddington luminosity [53, 54]. The RUBIES-EGS-55604 continuum fit prefers a luminous but sub-Eddington buried AGN, $\lambda_{\text{eff}} = 0.6$,

$$L_{\text{bol}} = 6.8 \times 10^{45} \text{erg s}^{-1}, \quad \dot{M}_{\text{BH}} = \frac{L_{\text{bol}}}{\epsilon c^2} = 1.2 M_{\odot} \text{yr}^{-1} \left(\frac{\epsilon}{0.1} \right)^{-1}, \quad (21)$$

where $\dot{M}_{\text{BH}}$ is the baryonic black-hole accretion rate, ϵ is the radiative efficiency, and c is the speed of light. Our model does not require a strongly super-Eddington radiative accretion ratio, $\lambda_{\text{eff}} \gg 1$, which remains observationally uncertain in high-redshift LRDs.

The disk columns in Equations 9 and 10 imply Thomson depths $\tau_{\text{T}} \gg 1$, so the directly escaping blue AGN continuum is strongly suppressed along high-covering sight lines [47]. Multiple electron scatterings, bound-free absorption, free-free absorption, metal opacity, and dust or dust-free thermalization channels can therefore reprocess a large fraction of the AGN power into a red continuum plus an obscured photosphere from the compact disk [47, 55]. We write the latter as

$$F_{\nu}^{\text{ph}} = A_{\text{ph}} B_{\nu}(T_{\text{gas}}) \left(\frac{\lambda_0}{\lambda} \right)^{\beta_{\text{ph}}} \mathcal{T}_{\text{B}}(\lambda), \quad (22)$$

where F_{ν}^{ph} is the obscured-photosphere component, B_{ν} is the Planck function, T_{gas} is the color or thermal-equilibrium temperature of the optically thick dense gas, λ_0 is a reference wavelength, β_{ph} is an effective wavelength-dependent escape or opacity index, and $\mathcal{T}_{\text{B}}$ is the transmission through the Balmer-active layer of the disk atmosphere or wind base. The normalization A_{ph} is fixed by the reprocessed luminosity budget,

$$\int L_{\nu}^{\text{ph}} d\nu = L_{\text{ph}} = f_{\text{ph}} L_{\text{bol}}, \quad (23)$$

where f_{ph} is the fraction of the buried AGN luminosity emerging through the obscured photosphere. Equation 22 is not a separate object from the gas disk; it is a compact spectral representation of unresolved radiative transfer through the optically thick disk atmosphere. Different LRDs need not have the same T_{gas} because dense gas has several plausible thermal states. Strong photoionization and efficient $\text{Ly}\alpha$ /recombination cooling drive photoionized gas toward a $T_{\text{gas}} \sim 10^4$ K plateau, while lower ionization parameter, stronger shielding, molecular or metal cooling, and a small amount of dust can shift the escaping continuum to a cooler $T_{\text{gas}} \sim 3000\text{--}5000$ K color temperature [47, 55]. In this paper T_{gas} is therefore an effective continuum-transfer parameter fitted together with the obscured AGN component.

Because the Balmer limit is not produced by a single cold slab, we represent the outer absorbing layer with a smooth effective transmission:

$$\mathcal{T}_{\text{B}}(\lambda) = (1 - C_{\text{B}}) + C_{\text{B}} \exp[-\tau_{\text{B}} S_{\text{B}}(\lambda)], \quad (24)$$

with

$$S_{\text{B}}(\lambda) = \frac{1}{1 + \exp[(\lambda - \lambda_{\text{B}})/\Delta\lambda_{\text{B}}]} \times \begin{cases} (\lambda/\lambda_{\text{B}})^3, & \lambda < \lambda_{\text{B}}, \\ 1, & \lambda \geq \lambda_{\text{B}}, \end{cases} \quad (25)$$

where $\lambda_B = 0.3646 \mu\text{m}$ is the Balmer edge, $\Delta\lambda_B$ is the effective smoothing width, and C_B is the covering factor of the Balmer-active layer. Physically, this smooth form represents a stratified disk atmosphere or disk-wind base with velocity, density, temperature, and ionization gradients. In Fig. 2, the fiducial RUBIES-EGS-55604 solution uses the smooth Balmer transmission with $C_B = 0.70$, $\tau_B = 15$, $\Delta\lambda_B = 0.006 \mu\text{m}$, and no extra trough term. The fit to another typical LRD, RUBIES-UDS-40579, uses the same functional form with a cooler color temperature and a broader effective Balmer layer; this is a change in the radiative-transfer state of the same thick-disk atmosphere, not a change in the underlying model.

We model the LRD continuum with four observable components:

$$F_\nu = F_\nu^{\text{obscured AGN}} + F_\nu^{\text{obscured photosphere}} + F_\nu^{\text{leaked young stars}} + F_\nu^{\text{old host}}. \quad (26)$$

The first term is radiation from the accreting black hole after propagation through the optically thick disk. The second term is the obscured photosphere defined in Equation 22. The third term represents a small fraction of recently formed young nuclear starlight that leaks or scatters through low-opacity channels. The fourth term represents weak pre-existing host starlight in the spectroscopic aperture. This is the moment when the model becomes an LRD: the black hole is already overmassive, the gas is still optically thick, and the observed continuum is dominated by reprocessed rather than direct AGN light.

The End of the LRD Mode

The LRD lifetime is the time over which the optically thick disk body and atmosphere remain replenished faster than they are consumed, accreted, expelled, or geometrically diluted. We write this column-clearing time as

$$t_{\text{LRD}} \simeq \frac{M_{\text{gas,d}} - M_{\text{gas,crit}}}{\dot{M}_{\text{BH}} + \dot{M}_\star + \dot{M}_{\text{out}} - \dot{M}_{\text{in}}}, \quad (27)$$

where $\dot{M}_\star$, $\dot{M}_{\text{out}}$, and $\dot{M}_{\text{in}}$ are the compact star-formation, outflow, and replenishing inflow rates, respectively. The critical mass $M_{\text{gas,crit}}$ is the disk mass below which the object no longer looks like an LRD. If the disk radius is roughly fixed, the vertical column evolves as

$$N_{\text{H},\perp}(t) \simeq N_{\text{H},\perp,0} \left[\frac{M_{\text{gas,d}}(t)}{10^8 M_\odot} \right] \left(\frac{R_d}{14 \text{pc}} \right)^{-2}. \quad (28)$$

The Compton-thick condition alone is lost only after the gas mass drops below $\simeq 7 \times 10^6 M_\odot$ for $N_{\text{H},\perp} = 10^{24} \text{cm}^{-2}$, but the V-shaped LRD continuum is more sensitive to the Balmer-edge layer. For the fiducial Balmer optical depth used in Section 3, reducing the effective column by a factor of a few is enough to weaken the Balmer-limit curvature. Taking $M_{\text{gas,crit}} \simeq (0.2\text{--}0.3)M_{\text{gas,d}}$ gives

$$t_{\text{LRD}} \sim 10\text{--}100 \text{Myr} \left(\frac{M_{\text{gas,d}}}{10^8 M_\odot} \right) \left(\frac{\dot{M}_{\text{net}}}{1\text{--}10 M_\odot \text{yr}^{-1}} \right)^{-1}, \quad (29)$$

where $\dot{M}_{\text{net}} \equiv \dot{M}_{\text{BH}} + \dot{M}_\star + \dot{M}_{\text{out}} - \dot{M}_{\text{in}}$. For RUBIES-EGS-55604, without replenishment, Eddington black-hole accretion alone would remove $10^8 M_\odot$ in $\simeq 16 \text{Myr}$, and star formation or outflows would shorten this time. A long LRD lifetime therefore requires continued supply from the larger nuclear disk or repeated compaction, consistent with gas-rich merger simulations that assemble compact disks rapidly but do not imply a static Gyr-lived obscurer [21, 22]. The abundance-based lifetime estimates for LRD populations are still uncertain, but duty-cycle arguments generally point to a short phase rather than a Hubble-time population [3, 4, 7].

The LRD mode ends when the high-covering column falls, the polar channel widens, or the disk expands so that the photospheric and Balmer-edge conditions are no longer satisfied. The descendant should then be a less-obscured compact AGN, a blue compact broad-line AGN if viewed through the polar direction, or a compact post-starburst/overmassive-black-hole system with weaker Balmer-edge curvature. In this picture, low-redshift analogues are uncommon because later small-galaxy mergers are less gas rich, have more feedback-regulated central gas reservoirs, and occur after the first large fractional black-hole mass jump has already passed.

MOCK SPECTRA AND OBSERVED LRD SPECTRA

Mock-Spectrum Construction

All spectra in this section are generated in absolute flux units from the same component library. The obscured-AGN continuum is an analytic red reprocessing template, $F_\nu^{\text{red}} \propto [1 + \exp[-(\lambda - \lambda_B)/w]]^{-1}(\lambda/0.55\,\mu\text{m})^{\alpha_{\text{red}}}$, optionally multiplied by the finite red-side self-absorption factor in Equation 35. The obscured photosphere is the modified Planck template in Equation 22. Both components are normalized by integrating the rest-frame L_ν template over frequency and setting the luminosity to the fitted fraction of L_{bol} . The Balmer curvature is then applied with the same smooth transmission function, Equation 24, for both fitted LRDs.

The stellar terms are deliberately continuum-only nuisance templates. For the leaked young component we use an FSPS-motivated young stellar-population shape with nebular emission switched off, normalized by the standard rest-UV continuous-star-formation conversion $L_\nu(1500\,\text{\AA}) \simeq 8 \times 10^{27}(\text{SFR}/M_\odot\,\text{yr}^{-1})\,\text{erg s}^{-1}\,\text{Hz}^{-1}$ and attenuated, when needed, with a Calzetti-law screen [56–59]. The weak old-host term uses the same stellar-population-motivated continuum shape and a simple rest-optical mass-to-light normalization. These stellar templates are included only to test whether aperture-level galaxy light can smooth the continuum; they contain no emission lines and are not allowed to dominate the LRD continuum.

We do not fit the emission lines in this paper. The grey masks in Figures 2 and 3 remove line complexes before computing continuum residuals. A full line and bound-free continuum calculation should be done with a photoionization/radiative-transfer code such as CLOUDY [47, 55]. Here CLOUDY is used only as the physical motivation for the dense-gas opacity discussion, not as an input library for the plotted mock spectra.

A Hot Dense-Gas LRD: RUBIES-EGS-55604

The thick disk is Compton thick in both its vertical and disk-body columns, but a polar funnel or low-covering channel can still allow a larger fraction of blue radiation to escape along the disk axis. For the RUBIES-EGS-55604 continuum we fit the high-column spectrum using the components in Equation 26. The dynamical and SIDM quantities are tied to the equations in Section 2. By contrast, the continuum-transfer quantities are effective parameters: the red reprocessed slope α_{red} , the red AGN fraction f_{red} , the obscured-photosphere fraction f_{ph} , the Balmer-layer covering factor C_B , and the Balmer-edge width depend on multidimensional radiative transfer through the disk atmosphere, viewing angle, scattering, thermalization depth, ionization structure, and possible dust or dust-free opacity. We therefore fit them to the line-masked continuum, while requiring subdominant stellar contributions and physically plausible columns.

This separation is important. The semi-analytic model determines the gravitational and energetic scale of the event: M_{h} sets C_{s} , the compact disk fixes R_{d} and N_{H} , the adopted galaxy-scale σ/m fixes the SIDM saturation mass through Equation 17, and λ_{eff} fixes the bolometric luminosity through Equation 20. What it does not determine is how a clumpy, optically thick, non-LTE disk atmosphere partitions that luminosity among a red scattered or reprocessed AGN continuum, a compact thermalization surface, Balmer-edge absorption, and small leakage channels. Those quantities require radiation-hydrodynamic transfer or photoionization calculations [47, 55]. In this paper they are therefore fitted effective parameters, not hidden dynamical parameters.

For RUBIES-EGS-55604 the best continuum solution has

$$\alpha_{\text{red}} = 4.3, \quad T_{\text{gas}} = 1.4 \times 10^4\text{K}, \quad f_{\text{red}} = 0.0245, \quad f_{\text{ph}} = 0.0455, \quad C_B = 0.70, \quad (30)$$

with $\tau_B = 15$ and an effective Balmer-edge width $0.006\,\mu\text{m}$. The fit uses a modest leaked young component, $\text{SFR}_{\text{eff}} = 6.9 M_\odot\text{yr}^{-1}$, age 10 Myr, and leakage fraction 0.2, plus a weak old host component, $M_{\star,\text{old}} = 7.6 \times 10^8 M_\odot$, contributing only a few percent of the continuum near $5500\,\text{\AA}$. These stellar terms are included because the observed spectrum is an aperture spectrum of the whole compact source; they are not allowed to dominate the LRD continuum.

The fitted Balmer depth is tied back to the disk columns derived above. Equations 9 and 10 give the hydrogen column available in the compact disk, so the effective column seen by an LRD sight line should lie between the vertical and disk-body values, depending on orientation and covering. If a fraction of that column lies in a cooler Balmer-active atmosphere or disk-wind base, the Balmer-edge optical depth is

$$\tau_B \simeq N_{\text{H,B}} \left(\frac{N_2}{N_{\text{H}}} \right) \sigma_B, \quad (31)$$

TABLE I. Continuum-model parameters for RUBIES-EGS-55604. “Derived” means fixed by the thick-disk/SIDM equations in Section 2 once the host halo, compact-disk mass, compact-disk speed, and σ/m are specified. “Effective fit” means constrained by the line-masked continuum because the parameter depends on unresolved radiative transfer, viewing angle, and ionization structure.

Parameter	Value	Role	Status
M_h	$5.0 \times 10^{10} M_\odot$	host halo scale	model input
C_s	84.9 km s^{-1}	SIDM sound speed	derived
$M_{\text{gas,d}}$	$10^8 M_\odot$	compact inflow disk	simulation-motivated input
R_d	14 pc	disk size	derived from $GM_{\text{gas,d}}/v_{c,d}^2$
$N_{\text{H,edge}}$	$2.5 \times 10^{25} \text{ cm}^{-2}$	high-column sight line	derived
σ/m	$1 \text{ cm}^2 \text{ g}^{-1}$	galaxy-scale SIDM cross-section	model input
$M_{\text{BH,SIDM}}$	$9.0 \times 10^7 M_\odot$	saturated SIDM-grown hole	black derived
λ_{eff}	0.6	buried accretion state	continuum constrained
L_{bol}	$6.8 \times 10^{45} \text{ erg s}^{-1}$	AGN power	derived from $\lambda_{\text{eff}} L_{\text{Edd}}$
α_{red}	4.3	obscured AGN slope	effective fit
f_{red}	0.0245	emergent obscured-AGN fraction	effective fit
T_{gas}	$1.4 \times 10^4 \text{ K}$	dense-gas color temperature	effective fit
f_{ph}	0.0455	obscured-photosphere luminosity fraction	effective fit
C_B	0.70	Balmer-layer covering factor	effective fit
τ_B	15	Balmer-edge optical depth	effective fit, but model constrained
N_2/N_H	3.0×10^{-6}	required excited-H fraction	derived from $\tau_B/N_{\text{H,edge}}\sigma_B$
$\Delta\lambda_B$	$0.006 \mu\text{m}$	Balmer-edge smoothing width	effective fit
SFR_{eff}	$6.9 M_\odot \text{ yr}^{-1}$	leaked young stars	aperture nuisance, subdominant
$M_{\star,\text{old}}$	$7.6 \times 10^8 M_\odot$	old host light	aperture nuisance, subdominant

where $N_{\text{H,B}}$ is the effective hydrogen column in that Balmer-active layer, N_2/N_H is the fraction of hydrogen in the first excited state, and $\sigma_B \simeq 2 \times 10^{-19} \text{ cm}^2$ is the Balmer-edge bound-free cross-section scale. The cross-section is fixed by atomic physics; the spectrum constrains τ_B , and the thick-disk model restricts the allowed $N_{\text{H,B}}$. The best-fit $\tau_B = 15$ is naturally associated with a high-column, near-edge-on sight line. Taking $N_{\text{H,B}} \simeq N_{\text{H,edge}} = 2.5 \times 10^{25} \text{ cm}^{-2}$ from Equation 10 implies

$$\frac{N_2}{N_H} \simeq \frac{\tau_B}{N_{\text{H,B}}\sigma_B} = 3.0 \times 10^{-6}. \quad (32)$$

This small excited-state fraction is the consistency requirement imposed by the observed Balmer curvature. If instead the vertical column were used, the same excited-state fraction would give $\tau_B \simeq 15N_{\text{H},\perp}/N_{\text{H,edge}} \simeq 8.7$, so the stronger fitted Balmer curvature supports the orientation picture in which RUBIES-EGS-55604 is viewed through the disk body or a high-covering disk atmosphere. This is not a first-principles ionization calculation, but it is a useful consistency check linking the spectral fit to the column-density model.

The fitted effective values have simple physical interpretations. The steep α_{red} is the rest-optical tail of radiation that has been scattered, absorbed, and re-emitted through a high-column disk body. The small f_{red} and f_{ph} keep the emergent continuum below the bolometric power available from the buried AGN, so no extra energy source is introduced. The fitted T_{gas} is close to the characteristic temperature of dense photoionized gas and sets the blue color of the obscured photosphere. The fitted C_B and τ_B encode the fact that the Balmer-active layer covers much, but not all, of the continuum source; increasing them deepens the Balmer-limit curvature, while changing the edge width controls how sharp the break appears after velocity gradients, scattering, and instrumental resolution are folded together. The stellar terms are allowed because the RUBIES aperture contains the compact galaxy, but their fitted amplitudes remain perturbative.

The plotted fluxes in Fig. 2 are absolute predictions for $z = 6.982$ and are not renormalized to the observed RUBIES-EGS-55604 spectrum. To match the convention used in the RUBIES spectral figures, the model (red spectrum in Fig. 2) and data (black spectrum) are displayed as rest-frame-density F_λ in units of $\text{erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1}$; the underlying model normalization is computed in observed F_ν , converted to observed F_λ using $F_{\lambda,\text{obs}} = F_{\nu,\text{obs}}c/\lambda_{\text{obs}}^2$, and then multiplied

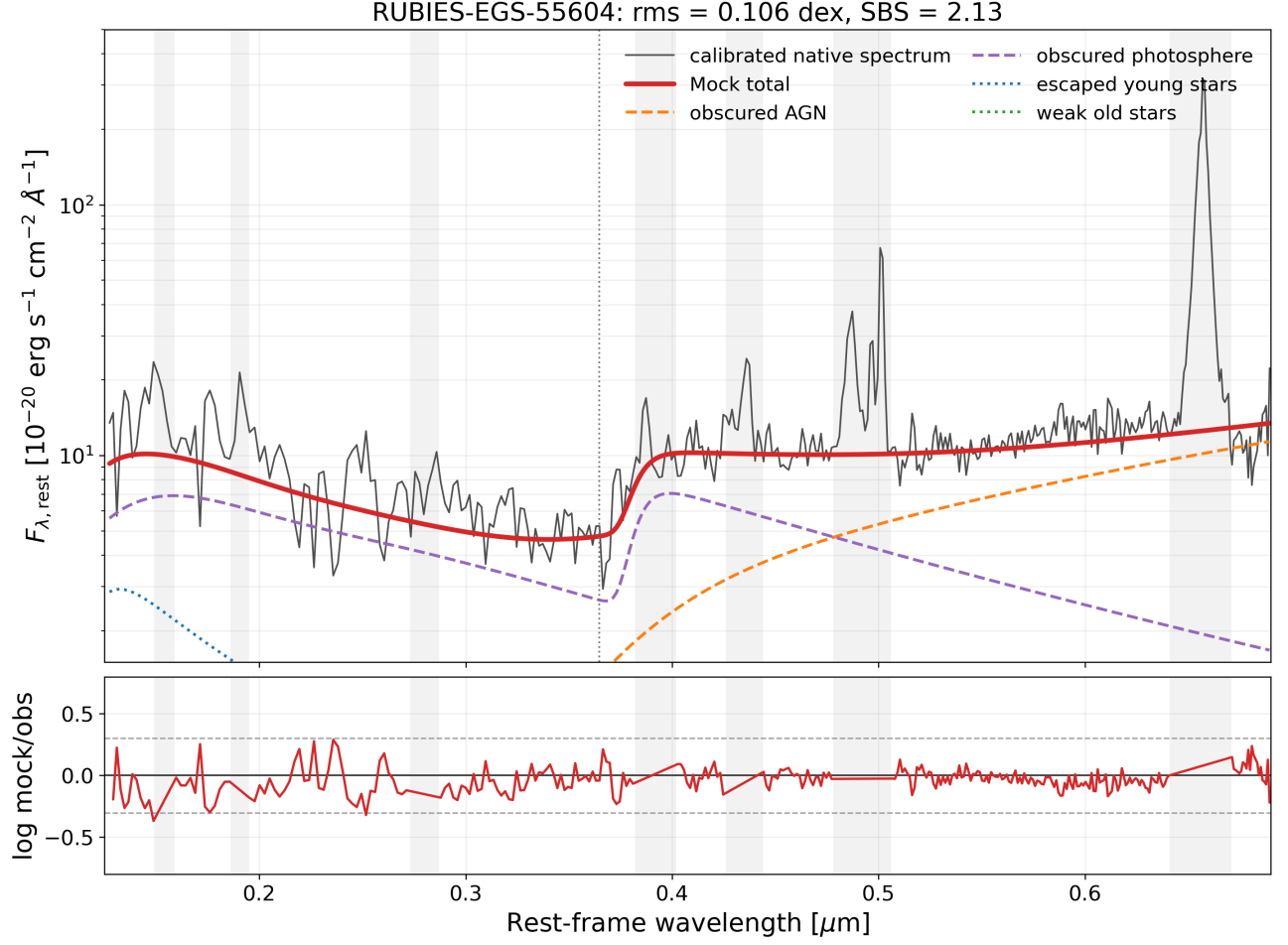


FIG. 2. **Mock spectrum versus the calibrated RUBIES-EGS-55604 spectrum.** The red curve is the Mock total; dashed or dotted curves show the obscured AGN, smooth Balmer-transmission obscured photosphere, leaked young-star component, and weak old host component. Grey regions are masked emission-line complexes. The fit has rms residual 0.106 dex outside the masked windows and spectral-break strength $SBS = 2.13$, matching the calibrated observed value.

by $(1+z)$ to report the rest-frame-density convention. Emission-line windows are masked when computing continuum residuals because reproducing line profiles requires photoionization and dynamical radiative-transfer modeling beyond the present continuum demonstration [47, 55]. The masked windows include the UV high-ionization features, C III], Mg II, the $H\zeta/H\epsilon/[Ne\ III]/He\ I$ complex, $H\gamma/[O\ III]$, $H\beta/[O\ III]$, and $H\alpha/[N\ II]$ complexes identified in the RUBIES spectra [5].

The fit uses the spectral-break-strength definition of Wang *et al.* [5], approximately the ratio of the continuum just redward and blueward of the Balmer limit. The model gives $SBS = 2.13$, while the calibrated spectrum gives $SBS = 2.13$. The stellar terms remain perturbative: the best fit uses a modest leaked young-star component and a weak old host contribution. The V-shaped continuum is therefore produced primarily by the obscured AGN plus the smooth Balmer-transmission photosphere, not by an old stellar population.

A Cooler LRD: RUBIES-UDS-40579

We next apply the same model to RUBIES-UDS-40579, one of the high-signal spectra in the de Graaff *et al.* LRD sample [42]. RUBIES-UDS-40579 represents a class of LRD that exhibit a red spectral bump and may have relatively low temperatures. This object has $z = 3.114$ and a fitted modified-blackbody temperature $T_{MBB} = 3648\text{ K}$ [42]. We use the public DJA v4 PRISM spectrum directly, convert the observed-frame F_ν to observed-frame F_λ , and compare the model in the same flux-density convention, as shown in Fig. 3. When a rest-frame luminosity density is needed, we

use

$$L_{\lambda,\text{rest}} = 4\pi D_L^2 (1+z) F_{\lambda,\text{obs}}. \quad (33)$$

As for RUBIES-EGS-55604, emission-line windows are masked and only the continuum is fitted.

The best continuum solution has

$$M_h = 1.0 \times 10^{11} M_\odot, \quad \sigma/m = 1.0 \text{cm}^2 \text{g}^{-1}, \quad \lambda_{\text{eff}} = 0.6, \quad (34)$$

which gives $C_s = 77.1 \text{km s}^{-1}$, $M_{\text{BH,SIDM}} = 6.1 \times 10^7 M_\odot$, $t_{\text{SIDM}} = 0.29 \text{ Myr}$, and $L_{\text{bol}} = 4.6 \times 10^{45} \text{erg s}^{-1}$. The compact disk parameters are $M_{\text{gas,d}} = 3.0 \times 10^7 M_\odot$, $v_{\text{c,d}} = 100 \text{km s}^{-1}$, $R_d = 12.9 \text{ pc}$, $N_{\text{H,edge}} = 5.1 \times 10^{24} \text{cm}^{-2}$, and $\tau_T = 3.4$.

For this source we use the AGN-dominated version of the same high-column smooth-Balmer model. The red continuum is not required to be dominated by the thermalized photosphere. Instead, an obscured AGN continuum provides most of the red-side continuum, while the obscured photosphere represents the fraction of absorbed power that escapes as a cooler compact thermal component. Both components propagate through the same outer Balmer-active disk atmosphere described by Equation 24. We write the AGN continuum as

$$F_\nu^{\text{obscured AGN}} \propto F_\nu^{\text{red}}(\alpha_{\text{red}}) \left[1 + \left(\frac{\lambda}{\lambda_{\text{turn}}} \right)^{s_{\text{turn}}} \right]^{-1} \mathcal{T}_B(\lambda), \quad (35)$$

where λ_{turn} and s_{turn} describe the red-side turnover caused by self-absorption, diffusion, or reprocessing in the optically thick disk body. This is still an effective radiative-transfer template; it is not a new dynamical component. In RUBIES-EGS-55604 the turnover is effectively unconstrained because any acceptable turnover lies far to the red of the fitted wavelength range. In RUBIES-UDS-40579 the available spectrum extends farther into the rest optical and therefore weakly constrains the turnover.

The best obscured-AGN continuum solution has

$$\alpha_{\text{red}} = 3.0, \quad \lambda_{\text{turn}} = 0.82 \mu\text{m}, \quad s_{\text{turn}} = 4.0, \quad (36)$$

for the obscured AGN and

$$T_{\text{gas}} = 3000 \text{K}, \quad \beta_{\text{ph}} = 0, \quad (37)$$

for the obscured photosphere. The fitted luminosity fractions are

$$f_{\text{AGN}} = 0.055, \quad f_{\text{ph}} = 0.051, \quad (38)$$

where f_{AGN} is the emergent obscured-AGN fraction and f_{ph} is the emergent obscured-photosphere fraction. The smooth Balmer layer has $\tau_B = 3.0$, $C_B = 0.95$, and $\Delta\lambda_B = 0.025 \mu\text{m}$. For the 40579 disk $H/R_d \simeq 1$, so the vertical and disk-body columns are both $N_{\text{H,B}} \sim 5.1 \times 10^{24} \text{cm}^{-2}$ and the corresponding Thomson depth is $\tau_{\text{T,edge}} \simeq 3.4$. The fitted Balmer depth is therefore of the same order as the column-density estimate and requires only $N_2/N_{\text{H}} \simeq 2.9 \times 10^{-6}$. The blue side is supplied by a modest leaked young-star component with effective escaped SFR = $6.8 M_\odot \text{yr}^{-1}$, age 50 Myr, and $E(B-V)_* = 0.2$. The old host component is again included from the seed-host prior, $M_{*,\text{old}} = 3.3 \times 10^7 M_\odot$, but is optically weak.

The line-masked continuum rms is 0.070 dex. The red-side AGN contribution is 58% over 0.75–0.90 μm and 58% over the masked continuum as a whole, so the fit is genuinely AGN dominated in the red continuum. This experiment shows that RUBIES-UDS-40579 does not require a photosphere-dominated red continuum; an Eddington-subcritical buried AGN with $\lambda_{\text{eff}} = 0.6$ can dominate the red side when it passes through the outer Balmer shell.

The fitted $T_{\text{gas}} = 3000 \text{ K}$ is consistent with a compact dense-gas color temperature. Since $f_{\text{ph}} = 0.051$ is only the observed escaping photosphere fraction, the obscuring disk must thermalize substantially more power than this in the unseen channels. Thus the cool photosphere temperature is set by the absorbed AGN heating and the unresolved radiative-transfer state of the disk atmosphere, not by equating the observed photosphere luminosity alone to the total heating power. The contrast with the $T_{\text{gas}} = 1.4 \times 10^4 \text{ K}$ 55604 solution is therefore allowed within the same model: 55604 is a hotter, more photoionized or shallower escaping layer, and possibly corresponds to an earlier stage of LRD, whereas 40579 is a cooler, more shielded, redder effective layer of the same compact thick disk, which may correspond to a later stage of LRD.

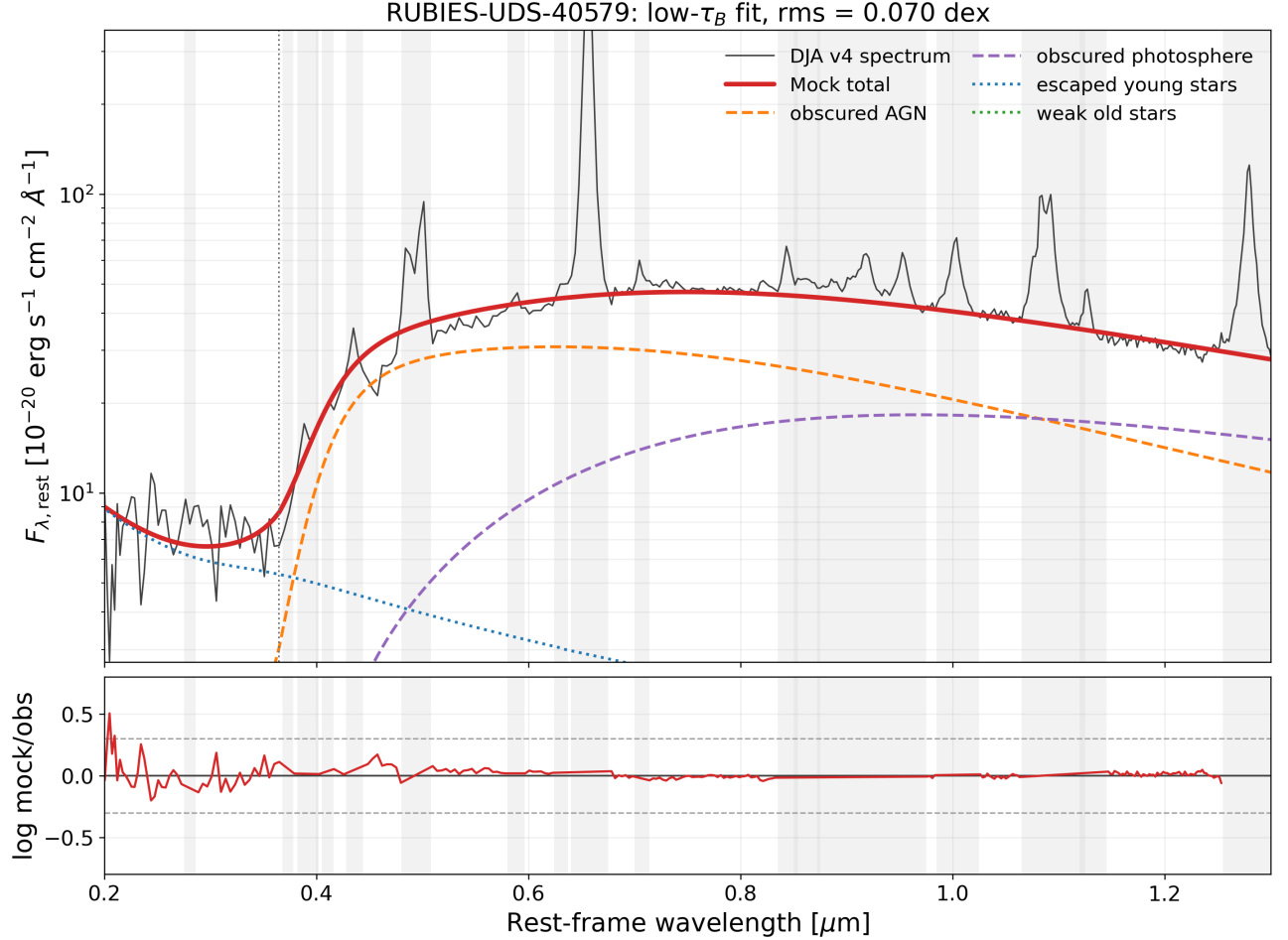


FIG. 3. **Mock spectrum compared with the observed spectrum of RUBIES-UDS-40579.** Analogous to Fig. 2, the red curve is the Mock total. The orange curve is the obscured AGN continuum after passing through the same smooth Balmer-active disk atmosphere as the photosphere, and it contributes 58% of the red continuum over $0.75\text{--}0.90\,\mu\text{m}$. The purple curve is the cooler obscured photosphere powered by absorbed AGN radiation. The blue curve is a modest leaked young-star component. Grey regions are masked emission-line complexes.

DISCUSSION

Comparison with Other Proposed LRD Channels

The model developed here belongs to a broader family of explanations in which LRDs are short-lived phases rather than a new long-lived galaxy class. The most useful way to compare these ideas is not only by asking whether they can make a red compact spectrum, but by asking what they require of the host halo population. Recent clustering and assembly-bias calculations show that several proposed LRD channels select very special halos and therefore predict strong or distinctive clustering: direct-collapse black holes, SIDM core collapse, low-spin compact-galaxy models, and primordial-black-hole-related channels can all carry a large host-halo or assembly-bias burden [29–31]. This is a serious discriminant because current LRD environment measurements do not yet require every LRD to live in the rarest, most biased peaks. Our model is designed to avoid this problem. It does not identify LRDs with a globally extreme halo class. The halo provides the SIDM reservoir and velocity scale, while the LRD switch is local: a gas-rich compaction event builds a compact nuclear thick disk, compresses the central SIDM, and then hides the gas-powered AGN along high-column sight lines.

This point is sharpest for the SIDM gravothermal core-collapse scenario of Jiang *et al.* [38]. That model elegantly produces massive black holes without luminous super-Eddington growth, but the collapse time is extremely sensitive to halo concentration, formation time, and assembly history. The same sensitivity that helps make early black holes also

TABLE II. Best-fit physical and effective continuum parameters for the two LRD spectra fitted in this work. The first block is tied to the thick-disk/SIDM model; the second block contains effective continuum-transfer parameters fitted to the line-masked continuum. RUBIES-UDS-40579 is shown with the obscured-AGN red-continuum solution.

Parameter	RUBIES-EGS-55604	RUBIES-UDS-40579
z	6.982	3.113558
M_{h}	$5.0 \times 10^{10} M_{\odot}$	$1.0 \times 10^{11} M_{\odot}$
$M_{\text{gas,d}}$	$10^8 M_{\odot}$	$3.0 \times 10^7 M_{\odot}$
$v_{\text{c,d}}$	175 km s^{-1}	100 km s^{-1}
R_{d}	14 pc	12.9 pc
$N_{\text{H,edge}}$	$2.5 \times 10^{25} \text{ cm}^{-2}$	$5.1 \times 10^{24} \text{ cm}^{-2}$
σ/m	$1.0 \text{ cm}^2 \text{ g}^{-1}$	$1.0 \text{ cm}^2 \text{ g}^{-1}$
C_{s}	84.9 km s^{-1}	77.1 km s^{-1}
$M_{\text{BH,SIDM}}$	$9.0 \times 10^7 M_{\odot}$	$6.1 \times 10^7 M_{\odot}$
t_{SIDM}	0.32 Myr	0.29 Myr
λ_{eff}	0.6	0.6
L_{bol}	$6.8 \times 10^{45} \text{ erg s}^{-1}$	$4.6 \times 10^{45} \text{ erg s}^{-1}$
α_{red}	4.3	3.0
λ_{turn}	—	$0.82 \mu\text{m}$
s_{turn}	—	4.0
T_{gas}	$1.4 \times 10^4 \text{ K}$	3000 K
β_{ph}	0	0
$f_{\text{red/AGN}}$	0.0245	0.055
f_{ph}	0.0455	0.051
C_{B}	0.70	0.95
τ_{B}	15	3.0
N_2/N_{H} required	3.0×10^{-6}	2.9×10^{-6}
$\Delta\lambda_{\text{B}}$	$0.006 \mu\text{m}$	$0.025 \mu\text{m}$
SFR_{eff}	$6.9 M_{\odot} \text{ yr}^{-1}$	$6.8 M_{\odot} \text{ yr}^{-1}$
$f_{\text{esc,young}}$	0.2	1.0
$E(B - V)_{\star}$	0	0.2
$M_{\star,\text{old}}$	$7.6 \times 10^8 M_{\odot}$	$3.3 \times 10^7 M_{\odot}$
AGN fraction, red continuum	0.72	0.58
Continuum rms	0.106 dex	0.070 dex

tends to select unusually early or concentrated halos, producing a strong environmental prediction [30]. It also leaves a redshift question: if halo age alone controls the visible event, lower-redshift halos have had more time to collapse, so the disappearance of LRDs toward low redshift is not automatic. In our picture SIDM still supplies the rapid dark mass growth, but the visible LRD phase is gated by high-redshift gas compaction and column density. Low-redshift halos may contain SIDM cores or black holes, but they usually lack the compact, gas-rich, Compton-thick nuclear disk needed to appear as an LRD.

Direct-collapse black-hole and quasi-star scenarios face a complementary problem [21–25]. They naturally produce buried accretion and dense gas, but the required initial conditions are severe: rapid inflow, suppressed fragmentation, special thermodynamic histories, and often a restricted halo population. If most LRDs were direct-collapse remnants, their abundance and clustering would trace the rarity of those conditions. Our model keeps the robust part of the direct-collapse simulations—the formation of a compact turbulent nuclear disk—but does not require the disk to form a black-hole seed already close to the observed LRD black-hole mass. It still requires a pre-SIDM seed massive enough for the compressed SIDM density to satisfy the ignition condition in Equation 14; in the fiducial model this is of order $10^5 M_{\odot}$, while denser nuclear disks can lower the requirement toward $10^4 M_{\odot}$ or even Pop III-remnant scales. This seed requirement is therefore comparatively mild and can plausibly be met in ordinary low-mass high-redshift galaxies or halos, rather than only in pristine direct-collapse sites. The large black-hole mass jump then occurs through SIDM accretion once the disk compresses the central dark matter, and the saturation mass is set by Equation 17. Thus the dense gas geometry is inherited from the simulations, while the need for a seed already comparable to the observed LRD black-hole mass is avoided.

Pure obscured-AGN or dusty-flow models solve the spectral problem with familiar physics but put too much work on gas accretion alone [2, 13, 14]. A particularly clear example is the super-Eddington unification model of Madau and Maiolino [15], in which LRDs are obscured little blue dots viewed through a dusty equatorial screen around a thick

accretion flow, while polar sight lines reveal the blue phase. This picture usefully captures orientation dependence and the connection between red and blue compact AGN. Its main burden is different from ours: the overmassive black hole must still be built by luminous gas accretion, usually in a super-Eddington state, and the model does not by itself explain why the phase is tied to a finite high-redshift gas-compaction event. More generally, in pure gas-accretion models the same gas must grow an overmassive black hole, power the observed continuum, remain X-ray weak, and avoid rapidly clearing itself by feedback. This often pushes the model toward high duty cycles, super-Eddington growth, or carefully tuned covering factors. Our model separates the tasks. SIDM accretion makes the black hole overmassive while radiating little after a finite ignition threshold is crossed; subsequent baryonic accretion can be sub-Eddington or near-Eddington; and the nuclear thick disk supplies the Compton-thick column that suppresses escaping X-rays. The X-ray weakness is therefore not an independent patch, but a consequence of the same gas structure that triggered the event.

Compact stellar, low-spin compact-galaxy, and forming-globular-cluster models emphasize real and important facts: LRDs are small, dense, and can contain stellar light [17–20]. Their weakness is that stellar compactness alone does not naturally explain broad permitted lines, dynamically or virially inferred overmassive black holes, weak X-rays, and a short high-redshift duty cycle. Low-spin models also risk becoming another special-halo selection, with corresponding clustering predictions [30]. In our fits the young and old stellar components are allowed, because the spectroscopic aperture includes the compact host, but they remain perturbative. Stars trace the inflow; they do not have to power the LRD.

Dense-gas or black-hole-star interpretations correctly identify an essential radiative-transfer clue: the Balmer-limit curvature and absorption can arise from dense gas rather than an old stellar population [8, 16]. We adopt this lesson, but add a dynamical trigger. In our model the Balmer-active layer is the atmosphere or wind base of the same Compton-thick nuclear disk that compressed the SIDM. The fitted Balmer optical depths are checked against the disk columns, and the required excited-state fractions are small. Dense gas is therefore not just a spectral fitting component; it is the physical mediator between baryonic inflow, SIDM compression, black-hole growth, and obscuration.

Finally, primordial-black-hole, dark-star, and other exotic channels can make early compact massive objects [26–28], but they still must explain why the observed objects have LRD spectra, why the phase is X-ray weak, why it fades at low redshift, and why its clustering is not obviously extreme. Our mechanism is less radical in its initial conditions and more restrictive in its observable trigger. It predicts that LRDs should correlate most strongly with local compaction, disturbed gas, compact high-column nuclei, and orientation, rather than only with rare halo labels. This is the central advantage of the model: it can produce overmassive black holes without demanding that every LRD occupy a globally exceptional halo, and it connects that mass growth to the same gas structure that makes the source red.

Observational Predictions

The model makes several falsifiable predictions. First, LRDs should be orientation dependent. High-column sight lines through the disk body, inner rim, or turbulent atmosphere appear red and X-ray weak, whereas polar sight lines should reveal bluer compact AGN with weaker Balmer curvature. Second, the strongest LRD signatures should correlate with compact gas, disturbed morphology, close companions, high attenuation, or other signs of recent inflow rather than only with large-scale halo overdensity. Third, the LRD phase should be short. Once the high-covering column declines, the polar channel widens, or the Balmer-active layer disappears, the descendant should be a less-obscured compact AGN, a blue compact broad-line AGN, or a compact post-starburst system with an overmassive black hole. Fourth, cooler LRDs such as RUBIES-UDS-40579 may represent more shielded or later radiative-transfer states of the same thick disk, while hotter V-shaped sources such as RUBIES-EGS-55604 may trace a more strongly photoionized or higher-column state.

This framework also explains why low-redshift analogues are uncommon. Later dwarf and small-galaxy mergers are less gas rich, more feedback regulated, and less able to maintain a compact Compton-thick nuclear disk. Moreover, once the first large fractional SIDM-fed black-hole growth event has occurred, later inflows change the black-hole mass by a smaller factor and are less able to produce the same potential-transformation phase. In more massive halos the velocity-dependent SIDM cross-section can be smaller, further suppressing efficient SIDM accretion. Thus the absence of abundant low-redshift LRDs follows from gas supply, orientation, duty cycle, and the one-time nature of the first large SIDM-fed mass jump.

Limitations and Next Steps

The present model is deliberately semi-analytic. The continuum parameters α_{red} , f_{red} , f_{ph} , T_{gas} , C_{B} , τ_{B} , and the Balmer-edge width obey energy and column-density constraints, but they are effective radiative-transfer parameters. A full radiation-hydrodynamic and photoionization calculation is required to predict them from first principles [55]. The stellar leakage and old-host terms are aperture-level nuisance components rather than central ingredients of the model. The SIDM compression calculation assumes an approximately isothermal SIDM response and parameterizes the disk potential with η_{Φ} . The decisive next step is a coupled simulation of a compact baryonic disk embedded in an SIDM halo with a growing central black hole, followed by radiative-transfer modeling of spectra, emission lines, and X-rays.

CONCLUSIONS

We have proposed a gas-triggered SIDM-accretion model for Little Red Dots. The central idea is that a compact nuclear thick disk can do two things at once: compress the central SIDM enough to ignite rapid dark accretion onto a pre-existing black hole, and provide the high-column gas that later reprocesses AGN radiation into an LRD spectrum. For a fiducial RUBIES-EGS-55604-like halo, the model gives $M_{\text{BH}} \simeq 9 \times 10^7 M_{\odot}$ in 0.32 Myr for $\sigma/m = 1 \text{ cm}^2 \text{ g}^{-1}$, while the subsequent gas-powered luminosity remains sub-Eddington.

The model naturally links several observed LRD properties. The black hole can become overmassive without requiring hyper-Eddington radiative growth. The compact disk provides Compton-thick columns, explaining weak escaping X-rays and strong reprocessing. A smooth Balmer-active disk atmosphere produces the Balmer-limit curvature. Orientation determines whether the same event appears as a red LRD or a bluer compact AGN. The LRD phase ends when the high-covering column is consumed, expelled, diluted, or no longer replenished, giving a natural lifetime of tens to hundreds of Myr.

Using the same thick-disk reprocessing framework, we fit two representative spectra: RUBIES-EGS-55604, a hotter V-shaped LRD, and RUBIES-UDS-40579, a cooler red-bump LRD. The required Balmer optical depths are consistent with the columns predicted by the disk model and imply excited-hydrogen fractions of order 10^{-6} . These fits are not a full radiative-transfer solution, but they demonstrate that the proposed physical picture can reproduce the observed continuum diversity with plausible parameters.

The most important observational tests are environmental and temporal. If our model is correct, LRDs should preferentially trace compact inflow events, close companions, disturbed morphologies, high covering columns, and orientation-dependent obscuration, rather than only the rarest high-concentration halos. Future JWST spectroscopy, X-ray constraints, MIRI photometry, resolved morphology, and clustering measurements can therefore distinguish gas-triggered SIDM accretion from direct collapse, gravothermal core collapse, compact stellar systems, and other proposed channels.

YR acknowledges support from the CAS Pioneer Hundred Talents Program (Category B), the NSFC grants 12522302, 12673017 and 12273037, and the USTC Research Funds of the Double First-Class Initiative.

* corresponding author: rongyua@ustc.edu.cn

- [1] J. Matthee, R. P. Naidu, G. Brammer, J. Chisholm, A.-C. Eilers, A. Goulding, *et al.*, [ApJ **963**, 129 \(2024\)](#).
- [2] J. E. Greene, I. Labbe, A. D. Goulding, L. J. Furtak, I. Chemerynska, V. Kokorev, *et al.*, [ApJ **964**, 39 \(2024\)](#).
- [3] D. D. Kocevski, S. L. Finkelstein, G. Barro, A. J. Taylor, A. Calabrò, B. Laloux, *et al.*, [ApJ **986**, 126 \(2025\)](#), [arXiv:2404.03576 \[astro-ph.GA\]](#).
- [4] H. B. Akins, C. M. Casey, E. Lambrides, N. Allen, I. T. Andika, M. Brinch, *et al.*, [ApJ **991**, 37 \(2025\)](#).
- [5] B. Wang, J. Leja, A. de Graaff, G. B. Brammer, A. Weibel, P. van Dokkum, *et al.*, [ApJL **969**, L13 \(2024\)](#), [arXiv:2405.01473 \[astro-ph.GA\]](#).
- [6] D. J. Setton, J. E. Greene, A. de Graaff, Y. Ma, J. Leja, J. Matthee, *et al.*, [ApJ **995**, 118 \(2025\)](#), [arXiv:2411.03424 \[astro-ph.GA\]](#).
- [7] R. E. Hviding, A. de Graaff, T. B. Miller, D. J. Setton, J. E. Greene, I. Labbé, *et al.*, [A&A **702**, A57 \(2025\)](#), [arXiv:2506.05459 \[astro-ph.GA\]](#).
- [8] A. de Graaff, H.-W. Rix, R. P. Naidu, I. Labbé, B. Wang, J. Leja, *et al.*, [A&A **701**, A168 \(2025\)](#), [arXiv:2503.16600 \[astro-ph.GA\]](#).
- [9] I. Labbe, J. E. Greene, J. Matthee, H. Treiber, V. Kokorev, T. B. Miller, *et al.*, [arXiv e-prints \(2024\)](#), [arXiv:2412.04557 \[astro-ph.GA\]](#).

- [10] L. J. Furtak, I. Labbé, A. Zitrin, J. E. Greene, P. Dayal, I. Chemerynska, *et al.*, *Nature* **628**, 57 (2024), [arXiv:2308.05735 \[astro-ph.GA\]](#).
- [11] I. Juodžbalis, C. Marconcini, F. D'Eugenio, R. Maiolino, A. Marconi, H. Übler, *et al.*, *Nature* **653**, 1017 (2026), [arXiv:2508.21748 \[astro-ph.GA\]](#).
- [12] P. G. Pérez-González, G. Barro, G. H. Rieke, J. Lyu, M. Rieke, S. Alberts, *et al.*, *ApJ* **968**, 4 (2024), [arXiv:2401.08782 \[astro-ph.GA\]](#).
- [13] F. Pacucci and R. Narayan, *ApJ* **976**, 96 (2024), [arXiv:2407.15915 \[astro-ph.GA\]](#).
- [14] Z. Li, K. Inayoshi, K. Chen, K. Ichikawa, and L. C. Ho, *ApJ* **980**, 36 (2025), [arXiv:2407.10760 \[astro-ph.GA\]](#).
- [15] P. Madau and R. Maiolino, *arXiv e-prints*, [arXiv:2602.22386](#) (2026), [arXiv:2602.22386 \[astro-ph.GA\]](#).
- [16] K. Inayoshi and R. Maiolino, *ApJL* **980**, L27 (2025), [arXiv:2409.07805 \[astro-ph.GA\]](#).
- [17] C. C. Williams, S. Alberts, Z. Ji, K. N. Hainline, J. Lyu, G. Rieke, *et al.*, *ApJ* **968**, 34 (2024), [arXiv:2311.07483 \[astro-ph.GA\]](#).
- [18] J. F. W. Baggen, P. van Dokkum, G. Brammer, A. de Graaff, M. Franx, J. Greene, *et al.*, *ApJL* **977**, L13 (2024), [arXiv:2408.07745 \[astro-ph.GA\]](#).
- [19] F. Pacucci and A. Loeb, *ApJL* **989**, L19 (2025), [arXiv:2506.03244 \[astro-ph.GA\]](#).
- [20] J. Chisholm, D. A. Berg, M. Boylan-Kolchin, A. de Graaff, L. J. Furtak, V. Kokorev, *et al.*, *ApJL* **1004**, L4 (2026), [arXiv:2602.15935 \[astro-ph.GA\]](#).
- [21] L. Mayer, D. Fiacconi, S. Bonoli, T. Quinn, R. Roškar, S. Shen, and J. Wadsley, *ApJ* **810**, 51 (2015), [arXiv:1411.5683 \[astro-ph.GA\]](#).
- [22] L. Mayer, P. R. Capelo, L. Zwick, and T. Di Matteo, *ApJ* **961**, 76 (2024), [arXiv:2304.02066 \[astro-ph.GA\]](#).
- [23] J. Jeon, B. Liu, V. Bromm, S. Fujimoto, A. J. Taylor, V. Kokorev, *et al.*, *ApJ* **998**, 148 (2026), [arXiv:2508.14155 \[astro-ph.GA\]](#).
- [24] M. C. Begelman and J. Dexter, *ApJ* **996**, 48 (2026), [arXiv:2507.09085 \[astro-ph.GA\]](#).
- [25] F. Pacucci, A. Ferrara, and D. D. Kocevski, *arXiv e-prints* (2026), [arXiv:2601.14368 \[astro-ph.GA\]](#).
- [26] B. Carr, F. Kühnel, and M. Sandstad, *Phys. Rev. D* **94**, 083504 (2016).
- [27] B. Zhang, W.-X. Feng, and H. An, *JCAP* **2026** (06), 097, [arXiv:2604.01684 \[astro-ph.GA\]](#).
- [28] C. Ilie, *arXiv e-prints* (2026), [arXiv:2606.02539 \[astro-ph.GA\]](#).
- [29] M. Carranza-Escudero, C. J. Conselice, N. Adams, T. Harvey, D. Austin, P. Behroozi, *et al.*, *ApJL* **989**, L50 (2025).
- [30] Z. Wang, F. Jiang, H. Zheng, X. Shen, Z. Jia, L. C. Ho, K. Inayoshi, and L. Jiang, *arXiv e-prints*, [arXiv:2603.15736](#) (2026), [arXiv:2603.15736 \[astro-ph.GA\]](#).
- [31] C. Zhang, H. Zhang, Q. Wu, L. C. Ho, and J.-M. Wang, *arXiv e-prints* (2026), [arXiv:2606.02773 \[astro-ph.GA\]](#).
- [32] D. N. Spergel and P. J. Steinhardt, *Phys. Rev. Lett.* **84**, 3760 (2000).
- [33] M. Kaplinghat, S. Tulin, and H.-B. Yu, *Phys. Rev. Lett.* **116**, 041302 (2016).
- [34] S. Tulin and H.-B. Yu, *Phys. Rep.* **730**, 1 (2018).
- [35] O. Sameie, P. Creasey, H.-B. Yu, L. V. Sales, M. Vogelsberger, and J. Zavala, *MNRAS* **479**, 359 (2018).
- [36] G. Despali, M. Sparre, S. Vegetti, M. Vogelsberger, J. Zavala, and F. Marinacci, *MNRAS* **484**, 4563 (2019).
- [37] Z. Zhang, Y. Chen, Y. Rong, H. Wang, H. Mo, X. Luo, and H. Li, *Nature* **642**, 47 (2025), [arXiv:2504.03305 \[astro-ph.GA\]](#).
- [38] F. Jiang, Z. Jia, H. Zheng, L. C. Ho, K. Inayoshi, X. Shen, M. Vogelsberger, and W.-X. Feng, *ApJL* **996**, L19 (2026), [arXiv:2503.23710 \[astro-ph.GA\]](#).
- [39] J. Hu, Y. Shen, Y.-Q. Lou, and S. Zhang, *MNRAS* **365**, 345 (2006).
- [40] N. Das and S. Kalita, *ApJ* **961**, 105 (2024).
- [41] V. M. Sabarish, M. Brüggen, K.-S. Hoberg, and M. S. Fischer, *A&A* **703**, A142 (2025).
- [42] A. de Graaff, R. E. Hviding, R. P. Naidu, J. E. Greene, T. B. Miller, J. Leja, *et al.*, *MNRAS* **551**, stag1567 (2026), [arXiv:2511.21820 \[astro-ph.GA\]](#).
- [43] P. Behroozi, R. H. Wechsler, A. P. Hearin, and C. Conroy, *MNRAS* **488**, 3143 (2019).
- [44] R. Barkana and A. Loeb, *Phys. Rep.* **349**, 125 (2001).
- [45] Planck Collaboration, N. Aghanim, Y. Akrami, *et al.*, *A&A* **641**, A6 (2020).
- [46] J. Binney and S. Tremaine, *Galactic Dynamics: Second Edition*, 2nd ed. (Princeton University Press, 2008).
- [47] D. E. Osterbrock and G. J. Ferland, *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei*, 2nd ed. (University Science Books, 2006).
- [48] G. R. Blumenthal, S. M. Faber, R. Flores, and J. R. Primack, *ApJ* **301**, 27 (1986).
- [49] O. Y. Gnedin, A. V. Kravtsov, A. A. Klypin, and D. Nagai, *ApJ* **616**, 16 (2004).
- [50] P. Madau and M. J. Rees, *ApJ* **551**, L27 (2001).
- [51] A. Heger, C. L. Fryer, S. E. Woosley, N. Langer, and D. H. Hartmann, *ApJ* **591**, 288 (2003).
- [52] V. Bromm and R. B. Larson, *ARA&A* **42**, 79 (2004).
- [53] E. E. Salpeter, *ApJ* **140**, 796 (1964).
- [54] M. Volonteri, *A&A Rev.* **18**, 279 (2010).
- [55] G. J. Ferland, M. Chatzikos, F. Guzmán, M. L. Lykins, P. A. M. van Hoof, R. J. R. Williams, *et al.*, *Revista Mexicana de Astronomía y Astrofísica* **53**, 385 (2017), [arXiv:1705.10877 \[astro-ph.GA\]](#).
- [56] C. Conroy, J. E. Gunn, and M. White, *ApJ* **699**, 486 (2009).
- [57] C. Conroy and J. E. Gunn, *ApJ* **712**, 833 (2010).
- [58] R. C. Kennicutt and N. J. Evans, *ARA&A* **50**, 531 (2012).
- [59] D. Calzetti, L. Armus, R. C. Bohlin, A. L. Kinney, J. Koornneef, and T. Storchi-Bergmann, *ApJ* **533**, 682 (2000).