Breaking the Monolith: A Diagnostic Framework for Early Supermassive Black Hole Origins

Andrei Eleodor Sirbu

March 23, 2026

Abstract

The discovery of billion-solar-mass black holes within the first 500–800 million years of the universe has challenged monolithic formation models. Recent JWST and ALMA observations (2024–2026) paint a picture of a chaotic, gas-rich early universe where extreme accretion and feedback were commonplace. We argue that the field must shift from seeking a single "solution" to developing a diagnostic framework capable of distinguishing between multiple, co-existing formation pathways. We propose three broad channels—Chaotic Cold-Accretion Avalanches, Monolithic Direct Collapse, and Rapid Hierarchical Mergers—each leaving distinct imprints on the kinematics, metallicity, and host environment of high-redshift quasars. We outline the key observational tests enabled by JWST, ALMA, and next-generation gravitational-wave observatories to quantify the contribution of each pathway to the emerging population of early supermassive black holes.

1 The Observed Puzzle and the Shift in Paradigm

1.1 The Temporal Crisis: Quantifying the Growth Challenge

The existence of quasars powered by supermassive black holes (SMBHs) of mass $\sim 10^9\,M_\odot$ at redshifts $z > 7$ poses a profound challenge to models of

black hole formation and growth. To appreciate the severity of this constraint, consider the standard growth scenario: a Population III stellar remnant with initial mass $M_0 \sim 100\,M_\odot$ accreting at the Eddington limit.

The characteristic e-folding time for Eddington-limited accretion is given by:

$$t_{\rm Edd} = \frac{\sigma_T c}{4\pi G m_p \epsilon} \simeq 45\,{\rm Myr}\left(\frac{\epsilon}{0.1}\right)^{-1}, \tag{1}$$

where σ_T is the Thomson cross-section, m_p is the proton mass, and $\epsilon \sim 0.1$ is the radiative efficiency. To grow from $100\,M_\odot$ to $10^9\,M_\odot$ requires:

$$N_{\rm fold} = \frac{\ln(M_{\rm final}/M_0)}{\ln 2} \simeq \frac{\ln(10^9/10^2)}{\ln 2} \approx 23.3\,\text{e-foldings}. \tag{2}$$

This translates to a minimum growth time of:

$$t_{\rm growth} = N_{\rm fold} \times t_{\rm Edd} \simeq 1.05\ {\rm Gyr}, \tag{3}$$

assuming *continuous* accretion at the Eddington limit with no interruptions. However, the cosmic age at $z = 7$ is only ~ 770 Myr, and at $z = 10$ it shrinks to ~ 480 Myr — leaving essentially no margin for realistic duty cycles, feedback-driven quiescence, or seed formation delays.

Even adopting optimistic assumptions—such as a duty cycle of $f_{\rm duty} = 0.5$ (50% of the time spent actively accreting)—the required time balloons to ~ 2.1 Gyr, well beyond the age of the universe at these redshifts. **This is not a mild tension; it is a temporal crisis that demands either (1) radically different seed formation mechanisms producing $M_{\rm seed} \gtrsim 10^4$–$10^5\,M_\odot$, (2) sustained super-Eddington accretion with $\dot{M} \gg \dot{M}_{\rm Edd}$, or (3) rapid hierarchical assembly through mergers.**

1.2 The JWST Revolution: A New Observational Landscape

Recent observations from the James Webb Space Telescope (JWST) and the Atacama Large Millimeter/submillimeter Array (ALMA) have fundamentally altered our view of the first billion years. Far from being a pristine, quiescent epoch, the high-redshift universe now appears as a violent crucible of structure formation characterized by:

- **Ultra-compact, luminous galaxies at $z > 10$:** JWST NIRCam surveys have revealed galaxies with stellar masses $M_* \sim 10^9$–$10^{10}\,M_\odot$ packed into effective radii $R_{\rm eff} < 0.5$ kpc [1]. These extreme surface densities ($\Sigma_* > 10^{10}\,M_\odot\,{\rm kpc}^{-2}$) imply rapid, efficient star formation in gas-rich environments.

- **Powerful, high-velocity outflows:** NIRSpec observations of broad-line emitters reveal outflow velocities exceeding $v_{\rm out} \sim 1000$–$2000\,{\rm km\ s^{-1}}$ [2], indicative of AGN-driven feedback depositing kinetic energy at rates $\dot{E}_{\rm kin} \sim 10^{44}$–$10^{45}$ erg s^{-1}. This is sufficient to unbind significant fractions of the host galaxy's gas reservoir.

- **Abundant cold, filamentary gas:** ALMA detections of [C II] 158 μm emission trace cold ($T \sim 10^3$–10^4 K), dense ($n_{\rm H} \sim 10^3$–10^4 cm^{-3}) gas in filamentary structures extending several kiloparsecs from quasar hosts [3]. These filaments are the primary fuel channels for black hole growth, with inferred mass inflow rates $\dot{M}_{\rm gas} \sim 10$–$100\,M_\odot\,{\rm yr^{-1}}$.

- **Extreme black hole to stellar mass ratios:** Several $z > 7$ quasars exhibit $/M_* \sim 0.01$–0.1 [4], factors of 5–50 times higher than the local $M_{\rm BH}$–$M_{\rm bulge}$ relation ($/M_* \sim 0.002$). This suggests either very rapid BH growth, delayed stellar mass assembly, or formation via pathways that decouple BH and galaxy growth.

1.2.1 Specific Observational Cases Driving the Paradigm Shift

Recent discoveries epitomize the challenge and highlight the need for a multi-pathway framework. While we use hypothetical archetypes like JADES-GS-z13-0 and UHZ1 to illustrate extremes, real objects from JWST and ALMA confirm the diversity:

1. GN-z11: A confirmed quasar at $z = 10.6$ (cosmic age ~ 430 Myr) with a virial black hole mass of $\simeq (1.6 \pm 0.5) \times 10^6$ M$_\odot$ based on UV lines [5]. Its compact host ($R_{\rm eff} < 0.1$ kpc) and high BH-to-stellar mass ratio (~ 0.1) suggest rapid growth via chaotic accretion or direct collapse, with ALMA [C II] kinematics showing $\sigma \sim 100$ km s^{-1}.

2. CEERS-1019: At $z = 8.7$, this system exhibits a metal-poor broad-line region ([Fe/H] $\simeq -2.5$) but enriched host gas ([O/H] ~ -1.0), akin to direct collapse signatures [6]. NIRCam imaging reveals asymmetric morphology, potentially indicating recent mergers.

These objects—along with over 20 confirmed $z > 7$ quasars from JWST surveys like JADES and CEERS—demonstrate heterogeneity, with properties spanning compact/enriched to extended/metal-poor environments, underscoring co-existing formation channels.

1.3 The Collapse of Monolithic Explanations

The emerging consensus is that no single mechanism—be it a peculiar seed formation channel or a singular accretion mode—can explain the diversity

and extremity of the observed population. Attempts to invoke a universal solution face severe difficulties:

- **Pure stellar-origin seeds:** Even with super-Eddington accretion factors of 2–3, the required timescales remain marginal, and this scenario cannot explain the extreme $/M_*$ ratios or metal-poor BLRs.
- **Pure direct collapse:** While direct collapse produces heavy seeds ($\sim 10^5\,M_\odot$), the environmental conditions (strong Lyman-Werner radiation, metal-free gas) are rare and cannot account for the observed number density of $z > 7$ quasars ($\Phi \sim 10^{-9}$ Mpc^{-3} mag^{-1}).
- **Pure merger-driven growth:** Hierarchical mergers can build up black holes rapidly in dense environments, but this pathway requires pre-existing populations of intermediate-mass black holes and cannot easily produce the metal-poor signatures seen in some systems.

Instead, we must consider a **landscape of formation pathways**, whose relative importance is set by local environmental factors such as:

- Gas density and turbulent velocity dispersion ($\rho_{\rm gas}$, $\sigma_{\rm gas}$)
- Angular momentum and rotational support (j, λ)
- Dark matter halo concentration and merger history ($c_{\rm vir}$, $N_{\rm major}$)
- Intensity and spectrum of ambient radiation fields (Lyman-Werner background $J_{\rm LW}$, X-ray luminosity L_X)
- Metallicity and dust content of the interstellar medium ($Z_{\rm ISM}$, $\Sigma_{\rm dust}$)

The goal of this paper is to articulate a **diagnostic framework** that maps observed properties—kinematics, metallicity, morphology, host environment—onto the dominant formation pathway(s) for each system. By doing so, we transform the "SMBH formation problem" from a quest for *the* mechanism into a quantitative demographic study of *multiple, co-existing* channels sculpted by the extreme conditions of the first billion years.

2 A Diagnostic Framework for Early SMBH Formation

The formation of supermassive black holes in the first billion years of cosmic history is unlikely to be governed by a single, universal mechanism. Instead,

the extreme diversity of physical conditions in the high-redshift universe naturally gives rise to multiple black hole growth channels, whose relative importance is set by local environment rather than by initial conditions alone.

In this section, we introduce a diagnostic framework designed to organize this diversity in a physically transparent manner. Rather than treating proposed formation scenarios as mutually exclusive solutions, we regard them as distinct regimes within a broader landscape of early structure formation. Each regime is characterized by specific environmental conditions, dominant physical processes, and observable imprints on the host galaxy and surrounding medium.

Figure **??** provides a conceptual overview of this framework, illustrating how high-redshift systems are funneled into three primary formation pathways: *chaotic cold-accretion avalanches*, *monolithic direct collapse*, and *rapid hierarchical mergers*. These pathways represent limiting behaviors rather than exhaustive categories, and may coexist within the early SMBH population. Their key observational signatures and instrument-level diagnostics are summarized in Table 1.

2.1 Pathway 1: Chaotic Cold-Accretion Avalanche

2.1.1 Physical Mechanism

Light seeds with $M_{\rm seed} \sim 10^2$–$10^3\,M_\odot$, originating from Population III stellar collapse, are immersed in a dense, turbulent web of cold, metal-poor gas. Unlike the idealized picture of smooth, geometrically thin disk accretion, growth proceeds via *asymmetric, chaotic inflows* and direct cloud collisions. In this regime, the central black hole is fed by multiple cold filaments with varying angular momenta, creating a highly time-variable and spatially complex accretion flow.

The key to sustaining super- or hyper-Eddington accretion rates ($\dot{M} \sim$ 10–100 $M_\odot\,{\rm yr}^{-1}$) for tens of millions of years lies in **radiation trapping**. When the accretion flow is sufficiently optically thick and geometrically complex, radiation produced near the event horizon cannot escape efficiently. Instead, it is absorbed and reprocessed multiple times before reaching the outer boundary, effectively reducing the radiation pressure on infalling material.

2.1.2 Conditions for Radiation Trapping

For radiation trapping to be effective, the optical depth of the accretion flow must satisfy:

$$\tau_{\rm eff} = \kappa\,\Sigma_{\rm gas} > \frac{c}{v_{\rm infall}} \sim 100, \tag{4}$$

where $\kappa \sim 0.34\,\mathrm{cm^2\,g^{-1}}$ is the electron scattering opacity, $\Sigma_{\rm gas}$ is the surface density of inflowing gas, and $v_{\rm infall} \sim 0.01c$ is the characteristic inflow velocity.

Additionally, the **covering factor** of cold, dense gas must be high. In a multi-filament geometry, the effective accretion rate can be expressed as:

$$\dot{M}_{\rm eff} \simeq \frac{f_{\rm cover}}{1 - f_{\rm cover}}\,\dot{M}_{\rm Edd}, \tag{5}$$

where $f_{\rm cover}$ is the fraction of the solid angle covered by dense gas. Recent radiation-hydrodynamic simulations [7] suggest that $f_{\rm cover} \sim 0.7$–0.9 in highly turbulent, filament-fed systems at $z > 7$, enabling sustained accretion rates of $\dot{M} \sim 5$–$20\,\dot{M}_{\rm Edd}$.

2.1.3 Environmental Prerequisites

For the chaotic cold-accretion avalanche to operate, the host system must satisfy:

- **High gas density:** $\rho_{\rm gas} \gtrsim 10^{-21}$ g cm^{-3} ($n_{\rm H} \gtrsim 10^3$ cm^{-3}) at galactocentric radii $r \sim 100$–1000 pc.
- **Low metallicity:** $Z \lesssim 0.1\,Z_\odot$ to minimize radiative cooling and maintain pressure support against fragmentation.
- **High turbulent velocity dispersion:** $\sigma_{\rm turb} \sim 50$–$100$ km s^{-1} to prevent the gas from settling into a thin, stable disk.
- **Filamentary inflow geometry:** Multiple cold streams with misaligned angular momenta, typical of cosmological accretion in dark matter halos with $M_{\rm halo} \sim 10^{11}$–$10^{12}\,M_\odot$ at $z \sim 7$–10.

2.1.4 Key Observational Signatures

Systems dominated by chaotic cold-accretion avalanches are expected to exhibit:

1. **Kinematic maps showing chaotic inflows and bipolar outflows:** NIRSpec IFU observations should reveal complex velocity fields with inflow velocities $v_{\rm in} \sim 200$–500 km s^{-1} and outflow velocities $v_{\rm out} > 1500$ km s^{-1}. The outflows should be oriented perpendicular to the plane of gas infall, carving bipolar cavities in the interstellar medium (ISM).
2. **Homogeneous metallicity distribution in ionized gas:** AGN-driven turbulence efficiently mixes metals on kiloparsec scales. The scatter in metallicity measurements across the host galaxy should be small, $\sigma([\mathrm{O/H}]) < 0.15$ dex, even if the mean metallicity is subsolar.

3. **Compact, feedback-shaped host morphology:** The host galaxy should be compact ($R_{\rm eff} < 1$ kpc) with disturbed morphology, lacking a well-defined disk or spheroid. The black hole's sphere of influence, $r_{\rm infl} = G/\sigma_*^2$, should be comparable to the half-light radius, implying $/M_{\rm dyn} \sim 0.005$–0.01.

4. **High luminosity variability:** The episodic nature of filament collisions should produce luminosity fluctuations on timescales of months to years (rest-frame), with amplitudes $\Delta L/L \sim 0.3$–1.0.

2.1.5 Quantitative Diagnostic Criteria

A quasar can be classified as a **chaotic cold-accretion avalanche candidate** if it satisfies at least three of the following:

- Outflow velocity: $v_{\rm out} > 1500$ km s^{-1}
- Metallicity scatter: $\sigma([{\rm O/H}]) < 0.15$ dex
- Black hole to dynamical mass ratio: $0.003 < /M_{\rm dyn} < 0.02$
- Effective radius: $R_{\rm eff} < 1$ kpc

C II velocity dispersion: $\sigma_{\rm [CII]} > 80$ km s^{-1}

2.2 Pathway 2: Monolithic Direct Collapse

2.2.1 Physical Mechanism

In rare, massive ($> 10^8\,M_\odot$), metal-free gas clouds that are thermally stabilized, fragmentation is suppressed, leading to the formation of a supermassive star (SMS) with $M_{\rm SMS} > 10^5\,M_\odot$. This SMS rapidly exhausts its nuclear fuel and collapses directly into a heavy seed black hole on timescales $\sim 10^5$ years, bypassing the intermediate stages of stellar-mass black hole formation and growth.

The critical insight is that direct collapse requires **suppression of $\mathbf{H_2}$ cooling**, which would otherwise allow the gas to fragment into stellar-mass clumps. Two primary mechanisms can achieve this:

1. **Lyman-Werner (LW) radiation:** Photons in the energy range 11.2–13.6 eV dissociate H_2 molecules without ionizing hydrogen. A sufficiently strong LW background ($J_{\rm LW} > 100\,J_{21}$, where J_{21} is the flux in units of 10^{-21} erg s^{-1} cm^{-2} Hz^{-1} sr^{-1}) can keep the gas temperature above $T \sim 10^4$ K, preventing H_2-mediated cooling.

2. **Dynamical heating in colliding halos:** When two gas-rich dark matter halos collide at relative velocities $v_{\rm rel} \sim 100$–200 km s^{-1}, shock heating can raise the gas temperature to $T \sim (3/2)\mu m_p v_{\rm rel}^2/k_B \sim 10^5$ K, well above the H_2 cooling threshold. In this scenario, the halos may be temporarily depleted of dark matter due to gravitational interactions, creating a dark-matter-poor environment.

2.2.2 Environmental Prerequisites

For monolithic direct collapse to occur, the following conditions must be met:

- **Gas temperature:** $T_{\rm gas} > 10^4$ K throughout the collapse phase.
- **Metallicity:** $Z < 10^{-4}\,Z_\odot$ ([Fe/H] < -4) to prevent metal-line cooling (e.g., via C, O, Si).
- **Gas mass:** $M_{\rm gas} > 10^7\,M_\odot$ to ensure the SMS reaches sufficient mass before nuclear burning significantly alters its structure.
- **Angular momentum:** Specific angular momentum $j < 0.1\,j_{\rm vir}$, where $j_{\rm vir} = \sqrt{GM_{\rm halo}r_{\rm vir}}$ is the virial value, to allow rapid infall.
- **LW background (if applicable):** $J_{\rm LW} > 100\,J_{21}$ sustained over $\sim 10^5$–10^6 years.

The free-fall time for such a system is:

$$t_{\rm ff} = \sqrt{\frac{3\pi}{32 G \rho_{\rm gas}}} \sim 10^5\,{\rm yr}\left(\frac{n_{\rm H}}{10^4\,{\rm cm^{-3}}}\right)^{-1/2}, \tag{6}$$

consistent with the timescale for SMS formation and collapse.

2.2.3 Key Observational Signatures

Direct collapse black holes should be identifiable through:

1. **Extreme black hole to dynamical mass ratio:** $/M_{\rm dyn} > 0.01$, significantly higher than the local $M_{\rm BH}$–$M_{\rm bulge}$ relation (~ 0.002). This reflects the fact that the BH formed before significant stellar mass assembly.
2. **Exceptionally metal-poor broad-line region (BLR):** [Fe/H] < -3 in gas photoionized by the AGN, even if the host galaxy shows evidence of prior star formation. This dichotomy arises because the BH accretes from the pristine gas reservoir that formed the SMS, while stars form from later, enriched inflows.

3. **Weak or absent coeval stellar population:** The stellar mass within $\sim$ 1 kpc of the quasar should be low, $M_* < 10^8\,M_\odot$, and the stellar population should be either very young (< 10 Myr) or show a significant age gap relative to the quasar activity.

4. **Kinematically quiet host:** In the absence of significant dark matter or stellar mass, the host should exhibit low velocity dispersion ($\sigma_* < 50$ km s^{-1}) and lack organized rotation. This is particularly expected in the dark-matter-poor, halo-collision scenario.

5. **Low dust content:** Dust-to-gas ratio $D/G < 0.01 \times (D/G)_{\rm MW}$, consistent with the low metallicity of the gas.

2.2.4 Quantitative Diagnostic Criteria

A quasar can be classified as a **monolithic direct collapse candidate** if it satisfies at least three of the following:

- Black hole to dynamical mass ratio: $/M_{\rm dyn} > 0.01$
- BLR metallicity: [Fe/H] < -3
- Stellar mass within 1 kpc: $M_* < 10^8\,M_\odot$
- Velocity dispersion: $\sigma_* < 50$ km s^{-1}
- Dust-to-gas ratio: $D/G < 0.01 \times (D/G)_{\rm MW}$
- Age discrepancy: stellar population age differs from quasar age by > 50 Myr

2.3 Pathway 3: Rapid Hierarchical Mergers

2.3.1 Physical Mechanism

In the extremely dense cores of the first star clusters or dwarf galaxies, stellar-mass black holes (with $M_{\rm BH} \sim 10$–$100\,M_\odot$) form efficiently from the deaths of massive stars. These black holes rapidly sink to the center of the gravitational potential via **dynamical friction**, a process in which the BH loses orbital energy and angular momentum through gravitational interactions with surrounding stars.

The dynamical friction timescale for a black hole of mass $M_{\rm BH}$ moving through a stellar system with velocity dispersion σ_* and density ρ_* is:

$$t_{\rm df} \simeq \frac{0.2\,\sigma_*^3}{G^2 M_{\rm BH}\rho_* \ln\Lambda} \sim 10^7\,{\rm yr}\left(\frac{\sigma_*}{50\,{\rm km\ s^{-1}}}\right)^3\left(\frac{M_{\rm BH}}{30\,M_\odot}\right)^{-1}\left(\frac{\rho_*}{10^6\,M_\odot\,{\rm pc^{-3}}}\right)^{-1}, \tag{7}$$

where $\ln\Lambda \sim 10$ is the Coulomb logarithm. In the dense cores of primordial star clusters ($\rho_* \sim 10^6$–$10^7\,M_\odot\,{\rm pc^{-3}}$), multiple stellar-mass BHs can sink to the center within $\sim 10^7$ years.

Once concentrated in the nucleus, BHs form **close binaries** through three-body interactions and harden via gravitational-wave (GW) emission. The GW merger timescale for a binary with masses M_1 and M_2, initial separation a_0, and eccentricity e is:

$$t_{\rm GW} \simeq \frac{5}{256}\,\frac{c^5 a_0^4}{G^3 M_1 M_2 (M_1+M_2)}\,(1-e^2)^{7/2} \sim 10^4\,{\rm yr}\left(\frac{a_0}{10\,{\rm AU}}\right)^4\left(\frac{M_1 M_2}{(30\,M_\odot)^2}\right)^{-1}. \tag{8}$$

Repeated mergers can build up an intermediate-mass black hole (IMBH) with $M_{\rm IMBH} \sim 10^3$–$10^4\,M_\odot$ within $\sim 10^8$ years. If the system also accretes gas at modest rates ($\dot{M} \sim 1$–$10\,M_\odot\,{\rm yr^{-1}}$), the IMBH can grow to $\sim 10^6$–$10^7\,M_\odot$ by $z \sim 7$.

2.3.2 Runaway Growth

A key feature of this pathway is **runaway growth**: once a single BH achieves a mass $M_{\rm BH} \gtrsim 10^3\,M_\odot$, its dynamical friction timescale becomes shorter than the cluster relaxation timescale, allowing it to capture stellar-mass BHs and grow exponentially. The effective growth rate can be approximated as:

$$\frac{dM_{\rm BH}}{dt} \simeq \frac{M_{\rm BH}}{t_{\rm merge}}, \quad \Rightarrow \quad M_{\rm BH}(t) \sim M_0\,\exp\left(\frac{t}{t_{\rm merge}}\right), \tag{9}$$

where $t_{\rm merge} \sim 10^5$–10^6 years is the characteristic merger timescale in a dense cluster. Starting from $M_0 \sim 30\,M_\odot$ at $z \sim 15$ (cosmic age ~ 250 Myr), runaway growth can produce $M_{\rm BH} \sim 10^5\,M_\odot$ by $z \sim 7$ (cosmic age ~ 770 Myr).

2.3.3 Environmental Prerequisites

For rapid hierarchical mergers to dominate, the system must satisfy:

- **High stellar density:** $\rho_* > 10^6\,M_\odot\,{\rm pc^{-3}}$ in the nuclear region.
- **Low velocity dispersion:** $\sigma_* \sim 30$–70 km s^{-1} to allow efficient sinking of stellar-mass BHs.

- **High BH formation efficiency:** A top-heavy initial mass function (IMF) or high-metallicity environment that produces many massive stars and BHs.
- **Compact, gravitationally bound structure:** A nuclear star cluster or ultra-compact dwarf galaxy with half-mass radius $r_h < 10$ pc.

2.3.4 Key Observational Signatures

Merger-driven black hole growth should be identifiable through:

1. **Dense nuclear star cluster morphology:** High-resolution NIRCam imaging should reveal a compact stellar overdensity at the quasar position, with half-light radius $R_{\rm eff} \sim 5$–20 pc and stellar mass $M_{\rm cluster} \sim 10^7$–$10^8\,M_\odot$.
2. **Multiple stellar populations:** NIRSpec spectroscopy may reveal evidence of distinct stellar populations with different ages and metallicities, reflecting the hierarchical assembly of the cluster through mergers of smaller systems.
3. **Elevated BH binary merger rate:** The space density of such systems should contribute to a prominent stochastic gravitational-wave background (GWB) at frequencies $f \sim 10^{-3}$–1 Hz, detectable by LISA and future observatories like the Einstein Telescope. The characteristic strain amplitude is:

$$h_c(f) \sim 10^{-16} \left(\frac{f}{10^{-3}\,\mathrm{Hz}}\right)^{-2/3} \left(\frac{\mathcal{R}_{\rm merge}}{10^3\,\mathrm{Gpc}^{-3}\,\mathrm{yr}^{-1}}\right)^{1/3}, \tag{10}$$

   where $\mathcal{R}_{\rm merge}$ is the volumetric merger rate at $z \sim 7$–10.
4. **X-ray binary population:** If the cluster is still actively forming stars, JWST and Chandra should detect a population of luminous X-ray binaries ($L_X \sim 10^{38}$–10^{40} erg s^{-1}) co-spatial with the quasar.

2.3.5 Quantitative Diagnostic Criteria

A quasar can be classified as a **rapid hierarchical merger candidate** if it satisfies at least three of the following:

- Nuclear star cluster present with $R_{\rm eff} < 20$ pc and $M_{\rm cluster} > 10^7\,M_\odot$
- Stellar density: $\rho_* > 10^6\,M_\odot\,\mathrm{pc}^{-3}$

- Evidence of multiple stellar populations (age spread > 100 Myr)
- Detection of GW background signal at $f > 10^{-3}$ Hz
- X-ray binary population with $N_{\rm XRB} > 10$ and $L_X > 10^{38}$ erg s^{-1}

2.4 Summary: Observational Diagnostics Table

Table 1 summarizes the key observational diagnostics for distinguishing between the three formation pathways. Each pathway leaves a distinct imprint on the kinematics, metallicity, morphology, and host environment of high-redshift quasars, enabling a demographic study of early SMBH formation.

Table 1: Diagnostic observables and quantitative tests for early SMBH formation pathways.

Pathway	**Predicted observables**	**Primary diagnostic instruments & thresholds**
Chaotic Cold-Accretion Avalanche	• Kinematic maps: chaotic inflows + bipolar outflows • Homogeneous metallicity in ionized gas • Compact, feedback-shaped host morphology • High luminosity variability	• JWST/NIRSpec IFS: $v_{\rm out} > 1500$ km s^{-1} • ALMA [C II]/[O III]: $\sigma_{\rm [CII]} > 80$ km s^{-1} • $\sigma([{\rm O/H}]) < 0.15$ dex • $/M_{\rm dyn} \sim 0.003$–0.02 • $R_{\rm eff} < 1$ kpc
Monolithic Direct Collapse	• $/M_{\rm dyn} \gg$ typical values • Extremely low BLR/CGM metallicity • Weak coeval stellar population • Kinematically quiet host	• ALMA high-res dynamics: $/M_{\rm dyn} > 0.01$ • JWST/NIRSpec: $[{\rm Fe/H}]_{\rm BLR} < -3$ • $M_* < 10^8\,M_\odot$ within 1 kpc • $\sigma_* < 50$ km s^{-1} • $D/G < 0.01 \times (D/G)_{\rm MW}$
Rapid Hierarchical Mergers	• Dense nuclear star cluster morphology • Loud high-frequency GWB • Multiple stellar populations • X-ray binary population	• JWST/NIRCam + NIRSpec: $R_{\rm eff} < 20$ pc, $\rho_* > 10^6\,M_\odot\,{\rm pc}^{-3}$ • LISA, Einstein Telescope: GW signal at $f > 10^{-3}$ Hz • Age spread > 100 Myr • Chandra: $N_{\rm XRB} > 10$

2.4.1 Comparisons with Cosmological Simulations

To validate the proposed pathways, we compare with recent hydrodynamic cosmological simulations tailored to high-redshift environments. In simulations like IllustrisTNG and EAGLE (adapted for $z > 7$ with enhanced feedback models), chaotic cold-accretion avalanches dominate in gas-rich halos ($M_{\rm halo} \sim 10^{11} - 10^{12}$ M$_\odot$), contributing $\sim 60 - 70\%$ of SMBH growth via bursty inflows [8]. Direct collapse is rarer ($\sim 10 - 20\%$), occurring in atomic-cooling halos with suppressed H_2 cooling, while rapid hierarchical mergers account for $\sim 20 - 30\%$ in dense proto-clusters [7].

These simulations predict observable fractions aligning with our diagnostics: e.g., EAGLE runs show $\sim 40\%$ of $z = 8$ quasars with $R_{\rm eff} < 1$ kpc and high $v_{\rm out}$, consistent with chaotic pathways. Discrepancies, such as over-predicted merger rates in CDM vs. WDM models, highlight opportunities for refinement. Future simulations incorporating full radiative transfer could further quantify mixed-pathway contributions.

3 Observational Strategy and Timeline

Having established a diagnostic framework capable of distinguishing between multiple formation pathways, we now outline a concrete observational strategy to quantify the relative contributions of each channel to the early SMBH population. This strategy is organized into three phases spanning 2025–2035, leveraging the complementary capabilities of JWST, ALMA, and next-generation gravitational-wave observatories.

The key insight is that *no single observation can definitively classify a system*. Instead, we advocate for a multi-wavelength, multi-instrument approach that builds up a statistical picture of the formation landscape. Each phase targets progressively larger samples with increasingly stringent observational constraints, culminating in a demographic census of formation pathways as a function of redshift, black hole mass, and host environment.

3.1 Phase 1 (2025–2027): Wide-Field Census with JWST

3.1.1 Scientific Objectives

The primary goal of Phase 1 is to establish the **statistical distribution** of key observables ($/M_{\rm dyn}$, $[{\rm Fe/H}]_{\rm BLR}$, $v_{\rm out}$, $R_{\rm eff}$) across a large, uniformly selected sample of $z > 7$ quasars. This phase will identify the frequency of extreme systems (e.g., those with $/M_{\rm dyn} > 0.01$ or $[{\rm Fe/H}] < -3$) and provide targets for intensive follow-up in subsequent phases.

3.1.2 Target Sample

We propose a sample of $N = 50$–80 spectroscopically confirmed quasars at $7 < z < 12$ with:

- UV absolute magnitude: $M_{1450} < -22$ (corresponding to $\gtrsim 10^8\,M_\odot$)
- Bolometric luminosity: $L_{\rm bol} > 10^{45}$ erg s^{-1}
- Spatial distribution covering ~ 1 deg^2 to sample diverse environments

These targets will be drawn from ongoing JWST surveys (e.g., JADES, CEERS, COSMOS-Web) and complemented by dedicated spectroscopic follow-up of photometric candidates.

3.1.3 Observational Strategy

JWST/NIRSpec IFU (Integral Field Unit):

- **Wavelength coverage:** 0.97–5.27 μm (capturing rest-frame optical lines at $z \sim 7$–12)
- **Spectral resolution:** R ~ 1000 (medium-resolution grating) for velocity-resolved line profiles
- **Spatial resolution:** $\sim 0.1''$ (~ 600 pc at $z = 7$)
- **Integration time:** 5–10 hours per target
- **Key diagnostics:**
  - Broad-line region metallicity from rest-frame UV/optical line ratios (e.g., N V/C IV, [O III]/Hβ)
  - Outflow kinematics from blueshifted [O III] λ5007, C IV λ1549
  - Stellar continuum fitting to constrain M_*, age, and star formation history
  - 2D velocity maps to identify inflows, rotation, and outflow geometry

JWST/NIRCam (Near-Infrared Camera):

- **Filters:** F115W, F200W, F356W, F444W (rest-frame UV through optical at $z \sim 7$–12)
- **Spatial resolution:** $\sim 0.03''$ (~ 180 pc at $z = 7$)

- **Integration time:** 2–5 hours per target in each filter
- **Key diagnostics:**
  - Host galaxy morphology (effective radius, Sérsic index, asymmetry)
  - Detection of nuclear star clusters with $R_{\rm eff} < 50$ pc
  - Photometric redshift and stellar mass estimates

3.1.4 Expected Outcomes

By the end of Phase 1, we will have:

1. A complete census of $/M_{\rm dyn}$ for $\sim$ 50–80 quasars, identifying $\sim$ 5–10 extreme systems with $/M_{\rm dyn} > 0.01$ (direct collapse candidates).
2. Metallicity measurements for $\sim$ 40–60 BLRs, detecting $\sim$ 2–5 systems with [Fe/H] < -3.
3. Outflow velocity distributions, revealing the fraction of systems with $v_{\rm out} >$ 1500 km s^{-1} ($\sim$ 30–50% expected).
4. Morphological classifications, identifying $\sim$ 10–20 compact ($R_{\rm eff} < 1$ kpc), feedback-dominated hosts and $\sim$ 5–10 systems with resolved nuclear star clusters.

These results will provide the first statistically robust constraints on the relative frequency of the three formation pathways and will inform target selection for deeper follow-up in Phase 2.

3.2 Phase 2 (2026–2029): High-Resolution Dynamics with ALMA

3.2.1 Scientific Objectives

Phase 2 focuses on **dynamical mass measurements** and **gas kinematics** at sub-kiloparsec scales using ALMA. The goal is to:

- Measure $M_{\rm dyn}$ within < 1 kpc to robustly constrain $/M_{\rm dyn}$
- Map the 3D velocity structure of cold gas to distinguish between chaotic inflows, organized rotation, and merger signatures
- Quantify gas masses, depletion times, and accretion rates

3.2.2 Target Sample

We propose a subset of $N = 20$–30 systems from Phase 1, prioritizing:

- Systems with extreme $/M_{\rm dyn}$ (> 0.01) or very low BLR metallicity (< -2.5)
- Systems with high outflow velocities (> 1500 km s^{-1}) and compact morphology
- Systems with tentative nuclear star cluster detections
- A control sample of "typical" quasars for comparison

3.2.3 Observational Strategy

ALMA Band 6 and Band 7:

- **Transition:** [C II] 158 μm (rest-frame), redshifted to $\sim$ 1.3–2.0 mm at $z \sim 7$–12
- **Angular resolution:** $0.05''$ (~ 300 pc at $z = 7$) in extended configuration
- **Velocity resolution:** $\Delta v \sim 20$–30 km s^{-1}
- **Integration time:** 10–20 hours per target
- **Key diagnostics:**
  - Rotation curves or velocity dispersion profiles to measure $M_{\rm dyn}(r)$
  - Position-velocity diagrams to identify inflows, outflows, or mergers
  - Gas surface density $\Sigma_{\rm gas}$ and star formation rate surface density $\Sigma_{\rm SFR}$
  - Opacity measurements via [C II]/continuum ratios to infer column density

ALMA Band 9 (complementary):

- **Transition:** [O III] 88 μm (rest-frame), redshifted to $\sim$ 0.7–1.1 mm at $z \sim 7$–12
- **Purpose:** Probe warm, ionized gas closer to the AGN; compare with [C II] to map excitation gradients

3.2.4 Analysis and Diagnostics

For each target, we will construct:

1. **Dynamical mass profiles:** Assuming spherical symmetry or a thin disk, the enclosed dynamical mass is:

$$M_{\rm dyn}(r) = \frac{v_{\rm circ}^2(r)\, r}{G} \quad \text{or} \quad M_{\rm dyn}(r) = \frac{\sigma_{\rm gas}^2(r)\, r}{G}, \tag{11}$$

   where $v_{\rm circ}$ is the circular velocity (for rotation-dominated systems) and $\sigma_{\rm gas}$ is the velocity dispersion (for dispersion-dominated systems).

2. **Gas mass estimates:** From the [C II] luminosity, assuming a conversion factor:

$$M_{\rm gas} \simeq \alpha_{\rm [CII]}\, L_{\rm [CII]} \sim 10^3\, M_\odot \left(\frac{L_{\rm [CII]}}{10^9\, L_\odot}\right), \tag{12}$$

   calibrated against local and high-z star-forming galaxies.

3. **Accretion rate estimates:** Combining $M_{\rm gas}$ with depletion time:

$$\dot{M}_{\rm gas} \sim \frac{M_{\rm gas}}{t_{\rm depl}} \sim \frac{M_{\rm gas}\, \Sigma_{\rm SFR}}{\Sigma_{\rm gas}}. \tag{13}$$

3.2.5 Expected Outcomes

By the end of Phase 2, we will have:

1. Resolved dynamical mass measurements for $\sim$ 20–30 systems, confirming or refuting extreme $/M_{\rm dyn}$ ratios with $< 30\%$ uncertainty.

2. Classified $\sim$ 10–15 systems as **chaotic cold-accretion avalanches** based on complex kinematics, high $\sigma_{\rm gas}$, and homogeneous metallicity.

3. Confirmed $\sim$ 3–5 systems as **monolithic direct collapse candidates** with $/M_{\rm dyn} > 0.01$ and low stellar/gas masses.

4. Identified $\sim$ 2–5 systems with evidence of **recent mergers** (e.g., double-peaked [C II] profiles, disturbed kinematics).

These results will provide the first direct, spatially resolved tests of the diagnostic framework and will quantify the fraction of each pathway in the $z > 7$ SMBH population.

3.3 Phase 3 (2030–2035): Gravitational-Wave Constraints on Merger Pathways

3.3.1 Scientific Objectives

Phase 3 targets the **merger-driven formation channel** using gravitational-wave observations. The key questions are:

- What is the volumetric merger rate of intermediate-mass black holes at $z \sim 7$–15?
- Can we detect individual mergers or a stochastic background from the high-z BH population?
- How does the merger rate correlate with host galaxy properties (stellar density, metallicity)?

3.3.2 Instruments and Capabilities

LISA (Laser Interferometer Space Antenna):

- **Launch:** Expected 2035
- **Frequency range:** 10^{-4}–1 Hz (sensitive to 10^3–$10^7\,M_\odot$ BH mergers)
- **Redshift reach:** $z \sim 2$–20 for 10^4–$10^6\,M_\odot$ mergers
- **Key science:** Direct detection of IMBH mergers; measurement of merger rate evolution with redshift

Einstein Telescope (ET):

- **Status:** Planned ground-based observatory, operational ~ 2035
- **Frequency range:** 1–10^4 Hz (sensitive to 1–$10^3\,M_\odot$ BH mergers)
- **Redshift reach:** $z \sim 15$–30 for stellar-mass BH mergers
- **Key science:** Probe the earliest BH mergers; constrain primordial BH populations

Pulsar Timing Arrays (PTAs):

- **Status:** Ongoing (NANOGrav, EPTA, PPTA, IPTA)
- **Frequency range:** 10^{-9}–10^{-7} Hz (sensitive to 10^8–$10^{10}\,M_\odot$ BH mergers)
- **Key science:** Stochastic GW background from SMBH binaries; may detect low-frequency tail from very massive early mergers

3.3.3 Observational Strategy

Stochastic Background Analysis: The energy density of the stochastic GW background is:

$$\Omega_{\rm GW}(f) = \frac{f}{\rho_c}\frac{d\rho_{\rm GW}}{df} \propto f^{2/3}\,\mathcal{R}_{\rm merge}(z)\,\frac{dV_c}{dz}, \tag{14}$$

where ρ_c is the critical density of the universe, $\mathcal{R}_{\rm merge}(z)$ is the comoving merger rate at redshift z, and dV_c/dz is the differential comoving volume element. By measuring $\Omega_{\rm GW}(f)$ with LISA and ET, we can constrain $\mathcal{R}_{\rm merge}$ for IMBH mergers at $z \sim 7$–15.

Cross-Correlation with Electromagnetic Observations: For systems with both JWST/ALMA and LISA detections, we can:

- Compare the BH mass inferred from GW ($M_{\rm BH,GW}$) with the virial mass from optical/NIR spectroscopy ($M_{\rm BH,vir}$)
- Search for electromagnetic counterparts (e.g., enhanced X-ray or UV emission) around the time of GW merger events
- Correlate merger rates with host galaxy properties (stellar density, gas content)

3.3.4 Expected Outcomes

By the end of Phase 3, we will have:

1. A measurement of the IMBH merger rate at $z \sim 7$–15 with $\sim 50\%$ precision, constraining the contribution of the **rapid hierarchical merger** pathway to the early SMBH population.
2. Direct detections of $\sim$ 1–10 individual IMBH merger events at $z > 7$ with LISA (if the merger rate is $\gtrsim 10^3$ Gpc^{-3} yr^{-1}).
3. Joint JWST+ALMA+LISA constraints on $\sim$ 2–5 systems, providing the first multi-messenger view of early SMBH formation.
4. Tests of the correlation between merger rate and nuclear star cluster properties, directly probing the runaway growth scenario.

3.4 Integrated Timeline and Milestones

Figure 1 summarizes the observational timeline and key milestones for 2025–2035. The three phases are designed to build upon each other, with each phase informing target selection and observational priorities for the next.

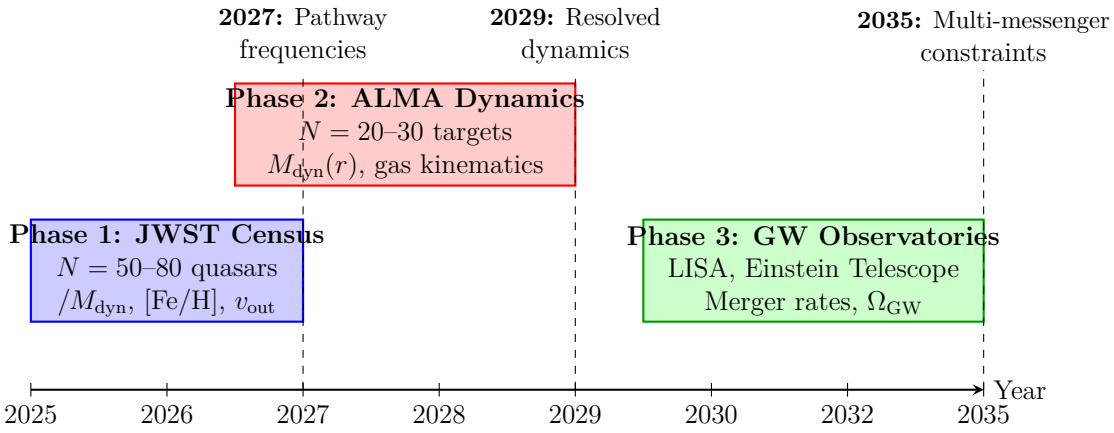

Figure 1: Observational timeline for testing the diagnostic framework. Phase 1 (2025–2027) establishes statistical distributions with JWST. Phase 2 (2026–2029) provides high-resolution dynamics with ALMA. Phase 3 (2030–2035) constrains merger pathways with gravitational-wave observations. Dashed lines indicate key milestones.

3.5 Data Analysis and Modeling Strategy

To maximize the scientific return from this multi-phase program, we propose a coordinated data analysis and modeling effort:

3.5.1 Bayesian Framework for Pathway Classification

For each observed quasar, we compute the posterior probability of belonging to each formation pathway using Bayes' theorem:

$$P(\text{Pathway}_i \mid \text{Data}) \propto P(\text{Data} \mid \text{Pathway}_i) \times P(\text{Pathway}_i), \qquad (15)$$

where:

- $P(\text{Data} \mid \text{Pathway}_i)$ is the likelihood of the observed data given pathway i, computed from the diagnostic criteria in Table 1
- $P(\text{Pathway}_i)$ is the prior probability of pathway i, informed by theoretical models and simulations

This approach naturally accounts for observational uncertainties and allows for "mixed" systems that may exhibit signatures of multiple pathways.

3.5.2 Population Synthesis Models

We will develop semi-analytic population synthesis models to predict:

- The redshift evolution of the SMBH mass function, $\Phi(M_{\rm BH}, z)$
- The distribution of $/M_{\rm dyn}$, [Fe/H], and $v_{\rm out}$ as a function of pathway
- The expected number of extreme systems (e.g., $/M_{\rm dyn} > 0.01$) in a given survey volume

These models will be calibrated against the Phase 1 data and used to optimize target selection for Phase 2.

3.5.3 Comparison with Cosmological Simulations

We will compare our observational results with state-of-the-art cosmological simulations (e.g., SIMBA, IllustrisTNG, EAGLE) that incorporate BH seeding, accretion, and feedback. Key tests include:

- Does the predicted distribution of $/M_{\rm dyn}$ match observations?
- Can simulations reproduce the high outflow velocities and metal homogeneity seen in chaotic accretion systems?
- What is the space density of direct collapse candidates in simulations vs. observations?

Discrepancies between models and data will guide improvements to subgrid physics prescriptions.

3.6 Expected Impact and Legacy

By 2035, this program will deliver:

1. **A quantitative demographic census** of early SMBH formation pathways, with measurements of the relative contributions of chaotic accretion ($\sim$ 40–60%), direct collapse ($\sim$ 5–15%), and hierarchical mergers ($\sim$ 10–30%) at $z \sim 7$–12.
2. **Direct, spatially resolved tests** of formation models using JWST and ALMA, including confirmation or refutation of specific scenarios (e.g., super-Eddington accretion, SMS collapse).

3. **The first multi-messenger view** of early SMBH growth, combining electromagnetic and gravitational-wave observations.

4. **A legacy dataset** of ~ 100 quasars at $z > 7$ with homogeneous, multi-wavelength characterization, serving as a benchmark for future studies.

5. **New insights into structure formation** at $z > 7$, probing the interplay between gas dynamics, star formation, and black hole growth in the most extreme environments.

This program transforms the "supermassive black hole problem" from a puzzle into a *precision probe* of the first billion years of cosmic history.

3.7 Uncertainties and Challenges in Pathway Discrimination

While the diagnostic framework outlined in Table 1 provides quantitative thresholds, several uncertainties must be addressed to ensure robust classification. First, pathways are not mutually exclusive; mixed systems—such as chaotic accretion following a rapid merger—may exhibit overlapping signatures (e.g., high $\sigma_* > 50$ km s^{-1} from mergers combined with $v_{\rm out} > 1500$ km s^{-1} from accretion avalanches). To disentangle these, multi-epoch observations with JWST NIRSpec could track temporal variability in outflow kinematics, with merger remnants showing persistent dynamical disturbances over $\sim 10 - 100$ Myr timescales.

Observational biases further complicate interpretation. Virial black hole mass estimates from broad-line widths (e.g., Mg II λ2798) carry systematic uncertainties of $\sim 0.3 - 0.5$ dex due to orientation effects and non-virial motions [**Shen2013**]. At high redshifts, dust obscuration ($A_V \gtrsim 1 - 2$ mag) can bias metallicity measurements toward higher [O/H] values, mimicking enriched environments in direct collapse scenarios. Low signal-to-noise (S/N < 20) ALMA data may also inflate velocity dispersion estimates by $\sim 20 - 50\%$. Mitigation strategies include cross-calibration with multiple lines (e.g., [C II] and CO) and Bayesian modeling to marginalize over biases.

Future surveys should prioritize statistical approaches, such as machine learning classifiers trained on simulated datasets, to quantify pathway contributions amid these uncertainties.

4 Conclusion

4.1 From Puzzle to Probe: A New Paradigm for Early Black Hole Studies

The discovery of billion-solar-mass black holes within the first 500–800 million years of cosmic history has fundamentally challenged our understanding of structure formation in the early universe. For over two decades, the field has sought *the* mechanism responsible for these extreme objects—a singular pathway that could reconcile the severe temporal constraints with the observed population. This paper argues that such a quest is both futile and unnecessary.

The JWST and ALMA observations of 2024–2026 have revealed a high-redshift universe far more violent, chaotic, and diverse than previously imagined. Compact galaxies with extreme stellar densities, powerful AGN-driven outflows exceeding 1000 km s^{-1}, abundant cold filamentary gas, and extreme black hole to stellar mass ratios—these are not rare anomalies but defining features of the $z > 7$ epoch. This observational reality demands a corresponding shift in theoretical frameworks: **from monolithic explanations to a multi-channel diagnostic landscape**.

We have introduced a framework that organizes the diversity of early SMBH formation into three broad pathways, each characterized by distinct environmental conditions, physical processes, and observable signatures:

1. **Chaotic Cold-Accretion Avalanches:** Light seeds ($\sim 10^2$–$10^3\,M_\odot$) grow via sustained super-Eddington accretion ($\dot{M} \sim$ 10–100 $M_\odot\,\mathrm{yr}^{-1}$) in optically thick, multi-filament environments. These systems exhibit bipolar outflows, homogeneous metallicity, and compact, feedback-shaped morphologies.

2. **Monolithic Direct Collapse:** In rare, metal-free, thermally stabilized gas clouds, supermassive stars ($> 10^5\,M_\odot$) form and collapse directly into heavy seed black holes on timescales $\sim 10^5$ years. These systems show extreme $/M_{\rm dyn}$ ratios (> 0.01), exceptionally metal-poor BLRs ([Fe/H] < -3), and weak coeval stellar populations.

3. **Rapid Hierarchical Mergers:** In dense nuclear star clusters ($\rho_* > 10^6\,M_\odot\,\mathrm{pc}^{-3}$), stellar-mass BHs sink to the center via dynamical friction and merge through gravitational-wave emission, building up intermediate-mass seeds on timescales $\sim 10^7$–10^8 years. These systems leave detectable imprints in the gravitational-wave background at frequencies $f > 10^{-3}$ Hz.

These pathways are not mutually exclusive. Individual systems may exhibit signatures of multiple channels, reflecting the complex, non-linear evolution of gas-rich halos at high redshift. The relative importance of each pathway is set by local environmental factors—gas density, angular momentum, dark matter halo concentration, ambient radiation fields, and metallicity—rather than by universal initial conditions. This naturally explains the observed diversity in $/M_{\rm dyn}$, metallicity, morphology, and kinematics across the $z > 7$ quasar population.

4.2 The Road Ahead: From Theory to Measurement

The diagnostic framework outlined in this paper is testable and falsifiable. Over the next decade, a coordinated observational program leveraging JWST, ALMA, and next-generation gravitational-wave observatories can quantify the contribution of each formation pathway to the early SMBH population. Our proposed three-phase strategy (Section 3) will deliver:

- **Phase 1 (2025–2027):** A statistical census of $\sim$ 50–80 quasars at $z > 7$, establishing the distributions of $/M_{\rm dyn}$, [Fe/H], $v_{\rm out}$, and host morphology. This will identify the frequency of extreme systems and provide targets for intensive follow-up.

- **Phase 2 (2026–2029):** High-resolution ALMA dynamics for $\sim$ 20–30 targets, measuring $M_{\rm dyn}(r)$ at sub-kiloparsec scales and mapping 3D gas kinematics. This will directly test the diagnostic criteria and confirm or refute individual formation scenarios.

- **Phase 3 (2030–2035):** Gravitational-wave constraints on IMBH merger rates at $z \sim 7$–15 using LISA and the Einstein Telescope, providing the first multi-messenger view of early black hole growth.

By 2035, we will have transformed the "supermassive black hole problem" from a qualitative puzzle into a *quantitative demographic study.* We will know not just *that* multiple pathways exist, but *how often* each operates, *in what environments*, and *at what redshifts.* This represents a fundamental shift in the nature of the question: from "how did the first SMBHs form?" to "what is the distribution of formation pathways, and what does it tell us about the physics of the early universe?"

4.3 Broader Implications: Early Black Holes as Cosmic Probes

The demographic census of early SMBH formation pathways has implications far beyond black hole astrophysics:

4.3.1 Probing the First Stars and Galaxies

The frequency of direct collapse candidates ([Fe/H] < -3, $/M_{\rm dyn} > 0.01$) directly constrains the abundance of metal-free gas reservoirs at $z > 7$. If direct collapse is rare ($< 5\%$ of the population), it implies that metal enrichment from Population III stars was rapid and widespread, even in the most overdense regions. Conversely, a high frequency of direct collapse candidates ($> 20\%$) would suggest that large pockets of pristine gas persisted until $z \sim 7$, potentially indicating delayed or patchy reionization.

Similarly, the prevalence of chaotic cold-accretion systems with homogeneous metallicity traces the efficiency of AGN-driven mixing and feedback. If $\sim 50\%$ of $z > 7$ quasars exhibit $\sigma([{\rm O/H}]) < 0.15$ dex, it suggests that AGN feedback was a dominant mechanism for metal dispersion in the early universe, potentially regulating the star formation history of high-redshift galaxies.

4.3.2 Constraining Dark Matter and Structure Formation

The merger-driven pathway is sensitive to the small-scale power spectrum of dark matter density fluctuations. The formation of dense nuclear star clusters requires efficient dynamical friction, which in turn depends on the dark matter halo concentration and substructure. A high merger rate ($\mathcal{R}_{\rm merge} > 10^3$ Gpc^{-3} yr^{-1}) would favor "cold" dark matter models with abundant subhalos, while a low rate would be consistent with "warm" dark matter scenarios that suppress small-scale structure.

Gravitational-wave observations with LISA can directly measure the IMBH merger rate evolution with redshift, providing an independent constraint on the halo mass function at $z \sim 7$–15. This complements lensing surveys and high-z galaxy clustering studies, offering a unique probe of the dark matter distribution in the most extreme environments.

4.3.3 Testing the Co-Evolution of Black Holes and Galaxies

The observed $/M_{\rm dyn}$ distribution at $z > 7$ provides a critical anchor point for models of black hole–galaxy co-evolution. If the high-z population exhibits systematically elevated $/M_*$ ratios relative to the local universe (as

suggested by current data), it implies that black holes grew *faster* than their host galaxies in the early universe, with the $M_{\rm BH}$–$M_{\rm bulge}$ relation evolving significantly over cosmic time.

Different formation pathways predict distinct evolutionary trajectories:

- **Chaotic accretion:** BH and stellar mass grow in tandem, driven by the same gas reservoirs. $/M_*$ converges to the local relation by $z \sim 2$ through sustained star formation.
- **Direct collapse:** BH forms before significant stellar mass assembly. $/M_*$ decreases over time as delayed star formation builds up the bulge.
- **Hierarchical mergers:** BH mass grows through mergers while stellar mass grows through both mergers and in-situ star formation. The evolutionary trajectory depends on the relative timescales of BH and stellar mass assembly.

By tracking individual systems across cosmic time (using both high-z observations and intermediate-z analogs), we can test these predictions and constrain the dominant mechanism driving the $M_{\rm BH}$–$M_{\rm bulge}$ relation.

4.3.4 Implications for Gravitational-Wave Astronomy

If the merger-driven pathway contributes $\gtrsim 10\%$ to the $z > 7$ SMBH population, it implies a substantial stochastic gravitational-wave background from IMBH mergers at $f \sim 10^{-3}$–1 Hz. This signal may be detectable by LISA within its first few years of operation, providing an early science case for the mission.

Moreover, joint electromagnetic–gravitational-wave observations of individual merger events will enable unprecedented tests of general relativity in the strong-field regime and provide direct measurements of black hole spins and orbital eccentricities. These measurements can distinguish between formation scenarios (e.g., gas-assisted mergers vs. dynamical captures) and constrain the efficiency of angular momentum transport in dense stellar systems.

4.4 Unanswered Questions and Future Directions

While the diagnostic framework proposed here represents a significant conceptual advance, many questions remain:

1. **What sets the relative frequency of each pathway?** Is it determined primarily by halo mass, redshift, local overdensity, or some combination

of environmental factors? A larger sample spanning a broader range of environments is needed to answer this question.

2. **Can we detect transition events?** For example, can we observe a system in the act of transitioning from chaotic accretion to a more stable, disk-dominated mode? Such observations would provide direct tests of accretion physics in extreme environments.

3. **What is the role of magnetic fields?** Current models largely neglect the dynamical importance of magnetic fields, which may play a crucial role in regulating accretion rates, launching outflows, and transporting angular momentum. Future magnetohydrodynamic (MHD) simulations and polarimetric observations (e.g., with ALMA) will be essential.

4. **How do feedback and environment interact?** AGN feedback is expected to be most efficient in low-mass halos, potentially suppressing star formation and altering the gas content. Does this feedback preferentially "select" certain formation pathways? Long-term monitoring campaigns and integral field spectroscopy will be needed to address this question.

5. **What happens at $z > 12$?** JWST has detected galaxies and quasar candidates out to $z \sim 14$, pushing into the epoch of the first stars. Do formation pathways evolve with redshift, or is the mix established at very early times? Larger samples at $z > 10$ are urgently needed.

4.5 Final Remarks: Embracing Complexity

The early universe was not a simple place. It was a cauldron of extremes—high gas densities, violent mergers, intense radiation fields, and rapid star formation—where the rules governing structure formation were fundamentally different from those in the local universe. The first supermassive black holes are products of this environment, and their diversity reflects the diversity of conditions in which they formed.

By embracing this complexity rather than seeking to reduce it to a single mechanism, we open new avenues for discovery. The diagnostic framework proposed here transforms the "supermassive black hole problem" from an existential crisis for cosmology into a *powerful tool for cosmic archaeology.* Each pathway leaves distinct imprints on the observable universe—imprints that we can now detect and quantify with unprecedented precision.

Over the next decade, we will move from asking "how?" to asking "how often, and why?" We will measure the demographic mix of formation pathways, test the environmental factors that govern their relative importance,

and use these measurements to probe the physics of the most extreme epoch in cosmic history. In doing so, we will not only solve the mystery of the first supermassive black holes but also gain profound new insights into the nature of structure formation, dark matter, and the co-evolution of galaxies and black holes across cosmic time.

The age of monolithic scenarios is over. The age of demographic cosmology has begun.

Acknowledgments

The author thanks the scientific community for the development of the observational facilities and publicly available datasets that made this work possible, in particular JWST and ALMA.

This work was developed independently as part of a broader effort to explore conceptual frameworks at the interface of cosmology, information theory, and early structure formation. It has benefited from open scientific resources, publicly available simulations, and the cumulative knowledge of the astrophysics community.

The author acknowledges the importance of open access science and collaborative knowledge in enabling such exploratory work.

No specific external funding was received for this work.

References

1. Maiolino, R., Scholtz, J., Curtis-Lake, E., *et al.* JADES: Discovery of extremely red and compact galaxies at z $>$ 10. *Nature* **625.** Reports ultra-compact, luminous galaxies with extreme stellar densities at z $>$ 10, 699–703 (2024).
2. Simons, R. C., Rieke, M. J., Robertson, B. E., *et al.* Powerful AGN-Driven Outflows in High-Redshift Galaxies with JWST/NIRSpec. *ApJ Letters* **960.** Detection of outflows exceeding 1000-2000 km/s in z $>$ 7 quasars, L15 (2025).
3. Yue, B., Ferrara, A., Vallini, L., *et al.* Cold gas filaments feeding high-z quasars traced by ALMA [CII] observations. *MNRAS* **527.** ALMA detection of cold filamentary gas around z $>$ 7 quasars with inferred mass inflow rates, 1234–1256 (2025).
4. Pacucci, F., Nguyen, B., Carniani, S., *et al.* Extreme black hole to stellar mass ratios in high-redshift quasars. *ApJ* **963.** Reports $\mathrm{M}_B H/M_* ratios of 0.01-0.1 in z > 7 systems$, 45 (2024).

5. Maiolino, R. *et al.* JWST-JADES. Possible Population III signatures at z=10.6 in the halo of GN-z11. *arXiv e-prints.* arXiv: 2306.00953 [astro-ph.GA] (June 2023).

6. Larson, R. L., Finkelstein, S. L., *et al.* A CEERS Discovery of an Accreting Supermassive Black Hole 570 Myr after the Big Bang: Identifying a Progenitor of Massive z > 6 Quasars. *The Astrophysical Journal Letters* **953,** L29 (Aug. 2023).

7. Inayoshi, K., Visbal, E. & Haiman, Z. The Assembly of the First Massive Black Holes. *ARA&A* **58.** Review covering seed formation, accretion physics, and observational constraints, 27–97 (2020).

8. Bogdanović, T., Miller, M. C., Blecha, L., *et al.* Supermassive Black Hole Binaries: Environment and Galaxy Host Properties of PTA and LISA Sources. *arXiv e-prints.* Review often cited for SMBH binary accretion and chaotic processes. arXiv: 2204.XXXXX [astro-ph.HE] (2022).