

Rise of the forsaken relics: connecting present-day stellar streams and phase-mixed galaxies to the Epoch of Reionization

ARITRA KUNDU ¹, ROBYN SANDERSON ¹, ADAM LIDZ ¹, PRATIK J. GANDHI ², ANDREW WETZEL ³,
ROBERT FELDMANN ⁴, NONDH PANITHANPAISAL ^{5,6}, JASJEEV SINGH ¹, AND MICHAEL BOYLAN-KOLCHIN ⁷

¹*Department of Physics & Astronomy, University of Pennsylvania, 209 S 33rd St., Philadelphia, PA 19104, USA*

²*Department of Astronomy, Yale University, P.O. Box 208101 New Haven, CT 06520-8101 USA*

³*Department of Physics & Astronomy, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA*

⁴*Department of Astrophysics, Universität Zürich, Winterthurerstrasse 190, Zurich, 8057, Switzerland*

⁵*The Observatories of the Carnegie Institution for Science, 813 Santa Barbara St, Pasadena, CA 91101, USA*

⁶*TAPIR, California Institute of Technology, Pasadena, CA 91125, USA*

⁷*Cosmic Frontier Center, Department of Astronomy, The University of Texas at Austin, 2515 Speedway, Austin, TX 78712, USA*

ABSTRACT

The ‘near-far’ approach to studying reionization leverages the star formation histories of the Milky Way (MW) or Local Group (LG) galaxies, derived from resolved photometry, to infer the low-mass/faint-end of the stellar mass functions (SMFs) or the ultraviolet luminosity functions (UVLFs) of high-redshift galaxies ($z \gtrsim 6$), beyond the current JWST detection limits ($M_{\text{UV}} \gtrsim -15$). Previous works considered only intact low-mass galaxies in the MW and LG, neglecting disrupted galaxies such as stellar streams and phase-mixed objects. Using the FIRE-2 simulations, we show that these disrupted galaxies contribute up to $\sim 50\%$ of the total stellar mass budget of the proto-MW/LG at $z = 6 - 9$. Including all the progenitors of these disrupted galaxies improves the normalization of the recovered SMFs/UVLFs by factors of $\sim 2-3$ and reduces the halo-to-halo variation in the slope by $\sim 20-40\%$. This enables robust constraints down to at least the resolution limit of the simulations, near $M_{\star} \sim 10^5 M_{\odot}$ or $M_{\text{UV}} \sim -10$ at $z \gtrsim 6$. We also show that ‘fossil record’ reconstructions—which assume each present-day system descends from a single reionization-era progenitor—are sensitive to the stellar mass/UV magnitude thresholds, which introduces bias in the inferred low-mass/faint-end slopes. Additionally, we demonstrate that neglecting disrupted systems underestimates the contribution of galaxies with $M_{\text{UV}} \lesssim -15$ to the reionization-era UV luminosity density. Finally, we estimate that a significant fraction ($\sim 50\%$) of streams with $M_{\star} \gtrsim 10^6 M_{\odot}$ at $z=0$ should be detectable from upcoming Rubin Observatory and Roman Space Telescope observations.

1. INTRODUCTION

The Epoch of Reionization (EoR) is the time period when the first stars, galaxies, and accreting black holes emitted ultraviolet (UV) radiation and gradually photoionized neutral hydrogen in the intergalactic medium (IGM) throughout the universe (Barkana & Loeb 2001; Furlanetto et al. 2006). The timing and nature of this process remain highly uncertain, while the properties of the ionizing sources are currently unknown. The most likely scenario is that the UV radiation from star-forming galaxies drives reionization, with the process completing around $z \sim 6$ (Shapiro et al. 1994; Madau et al. 1999; Faucher-Giguère et al. 2008; Becker & Bolton 2013), although precisely which galaxies dominate is unclear. The Hubble Space Telescope (HST) has directly observed faint reionization-era galaxies down to

$M_{\text{UV}} \sim -13$ in fields behind lensing clusters (Finkelstein et al. 2015; Bouwens et al. 2017; Livermore et al. 2017; Atek et al. 2018; Bouwens et al. 2022b), albeit with sizable uncertainties at the faint end. Ongoing and future observations with the James Webb Space Telescope (JWST) are pushing these measurements to $z \gtrsim 10$ and slightly fainter magnitudes (Finkelstein et al. 2023; Leung et al. 2023; Navarro-Carrera et al. 2023; Pérez-González et al. 2023; Atek et al. 2024; Finkelstein et al. 2024; Adams et al. 2024; Harikane et al. 2025; Chemerynska et al. 2025).

However, some reionization models predict that fainter yet more numerous low-mass galaxies, with, e.g. $M_{\text{UV}} \sim -15$ to -8 , are major contributors to the total budget of ionizing photons (Kuhlen & Faucher-Giguère 2012; Wise et al. 2014; Robertson et al. 2015;

Atek et al. 2015; Ma et al. 2018a,b; Finkelstein et al. 2019; Atek et al. 2024; Mascia et al. 2024; Simmonds et al. 2024; Harshan et al. 2024; Wu & Kravtsov 2024). Many of these sources will remain beyond the reach of even JWST (Boylan-Kolchin et al. 2015, 2016; Weisz & Boylan-Kolchin 2017). While observations of lensed galaxies can probe fainter magnitudes (down to $M_{UV} \sim -12$), they introduce significant systematic uncertainties due to magnification effects (Atek et al. 2018; Bouwens et al. 2022a). Additionally, low-mass galaxies provide valuable insights into the impact of stellar feedback and the nature of dark matter on galaxy formation, particularly at high redshifts (Sun & Furlanetto 2016; Bullock & Boylan-Kolchin 2017; Sales et al. 2022; Sippl & Lidz 2024; Sippl et al. 2025). Novel approaches for probing the faint/low-mass-end of the $z \gtrsim 6$ galaxy UV luminosity function (UVLF)/stellar mass function (SMF) are hence valuable - not only in determining the role of low-luminosity galaxies in reionizing the universe, but also in understanding feedback processes and the properties of dark matter.

One promising technique is the ‘near-far’ approach, which relies on studies of Milky Way (MW) satellite galaxies and Local Group (LG) low-mass galaxies at the present-day ($z = 0$) to infer properties of their $z \gtrsim 6$ progenitors (Boylan-Kolchin et al. 2015, 2016; Weisz & Boylan-Kolchin 2017; Weisz et al. 2014a). This employs color-magnitude diagrams (CMDs) of nearby MW/LG low-mass galaxies from resolved HST/JWST photometry to reconstruct their star-formation histories (SFHs) (Weisz et al. 2014b,a) and UV luminosities at $z \gtrsim 6$ (Weisz et al. 2014c; Weisz & Boylan-Kolchin 2017), using stellar population synthesis models. This indirect approach probes fainter magnitudes than direct high-redshift observations, thereby providing information about the faint-end/low-mass-end of the UVLFs/SMFs at $z \gtrsim 6$.

Previous works applied this technique to intact low-mass galaxies (Boylan-Kolchin et al. 2016; Weisz & Boylan-Kolchin 2017; Weisz et al. 2014c; Skillman et al. 2017; Savino et al. 2023; McQuinn et al. 2024) in the LG. This strategy assumes a one-to-one mapping from high- z low-mass galaxies to their present-day LG counterparts, ignoring the cosmological assembly histories of the low-mass galaxies and, in the case of satellite galaxies, their host halos. One consequence of this assumption is that the high- z SMF/UVLF inferred from the LG fossil record recovers only ≈ 15 -20% of the total amplitude (Gandhi et al. 2024). This is because it lacks information about the different progenitors that merged to form the present-day system. Gandhi et al. (2024) analyzed the impact of these mergers by including mul-

tipale progenitors of present-day low-mass galaxies in cosmological baryonic zoom-in simulations, and tested the near-far technique in terms of recovering the slope of the low-mass-end of the SMF at $z \gtrsim 6$. They found that the fossil record of surviving low-mass galaxies in the MW/LG can recover the slope of the faint-/low-mass-end of the UVLF/SMF of the proto-MW/LG at $z \gtrsim 6$, albeit with a large halo-to-halo scatter.

Although these studies illustrate the general promise of the near-far technique, they do not yet fully account for the fact that most high- z progenitors fail to survive intact as present-day MW/LG low-mass galaxies (Santistevan et al. 2020; Horta et al. 2024). In fact, most high- z progenitor galaxies will be completely disrupted by mergers and tidal forces ($\sim 50\%$ of the proto-MW/LG, Gandhi et al. 2024), leaving behind both coherent stellar streams and galaxy remnants that are now part of the equilibrium phase-space distribution defined by the host galaxy (‘phase-mixed galaxies’). These merged substructures have systematically different properties from the intact satellites, such as lower metallicities due to earlier star formation (e.g. Panithanpaisal et al. 2021; Cunningham et al. 2022; Horta et al. 2023). The near-far technique usually ignores these disrupted galaxies when inferring the high- z UVLF/SMF. However, the high- z progenitors of these substructures also contribute to the ionizing photon budget at $z \gtrsim 6$, and their exclusion may lead to inaccurate conclusions regarding the role of low-luminosity galaxies in reionizing the universe. The reconstructed SMFs/UVLFs from intact galaxies alone require large corrections to the normalizations of these functions. Furthermore, simulations show a large scatter in the shape of the reconstructed SMFs/UVLFs across different MW/LG-like halos (halo-to-halo scatter) (Gandhi et al. 2024). Incorporating disrupted galaxies into the analysis could potentially help reduce this scatter and provide more robust constraints on the shape and normalization of reconstructed high- z MW/LG SMF/UVLF. Hence, a comprehensive study of the impact of including disrupted galaxies on the near-far technique is crucial.

Concurrent with more comprehensive theoretical tests of the near-far technique, it is important to assess the detectability of the disrupted galaxies for use in this approach. To date, the *Gaia* mission and the Dark Energy Survey (DES) have detected significant numbers of low surface brightness ($\mu > 27$ mag/arcsec², Johnston et al. 2008) galaxies, such as stellar streams and phase-mixed galaxies, in the MW, greatly enhancing our understanding of the Milky Way’s formation history (e.g., Ibata et al. 2001; Newberg et al. 2002; Balbinot et al. 2016;

Shipp et al. 2018; Necib et al. 2020; Belokurov et al. 2018; Martin et al. 2022). However, the rich accretion history probed by the disrupted galaxies extends beyond $> 30 \text{ mag arcsec}^{-2}$ (Laine et al. 2018). Upcoming telescopes like the Vera C. Rubin Observatory (Ivezić et al. 2019), Roman Space Telescope (Spergel et al. 2013), and ARRAKIHS (Guzmán 2022) will enable detections of these fainter low surface brightness substructures both within the MW and beyond (Pearson et al. 2022a,c; Dey et al. 2023; Aganze et al. 2024). This presents us with an opportunity to utilize these substructures to infer the properties of their high- z progenitors and learn about their likely contribution to reionization, thereby also comprehensively testing the feasibility of the near-far technique in probing the EoR.

In this paper, we investigate these issues using a suite of simulated MW/LG-like galaxies from zoomed cosmological hydrodynamic simulations. In section 2, we describe the simulations used in this work. Section 3 discusses our criteria for classifying present-day substructures and our methodology for tracking them back to their simulated high- z progenitors. Through Sections 4 to 8, we illustrate the impact of adding disrupted galaxies to the near-far technique and also test the feasibility of using them by imposing detectability constraints. Finally, we summarize these results in Section 9.

2. SIMULATIONS

2.1. FIRE-2 simulations

We use cosmological zoom-in baryonic simulations from the Feedback in Realistic Environments (FIRE-2) project (Wetzel et al. 2023, 2025)¹. They are run with the GIZMO code (Hopkins 2015; Hopkins et al. 2018a), which uses a mesh-free, finite volume Lagrangian method for accurate hydrodynamics. Baryonic physical processes like star formation and stellar feedback are implemented using the FIRE-2 numerical prescriptions (Hopkins et al. 2018a). The multiphase interstellar medium (ISM) in each simulated galaxy is captured with realistic radiative heating and cooling processes across temperatures from 10 K to 10^{10} K.

Simulated star particles form out of gas parcels satisfying a Jeans instability criterion, and inherit the mass and elemental abundances from their respective gas cells. Each star particle in FIRE represents a single-age, single-abundance stellar population, whose evolution is based on stellar population synthesis models from STARBURST99 (Leitherer et al. 1999). FIRE also models stellar feedback processes like core-collapse

and Type Ia supernovae, radiation pressure, continuous mass loss, photoionization and photo-electric heating in a time-resolved manner. Furthermore, FIRE simulations implement a spatially uniform UV/X-ray ionizing background that evolves with redshift (Faucher-Giguère et al. 2009). Some of the caveats of this model are that it neglects the inhomogeneous nature of reionization and also has an earlier reionization time ($z \sim 10$) than recent measurements suggest ($z \sim 7$) (Planck Collaboration et al. 2020). But all the simulations that we use in this work self-consistently employ this model and hence will not result in any simulation-to-simulation difference due to these caveats. Additionally, these simulations include spurious heating of neutral gas at $\lesssim 1000$ K at $z \gtrsim 10$ due to cosmic rays, which suppresses star formation in low-mass halos at high redshifts. But this has no effect after reionization begins and hence our results at $z \lesssim 9$ are self-consistent (Gandhi et al. 2024). For a more detailed description of the physics implemented in FIRE simulations, see Hopkins et al. (2018a,b).

The FIRE-2 simulations successfully generate isolated MW-like galaxies and MW/M31-like galaxy pairs, reproducing the observed properties of MW stellar disks, bars, and spiral structures (Ma et al. 2017; Hopkins et al. 2018a; Sanderson et al. 2020; Ansar et al. 2025); as well as stellar mass functions, radial distance distributions, and star formation histories of their satellite galaxies (Wetzel et al. 2016; Garrison-Kimmel et al. 2019a,b; Samuel et al. 2020, 2021). The central massive MW-like galaxy that forms self-consistently in these simulations efficiently destroys satellite galaxies and low-mass sub-halos that orbit within the inner ~ 30 kpc through gravitational tidal forces (Garrison-Kimmel et al. 2017; Barry et al. 2023). The resulting radial distribution of MW satellite galaxies is consistent with observations (Samuel et al. 2020). The tidal disruption of well-resolved simulated satellite galaxies also leads to a realistic population of stellar streams (Panithanpaisal et al. 2021; Arora et al. 2022; Horta et al. 2023; Shipp et al. 2023), albeit with some discrepancies in their pericenter and apocenter distributions compared to observed MW low-mass galaxy streams (Shipp et al. 2023). However, streams from globular clusters do not naturally form in the FIRE-2 simulations due to their limited resolution.

In this work, we make a catalog of surviving low-mass galaxies, stellar streams, and phase-mixed galaxies, along with their surrounding low-mass galaxies (field galaxies), around 13 MW-mass halos: seven isolated MW-like halos from the *Latte* suite (Wetzel et al. 2016), and three LG-like MW/M31 pairs from the *ELVIS* on *FIRE* suite (Garrison-Kimmel et al. 2019a). All simulations follow a Λ CDM cosmology, with cosmological pa-

¹ <https://fire.northwestern.edu/>

rameters roughly consistent with the [Planck Collaboration et al. \(2020\)](#) results. Cosmological initial conditions are generated at $z = 99$ using MUSIC ([Hahn & Abel 2011](#)); the zoomed regions are selected from periodic cosmological boxes of sizes 70–172 Mpc. The star/gas particles in the isolated Latte suite have an initial mass resolution of $\sim 7100 M_\odot$, while in the paired simulations of the ELVIS suite, it is $3500\text{--}4000 M_\odot$. The gravitational softening length (spatial resolution) for the star particles is fixed at ~ 4 pc, while the gas particles have an adaptive softening length with a minimum of ~ 1 pc. The dark matter (DM) particles have a resolution of $19000\text{--}39000 M_\odot$ in mass, and a spatial resolution of ~ 40 pc. The main host halos have a total mass of $M_{200\text{m}} \approx 1\text{--}2 \times 10^{12} M_\odot$ ² and a stellar mass of $M_\star \approx 10^{10\text{--}11} M_\odot$, which spans measured values of MW’s halo and stellar mass ([Sanderson et al. 2020](#)). In each simulation, 600 snapshots are saved from $z = 99$ to $z = 0$ with a typical spacing of ~ 25 Myr after $z \sim 2$, allowing accurate tracking of disrupted galaxies.

Since an ultimate goal of the near-far technique is to infer the shape of the *universal* SMF/UVLF at high redshifts, we also use results from the FIREbox^{HR} simulation ([Feldmann et al. 2025](#)) as a benchmark. FIREbox^{HR}, a higher-resolution re-run of FIREbox ([Feldmann et al. 2023](#)), is a cosmological volume simulation with a box size of 22.1 cMpc and with a mass resolution ($m_b \sim 7800 M_\odot$) similar to the Latte suite. It contains ~ 2000 galaxies brighter than $M_{\text{UV}} = -14$ at $z \sim 6$, providing a good statistical sample. It uses the same FIRE-2 physics to model baryonic processes down to $z \sim 6$, and encapsulates the multi-phase nature of ISM because of its high dynamic range and multi-channel stellar feedback. The simulated SMFs/UVLFs at $z \sim 6\text{--}14$ are in good agreement with JWST observations ([Feldmann et al. 2025](#)). Furthermore, the observed cosmic UV luminosity density at these redshifts is well-reproduced by FIREbox^{HR}. The simulation naturally explains recent JWST observations, which suggest a relatively high level of star formation activity even during early phases of the reionization era, supporting the accuracy of the baryonic processes implemented in FIRE-2. We use the SMFs/UVLFs from FIREbox^{HR} as our benchmark for the high- z SMFs/UVLFs, and compare them against the ones derived from the progenitors of present-day substructures in the core suite. We also note that Gandhi et al. (in prep) compares their re-constructed SMFs/UVLFs to a universal SMF/UVLF

from the statistical sample of galaxies from the FIRE-2 high-redshift zoom-in suite ([Ma et al. 2018a,b, 2019](#)) as a benchmark for the high- z SMFs/UVLFs, which also been found to agree well with recent high- z JWST observations ([Sun et al. 2023b](#)). Both sets of simulations show good agreement with each other with respect to the SMFs/UVLFs and cosmic UV luminosity density, though they may not match exactly.

2.2. Identifying low-mass galaxies

The DM halos and subhalos (smaller virialized halos within a larger halo) in each snapshot of the core simulations are identified using the ROCKSTAR 6D phase-space halo finder ([Behroozi et al. 2013a](#)) to create a halo catalog for each of the 600 snapshots. The (sub)halos are identified according to their radius, such that it encloses 200 times the mean matter density of the Universe ($R_{200\text{m}}$), and only the ones with bound mass fraction > 0.4 and at least 30 DM particles are kept. This criterion reduces low-resolution artifacts, particularly contamination from the low-resolution DM particles (the low-resolution particles are a consequence of the zoom-in re-simulation technique employed). After the halo catalogs have been formed, a merger tree is constructed by connecting the (sub)halos in time using CONSISTENT-TREES ([Behroozi et al. 2013b](#)). To ensure numerical stability, the halo catalogs and merger trees are constructed using only DM particles ([Samuel et al. 2020](#)).

Star particles are then assigned to each (sub)halo in post-processing (adapted from the method described in [Necib et al. 2019](#) and [Samuel et al. 2020](#)), with slightly different criteria at $z = 0$ and $z \geq 6$:

1. At $z = 0$, for each (sub)halo with radius $R_{200\text{m}}$ and center-of-mass velocity $V_{\text{circ,max}}$ as obtained from ROCKSTAR, star particles are selected whose positions lie within $0.8R_{200\text{m}}$ (out to a maximum radius of 30 kpc) and whose relative velocity falls within $2V_{\text{circ,max}}$.
2. Star particles whose positions are within $1.5 R_{90}$ (the radius enclosing 90 percent of the mass of member star particles) of *both* the center-of-mass position of member stars and the DM halo center, and whose velocities are within twice the stellar velocity dispersion of the particles within the subhalo, are selected.
3. Steps (1) and (2) are repeated until the sum of the masses of all member star particles, $M_\star$, converges to within 1 percent.

² ‘200m’ indicates a measurement relative to 200 times the mean matter density of the Universe

4. Finally, all halos with at least six star particles and an average stellar density $> 300 M_{\odot} \text{ kpc}^{-3}$ are selected.

This method is conservative enough to keep galaxy properties stable across time at lower redshifts, and excludes nearly all stream stars (Panithanpaisal et al. 2021; Samuel et al. 2020). These assignments are then used as a starting point for identifying streams, as described in Section 3.1. However, these criteria can be too stringent when assigning star particles to low-mass (sub)halos at $z \geq 6$, which in turn can lead to an incomplete low-mass (sub)halo population. In order to improve the overall completeness of the star particle assignment to each (sub)halo at $z \geq 6$, we follow the method implemented in Gandhi et al. (2024): we change the enclosing radius from $0.8 R_{200\text{m}}$ to $1.0 R_{200\text{m}}$ and do not apply any constraints on the velocity relative to $V_{\text{circ,max}}$. Furthermore, we keep all the halos with at least two star particles instead of six, since most of the galaxies that we analyze at high redshifts are progenitors of well-resolved galaxies at $z = 0$.

3. PRESENT-DAY SUBSTRUCTURES AND THEIR HIGH REDSHIFT PROGENITORS

3.1. Substructure classification algorithm

We classify present-day substructures into four categories:

1. **Surviving satellite galaxies:** Intact low-mass galaxies within the virial radius (radii) of the isolated (paired) MW-like (MW/M31-like) host(s) at $z = 0$. In this work, we consider the virial radius to be $R_{200\text{m}}$.
2. **Stellar streams:** Disrupted low-mass galaxies which are torn and stretched along their orbits by tidal forces from the host galaxy.
3. **Phase-mixed galaxies:** Disrupted low-mass galaxies which were once stellar streams, but are now merged with the host galaxy and share its equilibrium phase-space distribution.
4. **Field galaxies:** Intact galaxies that are outside the virial radius (radii) of the isolated (paired) MW-like (MW/M31-like) host(s) at $z = 0$, but within a maximum distance as discussed later.

For the intact galaxies (satellite + field) at $z = 0$, we select galaxies with stellar mass $M_{\star} \geq 10^{4.5} M_{\odot}$ (or number of star particles $\gtrsim 6$) within 2 Mpc of the center

of each isolated MW-like galaxy, or from the geometric center of mass of the LG-like pairs. These criteria ensure that our sample does not suffer from numerical issues due to very low-mass subhalos, and also avoid contamination by low-resolution DM particles (same as in Gandhi et al. 2024). We also exclude star particles associated with the $z = 0$ host galaxies themselves.

We use a modified version of the method described in Panithanpaisal et al. (2021) to identify the satellites, streams, and phase-mixed galaxies in each of the halos. We track each bound subhalo with non-zero stellar mass within the virial radius of the main host at $z = 0$ (~ 350 kpc). We start tracking the subhalos from the redshift ($z_{\text{MR}_{3:1}}$) values calculated in Horta et al. (2024). This is the redshift at which the stellar mass ratio of the most massive halo to the second most massive subhalo is 3:1 (see also Santistevan et al. 2020), and is used to define the time when the dominant host galaxy emerges. The subhalos that merge with the main progenitor galaxy before this redshift are considered to be part of the host galaxy and we do not include them. This is the earliest point in time that we can reasonably distinguish the host from satellites. Table 1 lists $z_{\text{MR}_{3:1}}$ for each of the host halos. We also update the selection of member stars from Panithanpaisal et al. (2021). Following Shipp et al. (2023), we require that each stream member star belongs to its progenitors for at least 10 snapshots. This removes contamination from unassociated satellites and from host disk stars that may be erroneously picked up by the halo finder due to coincidental proximity in phase-space.

We use three criteria to select and classify present-day substructures. We first require that the number of star particles be larger than 100 ($M_{\star} \gtrsim 5 \times 10^5 M_{\odot}$).³ To be considered disrupted rather than intact, the maximum pairwise separation between any two star particles must also be greater than 120 kpc. Finally, the local velocity dispersion σ of each disrupted substructure is compared to a stellar-mass dependent threshold $\sigma_{\text{th}}(M_{\star})$ (Panithanpaisal et al. 2021, Eq. 2). Substructures with $\sigma < \sigma_{\text{th}}(M_{\star})$ are considered streams, while those with $\sigma > \sigma_{\text{th}}(M_{\star})$ are considered phase-mixed galaxies. In some borderline cases where a substructure barely satisfies some criterion, but looks distinctively like another category, we reclassify the substructure manually. Table 1 shows the number of different substructures in each

³ We manually reduce the stellar mass requirement to $10^{4.5} M_{\odot}$ for the case of satellite galaxies to be consistent with Gandhi et al. (2024). This choice seems reasonable since these galaxies are more compact than streams and therefore easier to detect. In Section 8 we quantify detectability in terms of surface brightness.

Table 1. Properties of the FIRE-2 simulations at $z = 0$.

Simulation name	$M_{200\text{m}}$ ($\times 10^{12} M_{\odot}$)	$M_{\star,90}$ ($\times 10^{10} M_{\odot}$)	$R_{200\text{m}}$ (kpc)	$z_{\text{MR}_{3:1}}$	Number of			
					satellite galaxies	stellar streams	phase-mixed objects	field galaxies
m12i ^A	1.2	6.3	336	3.18	20	13	17	8
m12b ^B	1.4	8.5	358	1.59	19	13	10	21
m12f ^C	1.7	7.9	380	3.77	28	12	29	23
m12c ^B	1.3	5.8	351	1.32	33	11	7	42
m12m ^D	1.6	11.0	371	1.68	31	13	6	32
m12r ^E	1.1	1.7	321	1.02	24	5	6	20
m12w ^E	1.1	5.7	319	2.92	31	5	14	24
Romeo ^B	1.3	6.6	341	5.70	23	20	37	74
Juliet ^B	1.1	3.8	321	4.72	22	15	29	
Romulus ^F	2.1	9.1	406	1.70	38	14	25	59
Remus ^F	1.2	4.6	339	0.98	20	15	6	
Thelma ^F	1.4	7.1	358	4.29	20	11	24	64
Louise ^F	1.1	2.6	333	3.32	27	12	24	

NOTE—*Simulation name*: References listed below; top 7 are isolated MW-like and bottom three are paired MW/M31-like. $M_{200\text{m}}$ and $R_{200\text{m}}$: total mass and spherical radius within which the mean density is 200 times the matter density of the universe. $M_{\star,90}$: stellar mass within a spherical radius that encloses 90% of the stellar mass within 20 kpc. $z_{\text{MR}_{3:1}}$: the redshift at which the stellar mass ratio of the most massive halo to the second most massive subhalo is 3:1, which is when the dominant host galaxy emerges. Numbers of surviving satellite galaxies ($M_{\star} \gtrsim 10^{4.5} M_{\odot}$), stellar streams ($M_{\star} \gtrsim 10^6 M_{\odot}$), phase-mixed galaxies ($M_{\star} \gtrsim 10^6 M_{\odot}$), and field galaxies ($M_{\star} \gtrsim 10^{4.5} M_{\odot}$) are determined as described in Section 3.1.

References—A: Wetzel et al. (2016), B: Garrison-Kimmel et al. (2019a), C: Garrison-Kimmel et al. (2017), D: Hopkins et al. (2018a), E: Samuel et al. (2020), F: Garrison-Kimmel et al. (2019b).

of the FIRE-2 simulations, along with the properties of the host(s). We show examples of the substructures classified by the algorithm as satellite galaxies, stellar streams, phase-mixed galaxies, and field galaxies, in Figure 1, along with cases where we manually reclassify the original assignment according to the algorithm.

Unlike Panithanpaisal et al. (2021), we do not set any upper bound on the number of star particles, which leads us to include very massive objects ($M_{\star} \sim 5 \times 10^9 M_{\odot}$). Clearly, the progenitors of these substructures also contribute to the UV and ionizing photon budget during reionization. However, these disrupted systems can exhibit complex morphologies, with some parts displaying a coherent tidal tail and other parts appearing more phase-mixed (Figure 1, bottom center panel). Most of these cases are classified as phase-mixed according to our criteria, but if the object has a prominent tidal tail, we manually reclassify it as a stream, as was done for the stream in Figure 1.

Another common scenario for manual reclassification is shown in the bottom left panel of Figure 1. This satellite galaxy was originally classified as a stream by the algorithm because of a single pair of particles separated by more than 120 kpc in distance. We reclassify it as a low-mass satellite galaxy since the bound part clearly dominates; such a substructure would almost certainly be detected as a satellite.

Finally, in some low mass phase-mixed galaxies, the local velocity dispersion is below the threshold, while the pairwise particle separations are smaller than 120 kpc. These objects are classified as satellites by the algorithm, but their star particles are distributed around the center of the host galaxy (Figure 1, bottom right panel). These are manually reclassified as phase-mixed.

We stress that our classifications in this work are based on the entire distribution of simulated star particles. A more robust classification requires taking observability effects into the picture as in Shipp et al. (2023),

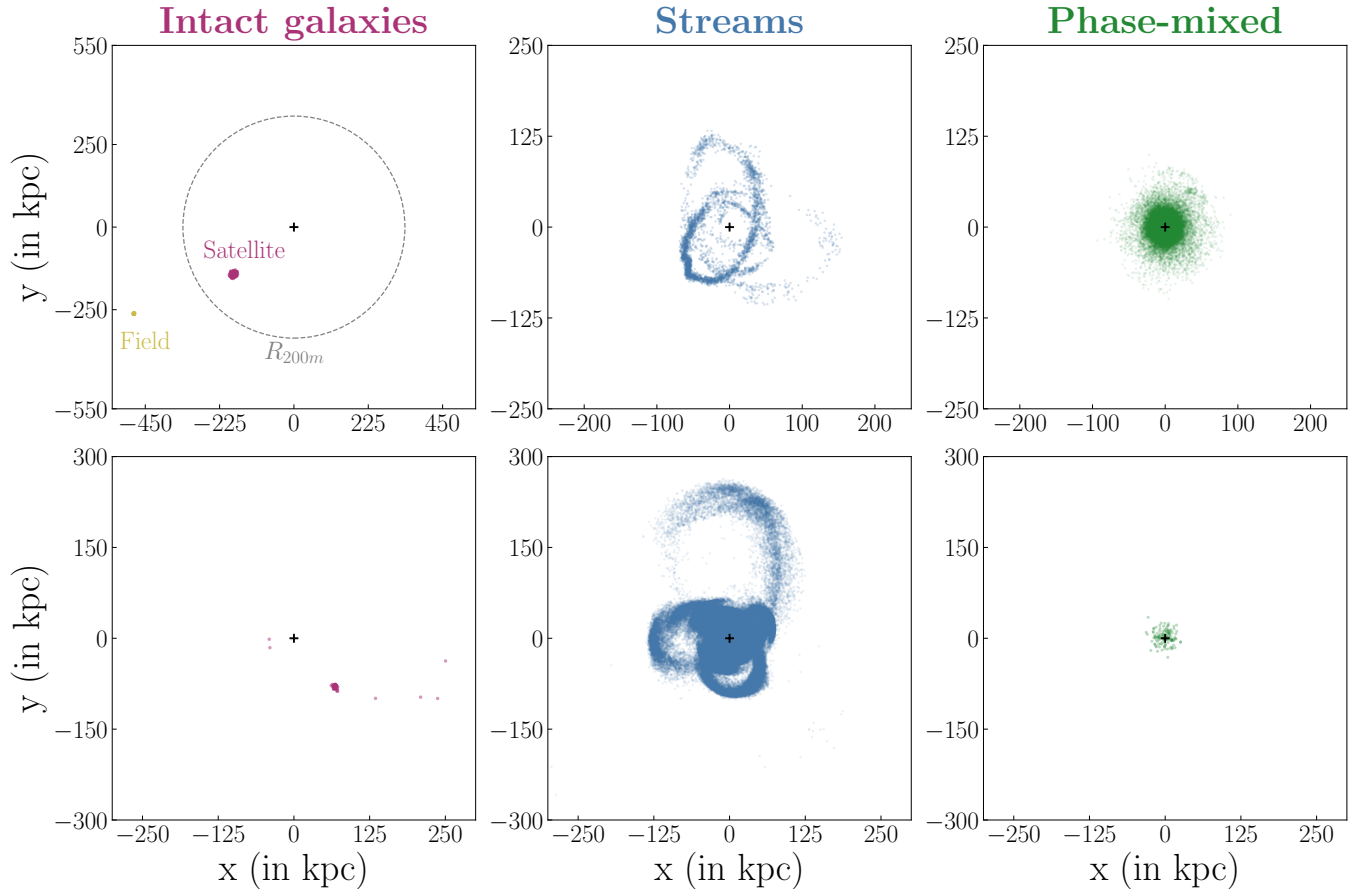


Figure 1. Present-day substructures classified by our algorithm, with manual reclassification. The center of the host galaxy is denoted by the black cross in each panel. **Top:** Three examples correctly classified by the automated algorithm described in Section 3.1. *Left:* a satellite within R_{200m} (black dashed circle) and a field galaxy outside R_{200m} . *Center:* a stellar stream. *Right:* a phase-mixed galaxy. **Bottom:** Manual reclassification of substructures with complex morphologies. *Left:* a satellite galaxy which was classified as a stream by the algorithm due to a handful of unbound star particles. *Center:* a stream which was classified as a phase-mixed galaxy due to its large mass. *Right:* a phase-mixed galaxy which was classified as a satellite galaxy due to its small size.

which can partially mitigate the problem of these morphologically complex substructures. We will explore this in future work.

3.2. Identifying the high-redshift progenitors

We track the $z = 0$ star particles assigned to present-day substructures back to $z \gtrsim 6$ and determine their progenitor host halos from the corresponding halo catalogs. We assign a galaxy (a halo containing star particles) as a progenitor of a substructure if at least one of its star particles ends up in that substructure at the present day. We do not include the progenitors of the present-day central MW/M31-mass galaxies. We consider a progenitor galaxy to be a high-redshift contributor to the proto-MW/LG if it has at least one star particle that ends up within 2 Mpc from the center of the host galaxy (or the geometric center of mass of the paired hosts) at $z = 0$.

Star particles formed in the same progenitor galaxy at high redshift can end up in different galaxies at the present day. For example, a large fraction of the star particles from a progenitor galaxy may end up in a particular stream, while a smaller fraction may end up in a second stream, usually due to a close dynamical interaction during its evolution. In such cases, if a progenitor contributes stars to two $z = 0$ objects with the same classification, it is counted once. If a single progenitor contributes to multiple objects with different present day classifications (e.g., a stream and a phase-mixed galaxy), then the progenitor is counted as a predecessor of each object. However, when we count progenitors among multiple different $z = 0$ categories (e.g., stream *and* phase-mixed galaxies), we count each progenitor only once to avoid double-counting.

A small number of star particles in the $z = 0$ substructures exist at $z \gtrsim 6$ but are not associated with

any progenitor galaxy in the halo catalogs. Since this is a small fraction of the total stellar mass, we ignore these star particles when we assign progenitors. There are also a small number of $z = 0$ galaxies containing star particles that can be traced back to $z \gtrsim 6$ but are not assigned to any galaxy at this epoch. These are mainly the lowest-mass galaxies close to the limit of our resolution. Like Gandhi et al. (2024), we assert that each of these galaxies has one progenitor at that redshift, and assign all the traceable stellar mass to this single progenitor.

4. STELLAR MASS DISTRIBUTION OF PROGENITOR GALAXIES AT HIGH REDSHIFTS

Having assigned progenitor galaxies to present-day substructures, we next analyze the contributions of each classification to the total stellar mass distribution of the proto-MW/LG during the reionization epoch. We then reconstruct the high- z SMFs/UVLFs and determine the effect on their amplitude and shape of including the present-day disrupted galaxies.

4.1. Stellar mass budget during the EoR

The amount of stellar mass contained by the proto-MW/LG and formed before $z \sim 6$, often referred to as the “stellar mass budget” at the EoR, controls the number of photons available to reionize the Universe. Thus, we first explore how the total proto-MW/LG stellar mass is distributed among the progenitors of each present-day classification.

Figure 2 shows the total progenitor stellar mass budget at $z = 6 - 9$ from each classification, as well as their fractional contributions to the stellar mass of the proto-MW/LG. We show the results from the isolated (MW-like) and paired (LG-like) galaxy suites in the left and right panels, respectively, since the number of galaxies is different in each suite. Each line shows the median stellar mass budget across all of the simulations in that particular suite, while the shaded region shows the 68% halo-to-halo scatter. The black line indicates the total stellar mass from all of the progenitors (proto-MW/LG); the red, blue, green, and yellow lines show the contributions from the progenitors of present-day satellite galaxies, stellar streams, phase-mixed galaxies, and field galaxies, respectively. We define the proto-MW by including the progenitors of the present-day host galaxy in addition to the classified substructures (up to 2 Mpc from the center of the host galaxy). For the proto-LG, we include the progenitors of both hosts along with the classified substructures (up to 2 Mpc from the geometric center of mass of the two hosts).

In the isolated suite, the progenitors of phase-mixed galaxies contribute the most to the total stellar mass

budget, followed by the stellar stream progenitors. In the paired suite, progenitors of phase-mixed galaxies again contribute the most to the total stellar mass budget, followed by the progenitors of field galaxies, streams, and satellites, respectively. Though the stream progenitors contribute less than that of the field galaxies in the paired suite, in both cases, the progenitors of the present-day disrupted galaxies yield more stellar mass during the EoR than the progenitors of intact low-mass galaxies. Even though the stream progenitors contribute less than phase-mixed galaxies, they play a larger role than intact low-mass galaxies. The fractional stellar mass contribution from progenitors of each classification, relative to progenitors of the main galaxy, is roughly constant from $z = 6 - 9$, with disrupted galaxies contributing at least as much as intact satellites across this entire range (Figure 2, bottom panels). There is no strong mass-dependent bias across different present-day substructures. This suggests that neglecting the progenitors of streams and phase-mixed galaxies can underestimate the reionization-era ionizing photon budget derived from the near-far technique. However, even with adding the disrupted galaxies we can recover only up to $\sim 50\%$ of the total budget.

The field galaxy progenitors contribute more to the stellar mass budget than the stream progenitors in the paired suite because of our present-day distance selection criterion (2 Mpc). In the paired simulations, the cumulative number of low-mass galaxies beyond 1 Mpc continues to increase faster than in the isolated suite (Gandhi et al., in prep), leading to more field galaxies (Table 1) and hence a greater contribution from field galaxy progenitors at high redshift. If we instead use a distance cutoff of 1 Mpc, we find consistent behavior between the isolated and paired galaxy suites.

4.2. Hierarchical structure formation at high redshifts

The near-far technique assumes that the stellar mass inferred from a present-day fossil record came from a single progenitor at all redshifts. In Gandhi et al. (2024) and in this work, we similarly define the *fossil record* by combining the total stellar mass from all the progenitors of each present-day structure into one. We also consider instead assigning each present-day galaxy to its main progenitor, defined to be the progenitor with the highest stellar mass. We discuss the main progenitor only in this section, for purposes of illustration, and proceed with the fossil record assumption in the rest of the paper when assuming a single progenitor.

The viability of the fossil record assumption is important because the reconstructed SFHs alone do not directly supply the complete merger histories of present-

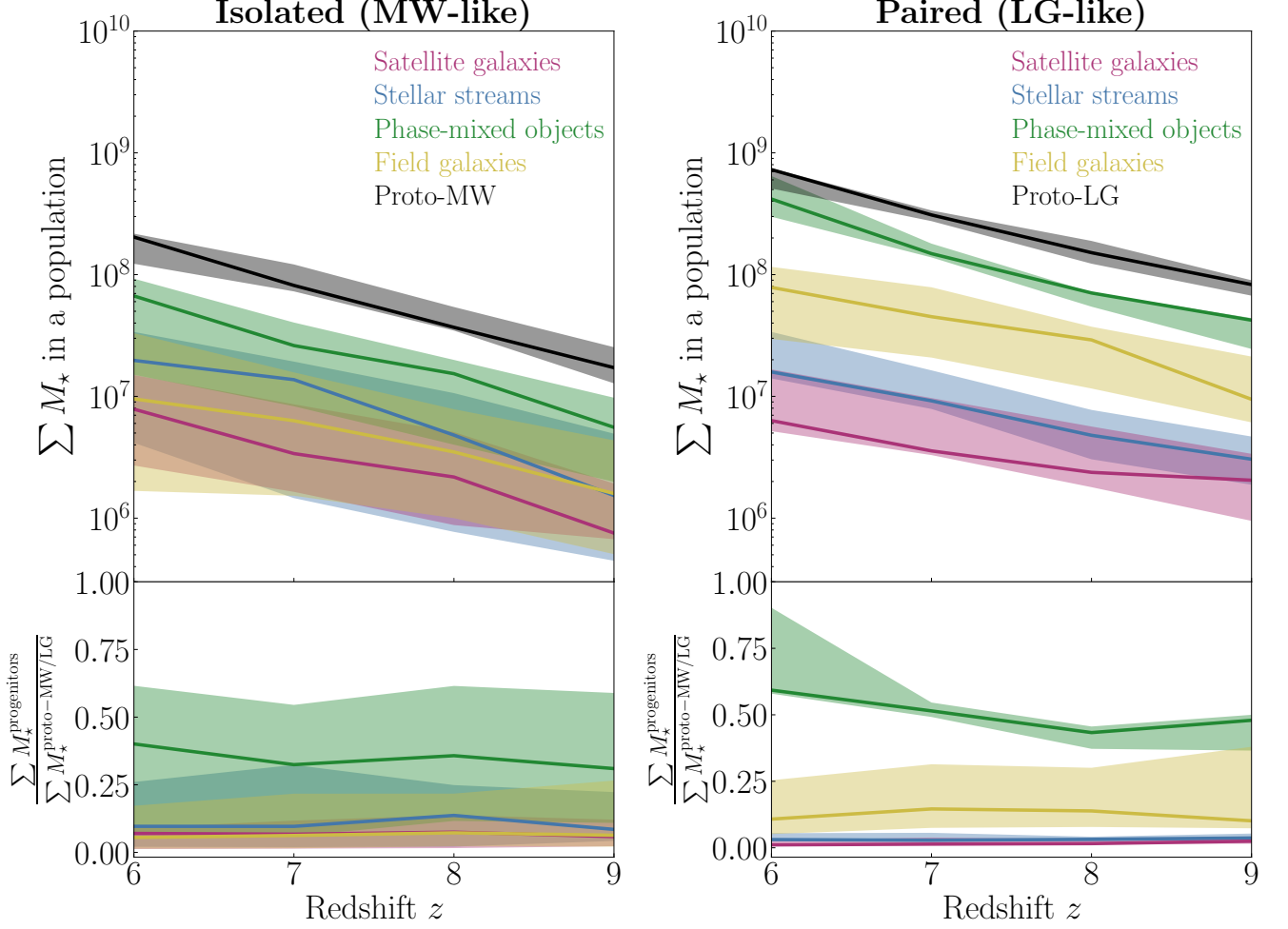


Figure 2. Streams and phase-mixed galaxies contribute more to the stellar mass budget at the EoR than intact low-mass galaxies. In all panels, solid lines show the median across the simulations and shaded regions show the 68% halo-to-halo scatter, for the isolated (left) and the paired (right) galaxy suites. The red, blue, green, and yellow lines show the total stellar mass of the progenitors of satellite galaxies, stellar streams, phase-mixed galaxies, and field galaxies, respectively. The black line shows the stellar mass of the proto-MW/LG, defined here as the total stellar mass budget from all progenitors, including those associated to the present-day main host galaxy (or host galaxies, in the paired case). **Top:** total stellar mass at EoR for each type of substructure. **Bottom:** fractional contribution to the stellar mass from each substructure type (i.e., the ratio of the colored lines in the top panel to the solid black line).

day galaxies. However, structure formation proceeds hierarchically (at least in cold dark matter models): present-day substructures were generally built up through mergers, so their stellar mass at high redshift may be distributed across multiple progenitors. Thus, the fossil record assumption could potentially bias inferences of the high- z SMFs/UVLFs.

Gandhi et al. (2024) explored the effects of mergers and disruptions for present-day surviving low-mass galaxies by quantifying their progenitor counts during the reionization era. They found that the lowest mass simulated galaxies ($M_* \sim 10^5 M_\odot$) have ~ 1 – 2 progenitors at $z \gtrsim 6$, while the progenitor number increases

with present-day stellar mass, with the most massive galaxies ($M_* \sim 10^9 M_\odot$) having ~ 15 – 40 progenitors. Furthermore, they showed that the fractional contribution of the main progenitor galaxy relative to the total stellar mass from all progenitor galaxies at that redshift, is $\gtrsim 50\%$ dominant across a wide range of present-day stellar masses ($M_* \sim 10^{4.5-8} M_\odot$). The present-day lowest-mass galaxies have an order unity contribution from a single main progenitor since they have at most ~ 1 – 2 progenitors. On the other hand, high-mass present-day substructures are typically assembled from many smaller units, and the main progenitor makes a smaller fractional contribution to the

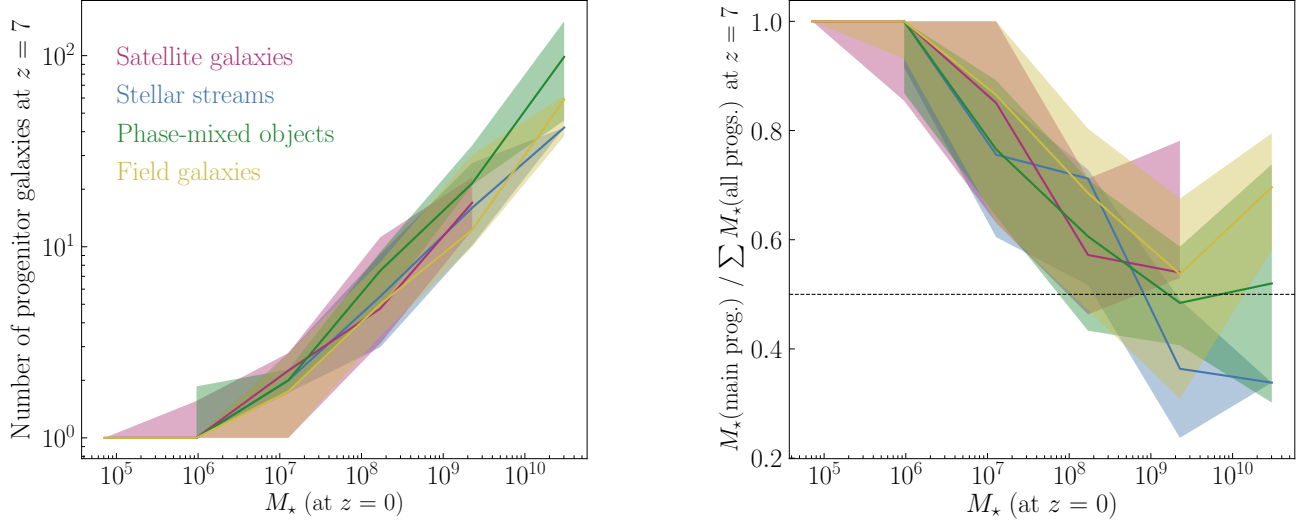


Figure 3. Hierarchical assembly at high redshifts is similar for all substructure types. In both panels, solid lines show the median across all simulations (both isolated and paired) at $z = 7$; shaded regions show 68% halo-to-halo scatter. **Left:** number of progenitors of satellite galaxies (red), stellar streams (blue), phase-mixed galaxies (green), and field galaxies (yellow) as a function of present-day stellar mass. The number of progenitors at high redshifts increases with $z = 0$ stellar mass. **Right:** fractional contribution of the main progenitor to total progenitor stellar mass. The main progenitor dominates the fractional contribution for masses between $M_* = 10^{4.5-9} M_\odot$, with $\sim 100\%$ contributions at $M_* \sim 10^{5-6} M_\odot$.

total progenitor stellar mass at high redshift. Nevertheless, the main progenitor of even the most massive galaxies contributes $\sim 50\%$ to the total progenitor stellar mass. Ultimately, [Gandhi et al. \(2024\)](#) concluded that the assumption of a single progenitor in case of the surviving low-mass galaxy fossil record still allows an accurate recovery of the low-mass/faint-end slope of the high- z proto-MW/LG SMFs/UVLFs.

Stellar streams and phase-mixed galaxies have different dynamical and formation histories than intact low-mass galaxies ([Panithanpaisal et al. 2021](#); [Cunningham et al. 2022](#)). Thus, it is not clear that the fossil record assumption will extend trivially to these other classifications even if it works for intact galaxies. To confirm that disrupted galaxies also follow the fossil record, we examine the hierarchical formation histories of the different types of present-day substructures back to $z = 7$ (Figure 3). We consider both the number of progenitors at $z = 7$ as a function of their present-day stellar mass (left panel), and the fractional contribution of the main progenitor (the ratio of the main progenitor’s stellar mass to the total mass from all progenitors evaluated at $z = 7$; right panel), as a function of present-day stellar mass. Since these relations are similar for the isolated and paired simulations, we show the median trend over all 13 halos, with the shaded regions denoting the 68% halo-to-halo scatter.

The progenitors of stellar streams (blue) and phase-mixed galaxies (green) follow a pattern of hierarchical

assembly that is indistinguishable from that of satellite (red) and field galaxies (yellow). As for intact galaxies, the least massive streams and phase-mixed galaxies ($M_* \sim 10^{6-7} M_\odot$ at the present day) have ~ 1 – 2 progenitors at $z = 7$; while the most massive ones ($M_* \sim 10^{9-10} M_\odot$) have ~ 30 – 70 progenitors. The slope of these curves is ≈ 0.38 for all classifications. The main progenitors of the present-day streams and phase-mixed galaxies (right panel) are also the greatest contributors to the stellar mass budget at $z = 7$ across a wide range of present-day stellar masses ($M_* = 10^{4.5-9}$). Although not shown here, we find similar behavior at $z = 6, 8$, and 9 . Thus, the assumption of a single main progenitor appears as accurate for the streams and phase-mixed galaxies as it is for the intact low-mass galaxies.

5. STELLAR MASS FUNCTIONS AND UV LUMINOSITY FUNCTIONS AT HIGH-REDSHIFTS

We further test the fossil record assumption by reconstructing the high- z SMFs/UVLFs including disrupted galaxies. We examine how well the SMFs/UVLFs of the proto-MW/LG (and by extension the universal SMF/UVLF) at $z \gtrsim 6$ can be recovered from each present-day classification category, as well as from several different combinations of intact and disrupted galaxies. For now, our analysis assumes that the SFHs of these substructures can be precisely reconstructed from their CMDs at $z = 0$. We compute the reconstructed

SMFs/UVLFs in both the isolated and paired suite, and compare with the SMF/UVLF from FIREbox^{HR}, which we treat as a benchmark for the universal SMF/UVLF. We do this for both the differential and cumulative functions.

We determine the SMFs of high- z progenitor galaxies by counting the number of halos in logarithmic mass bins. In order to account for the increasing number of halos at small masses and to reduce the noise at the high-mass end, we adopt variable bin widths: 0.25 dex over [4.5, 5.5], 0.5 dex over [5.5, 6.75], and 0.75 dex over [6.75, 9.0]. This adaptive binning scheme ensures adequate sampling across the full dynamic range in mass: these bin widths capture the shape of the SMFs, without excessive noise. We then divide by a volume factor, which we calculate by constructing a convex hull using all the halos at that redshift, including non-luminous ones, to avoid selection bias and account for voids to the extent permitted by simulation volume. This volume is distinct from the periodic box size used to calculate the FIREbox^{HR} SMFs/UVLFs. Since FIREbox^{HR} is a non-zoomed cosmological simulation with periodic boundary conditions, it has a well-defined total volume; for zoom-in simulations, which do not have periodic boundary conditions, the convex hull is the closest equivalent (Rose et al. 2025).

The main goal of the near-far technique is to probe the faint-end of the UVLF during reionization. Hence, in addition to the SMFs, we also assess the impact of the progenitors of present-day disrupted galaxies on the reconstructed high- z UVLF. Following the approach in Sun et al. (2023a), we compute the rest-frame UV magnitude (M_{UV}) at 1500 Å for each progenitor galaxy at high redshift. We then construct the differential and cumulative UVLFs for the same set of progenitor populations as the SMFs. We follow the same technique used to construct the SMFs, but now we count the number of progenitor galaxies in bins of UV magnitudes. Again, we adopt variable bin widths: -0.75 dex over $[-8.00, -10.25]$, -1.50 dex over $[-10.25, -14.75]$, and -2.00 dex over $[-14.75, -21.75]$.

These procedures result in reconstructed high- z SMFs/UVLFs for both the *all progenitors* and *fossil record* scenarios. We compare them to universal benchmark SMFs/UVLFs in sections 5.2 and 5.3.

5.1. Reconstructed differential SMFs and UVLFs for each present-day classification

Figure 4 compares the differential SMFs (top panel) and UVLFs (bottom panel) of the proto-MW/LG (black) at $z = 7$ with the ones constructed from the progenitors of present-day satellite galaxies (red), stel-

lar streams (blue), phase-mixed galaxies (green), and field galaxies (yellow). The solid lines show reconstructions from the full set of progenitors and the dashed lines show reconstructions obtained using a simple model for the detectability of present-day substructures with the LSST (section 8). We show results for the isolated suite and the paired suite in the left and right panels, respectively. The comparison between the two panels illustrates how effective the near-far technique is when applied to each of the MW and the LG alone. The curves show the median across all of the simulations in the respective galaxy suite. To highlight differences between classifications, the SMFs/UVLFs are computed in the *all progenitors* scenario rather than the *fossil record* scenario.

The shape of the SMFs/UVLFs in the low-mass/faint-end ($M_{\star} \lesssim 10^6 M_{\odot}$ or $M_{UV} \gtrsim -13$) reconstructed from the progenitors of present-day stellar streams and phase-mixed galaxies is qualitatively similar in representing in the proto-MW/LG SMFs/UVLFs, when compared to the SMFs/UVLFs from just the progenitors of intact galaxies. Furthermore, the disrupted galaxy progenitors recovers more of the normalization of the SMFs/UVLFs at least in comparison to those from intact present-day satellite galaxies. This underlines the need to include present-day disrupted galaxies in near-far analysis. The SMFs/UVLFs from the disrupted galaxies tend to extend to more massive/brighter galaxies, because more of these systems have massive progenitors at high redshifts than in the intact galaxy population. The larger normalization of the SMFs/UVLFs from these disrupted systems, compared to that recovered from the low-mass satellites alone, is partly because some of these massive systems have multiple high- z progenitors. The overall shapes do not change drastically even when detectability constraints are imposed, suggesting that it may be feasible to include disrupted galaxies in future near-far technique analyses. The SMFs and UVLFs reconstructed from the stream progenitors remain almost unchanged, because the detectable streams are the most massive ones with the largest number progenitors, comprising a large portion of the distribution at high redshifts.

5.2. Reconstructed differential SMFs and UVLFs using the full merger histories of present-day substructures

In this section, we explore how accurately we can recover the SMFs/UVLFs of the proto-MW/LG at $z \gtrsim 6$ by adding the progenitors of present-day disrupted galaxies along with the intact galaxies in the *all progen-*

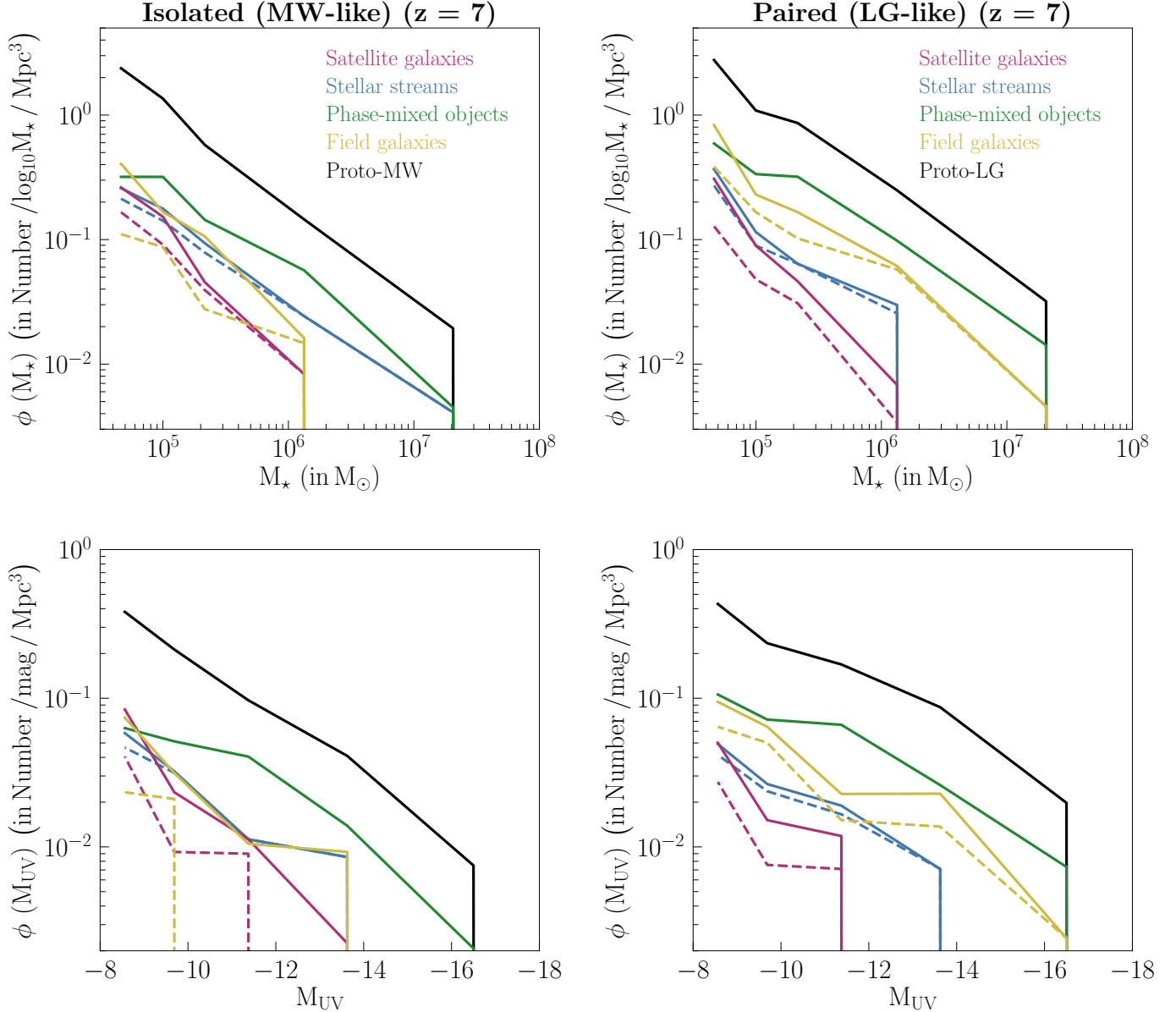


Figure 4. Present-day disrupted and intact galaxies recover different differential SMFs and UVLFs at $z = 7$. All panels show median reconstructed SMFs (top) and UVLFs (bottom) for the progenitors of various present-day substructure types, without (solid curves) and with (dashed curves) detectability constraints (Section 8). The isolated galaxy suite is shown in the left column; paired galaxies in the right column. The proto-MW/LG (black), which includes the progenitors of the present-day main host(s), is compared with reconstructions from the progenitors of satellite galaxies (red), stellar streams (blue), phase-mixed galaxies (green), and field galaxies (yellow).

itors scenario. Here we divide our present-day substructures into three samples:

Sample I: Intact galaxies only (satellite and field galaxies)

Sample II: Intact galaxies and stellar streams

Sample III: Intact galaxies, stellar streams, and phase-mixed galaxies.

Sample II is likely more attainable than Sample III, since stellar streams usually require less observational data to identify and reconstruct than phase-mixed galaxies.

Figures 5 and 6 (top panels) compare the differential SMFs and UVLFs at $z = 7$ for Samples I (red), II (blue), and III (green) to those of the proto-MW/LG (black) and the universal SMFs/UVLFs from FIREbox^{HR} (yellow). As discussed in section 5.1, we show the isolated suite (left) and the paired suite (right) separately. Solid

All progenitors scenario

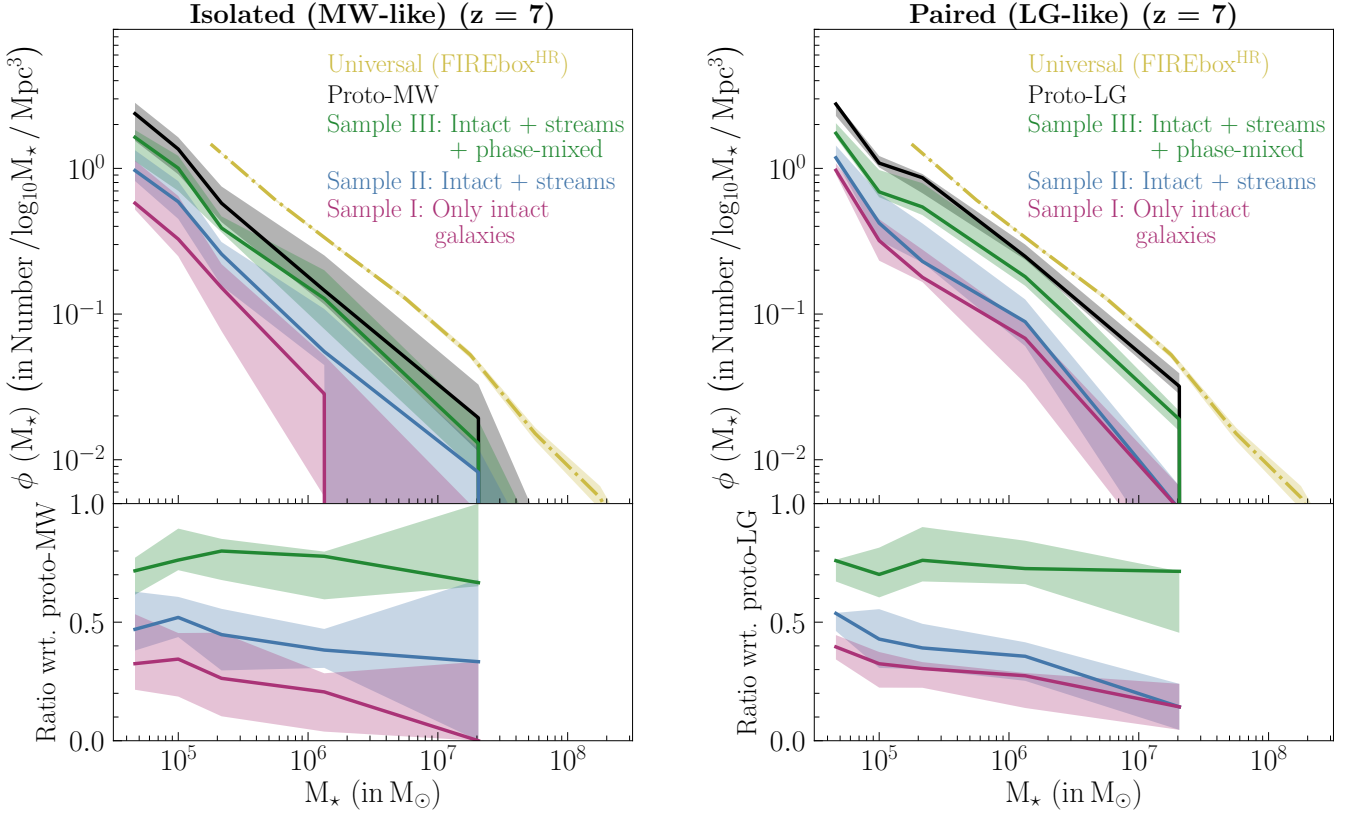


Figure 5. Differential SMFs at $z = 7$ in the *all progenitors* scenario. In all the panels, red, blue, and green solid lines show the median across all the simulations for sample I, II, and III, respectively in the isolated (left) and paired (right) galaxy suites. The shaded regions show the corresponding 68% halo-to-halo scatter. In this scenario we measure all progenitors separately, instead of treating all progenitors as a single galaxy, as in the case of the fossil record. **Top:** reconstructed differential SMFs at $z = 7$ from all of the progenitors in each present-day sample, compared to the proto-MW/LG (black; includes the present-day main host(s)) and the universal differential SMF from FIREbox^{HR} (dashed yellow). **Bottom:** ratio of the SMF from each sample to the SMF of the proto-MW/LG. Reconstructed SMFs from sample II and III represent both the shape and normalization of the proto-MW/LG SMF better than sample I at all stellar masses, and generally appears to have smaller halo-to-halo scatter with some exceptions (see text for details).

lines show median SMFs/UVLFs from all of the simulations in the respective suite. Shaded regions show the 68% halo-to-halo scatter except for the universal FIREbox^{HR} curve (yellow); here the shading denotes the 68% confidence interval (CI) obtained via bootstrapping (Feldmann et al. 2025). We also compare the ratio of the SMF/UVLF from each sample with respect to the SMF/UVLF of the proto-MW/LG (Figures 5 and 6, bottom panels) to help visualize their shapes as a function of stellar mass/UV magnitude. We compute the ratios for each simulation first and then take the median and 68% CI.

The proto-MW/LG SMFs/UVLFs have shapes similar to the universal SMF/UVLF from FIREbox^{HR}, lending confidence to the assumption that the proto-MW/LG provides a representative sample of the global average SMF/UVLF during reionization. However, the

amplitude of the proto-MW/LG SMF/UVLF is lower than that of FIREbox^{HR}, particularly in the isolated suite. This could be attributed to the difference in the calculation of the volume that goes into the denominator while computing the SMFs/UVLFs (see section 5). The stellar masses and UV magnitudes in FIREbox^{HR} are also computed differently than in this work (see Feldmann et al. (2025) for details). Additionally, the FIREbox^{HR} box has a higher mean matter density (exactly matching the mean density of the Universe) across the simulation volume than the MW-like zoomed simulations at high redshifts, which could also lead to this difference in amplitude. This discrepancy is smaller in the case of the paired LG-like suite because these simulations have a higher mean matter density than the isolated MW-like simulations, and the galaxies form stars at earlier times in the paired suite, leading to more star-

All progenitors scenario

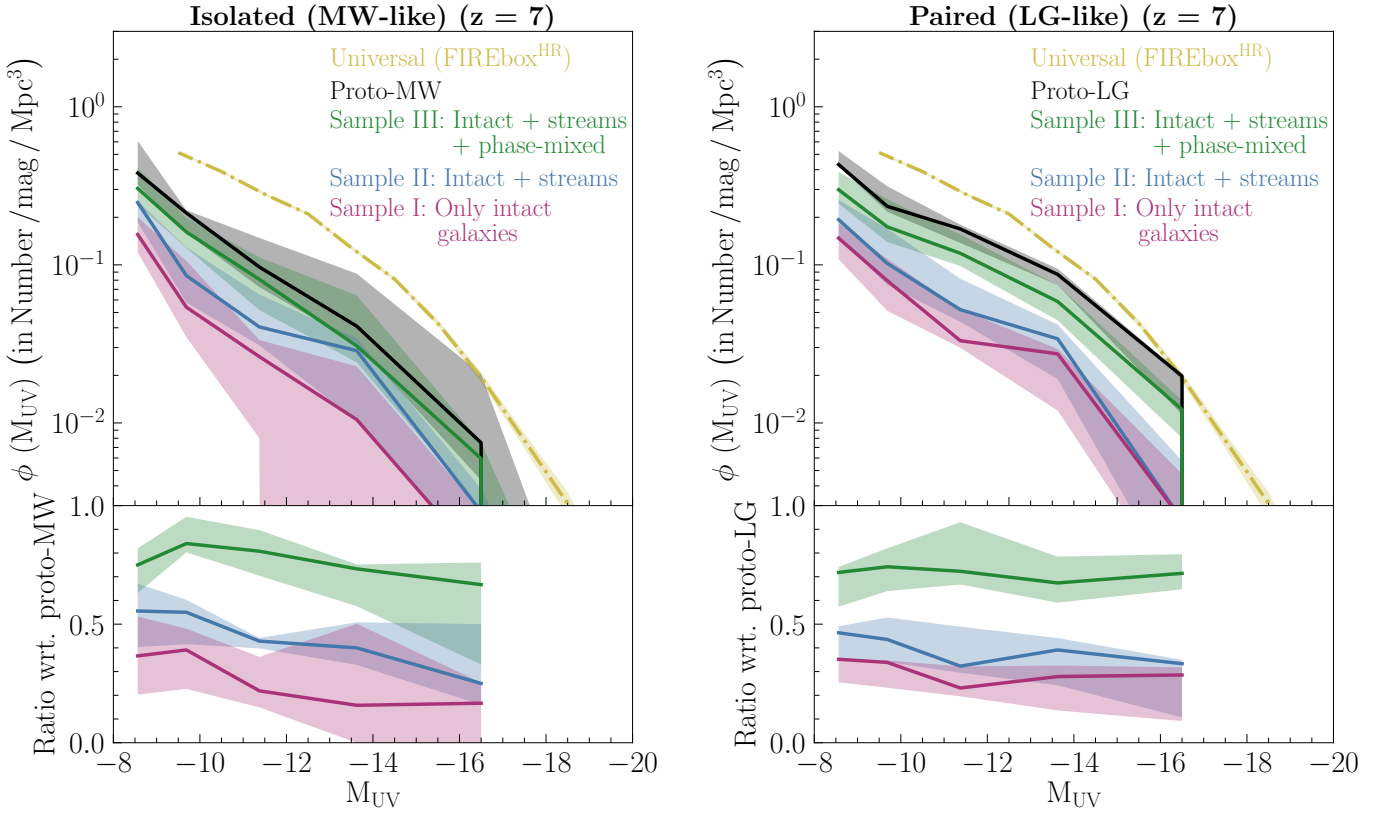


Figure 6. Differential UVLFs at $z = 7$ in the *all progenitors* scenario. Same as Figure 5, but showing the differential UVLFs. As for the SMFs, including present-day streams (Sample II, blue) and phase-mixed galaxies (Sample III, green) improves the shape and normalization of the reconstructed UVLF compared to intact dwarf galaxies alone (Sample I, magenta).

forming galaxies at high redshifts. We plan to explore this difference further in future work.

The overall shapes of the reconstructed SMFs/UVLFs from the three samples are qualitatively similar to that of the proto-MW/LG. Adding the progenitors of the stellar streams recovers the SMF/UVLF of the proto-MW/LG more accurately, particularly in terms of normalization, up to $\sim 50\%$. Adding the progenitors of the phase-mixed galaxies improves the reconstruction even more. Including the disrupted systems helps most in recovering intermediate- and high-mass progenitors that are missed when using only intact low-mass galaxies. Furthermore, the halo-to-halo scatter in samples II and III generally appears lower than sample I, with the exception of bottom right panels in Figures 5 and 6. But there the scatter is from just 3 paired halos and hence may not be robust. This is even more apparent when the low-mass/faint-end slope is estimated (Section 6.1). Adding more disrupted galaxies to our present-day samples leads to more robust recovery of our proto-MW/LG SMFs/UVLFs by potentially reducing the halo-to-halo scatter, which is sensitive to the variations in accretion

histories. This is particularly prominent in the case of the isolated suite, since it probes a wider range of accretion histories than the paired suite, leading to more scatter in the case of Sample I.

The ratios of the reconstructions from each sample to the proto MW/LG are nearly flat over a wide range of stellar masses/magnitudes. The scatter is larger in the high-mass end because of low-number statistics. Samples I and II have a slight bias at lower (fainter) masses (magnitudes), but remain almost flat at the low-mass/faint-end ($M_\star \lesssim 10^{6.5} M_\odot$ or $M_{UV} \gtrsim -14$). Sample II recovers $\sim 50\%$ of the proto-MW/LG SMF/UVLF while sample I recovers only $\sim 25\%$, a factor of ~ 2 improvement in the normalization. Sample III does even better, recovering $\sim 75\%$ of the proto-MW/LG with reduced bias. These results argue that the progenitors of the stellar streams and phase-mixed galaxies are crucial for a more accurate and precise representation of the proto-MW/LG and the universal SMFs/UVLFs at high redshifts.

5.3. Reconstructed differential SMFs and UVLFs from the fossil record of present-day substructures

Fossil record scenario

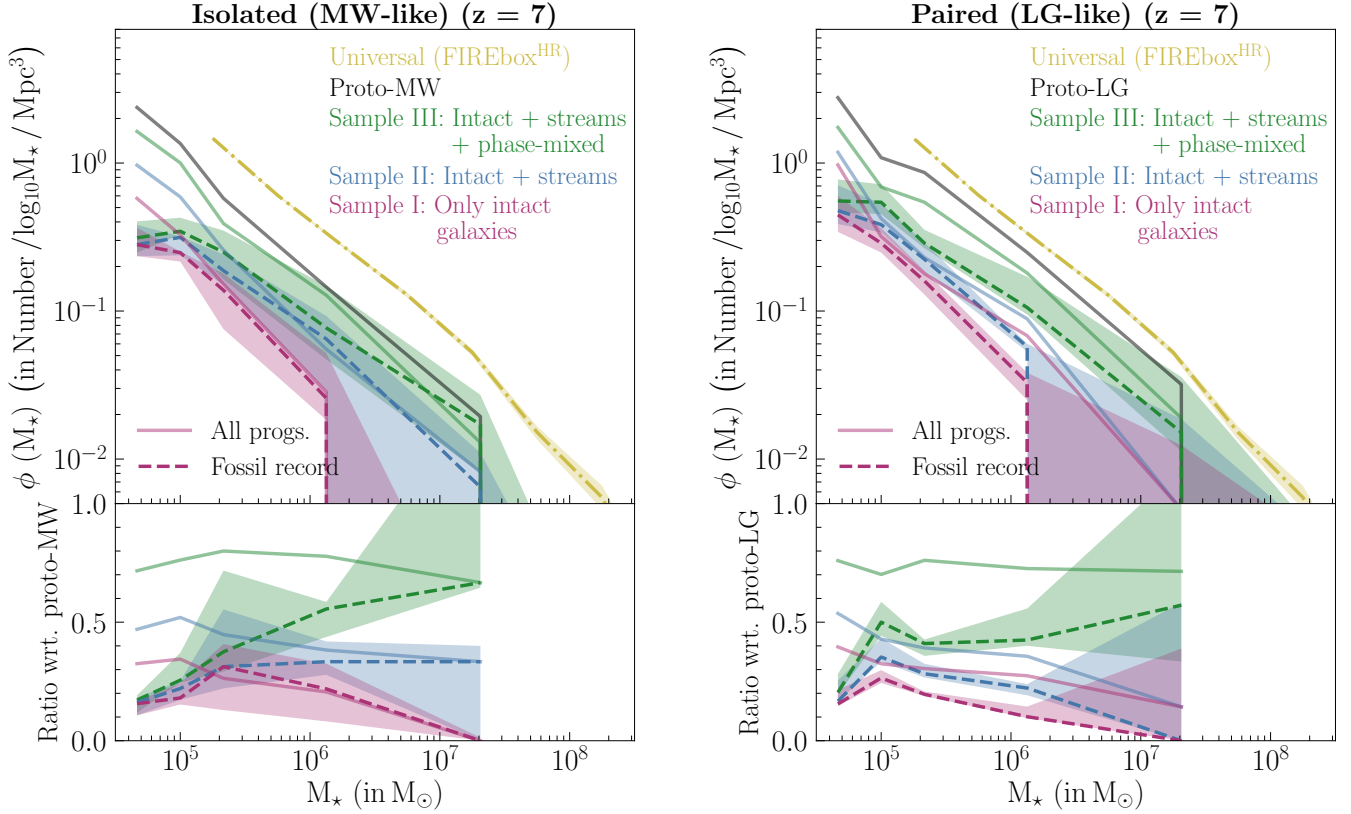


Figure 7. Differential SMFs at $z = 7$ in the *fossil record* scenario. Similar to Figure 5, but for the SMFs in the *fossil record* scenario (curves with shading). This scenario assumes all the stellar mass from multiple progenitors in enclosed in a single one, as we would infer observationally. The SMFs and ratios for the *all progenitors* scenario shown in Figure 5 are included for comparison (solid curves with no shading). Unlike the *all progenitors* scenario, the *fossil record* scenario has a mass-dependent bias: the reconstruction is less complete for lower-mass galaxies than for higher-mass galaxies. The effect is most prominent in sample III (green), followed by samples II (blue) and I (magenta) respectively.

We next reconstruct the SMFs from the fossil record, by combining the stellar masses of all the progenitors of each galaxy into a single progenitor (Figure 7). We use this combined stellar mass to calculate UV magnitudes and reconstruct the UVLFs (Figure 8). The color scheme in these figures remains the same as in Figure 5, but the dashed lines now show the median SMFs/UVLFs from the fossil record, and the shaded regions show the 68% halo-to-halo scatter in the fossil record. We show the SMFs/UVLFs from the *all progenitors* scenario as solid lines for reference but (for the sake of readability) do not show their halo-to-halo scatter.

Using the fossil record instead of the complete merger history has several effects on reconstructing the $z = 7$ SMFs/UVLFs. First, the overall normalization is reduced because there are fewer progenitors in the *fossil record* scenario. Second, the SMFs/UVLFs from the *fossil record* scenario in Figures 7 and 8 (top panel) and their ratios with respect to that of the proto-MW/LG (bottom panel) decrease towards the low-mass/faint-

end because many of the low-mass/faint progenitors are combined into a single more massive progenitor in the fossil record assumption. Since the present-day disrupted galaxies can be more massive than intact galaxies, and thus can have multiple progenitors at high redshifts, the fossil record approach leads to larger differences in shape towards the low-mass/faint-end when they are included. Furthermore, the number of massive systems with multiple high- z progenitors also increases when we add more types of present-day substructures. Hence the SMFs/UVLFs of sample III, which contains the most substructures in general as well as the largest number of disrupted galaxies, show the largest difference, followed by sample II and I, respectively.

Except for the case of the sample III SMF in the isolated suite, the ratios in the *fossil record* scenario remain nearly flat between $10^5 M_\odot \lesssim M_* \lesssim 10^{6.5} M_\odot$ for the other SMFs. In other words, the bias is not mass dependent over this range. The UVLFs between $-14 \lesssim M_{UV} \lesssim -10$ also remain nearly flat. However,

Fossil record scenario

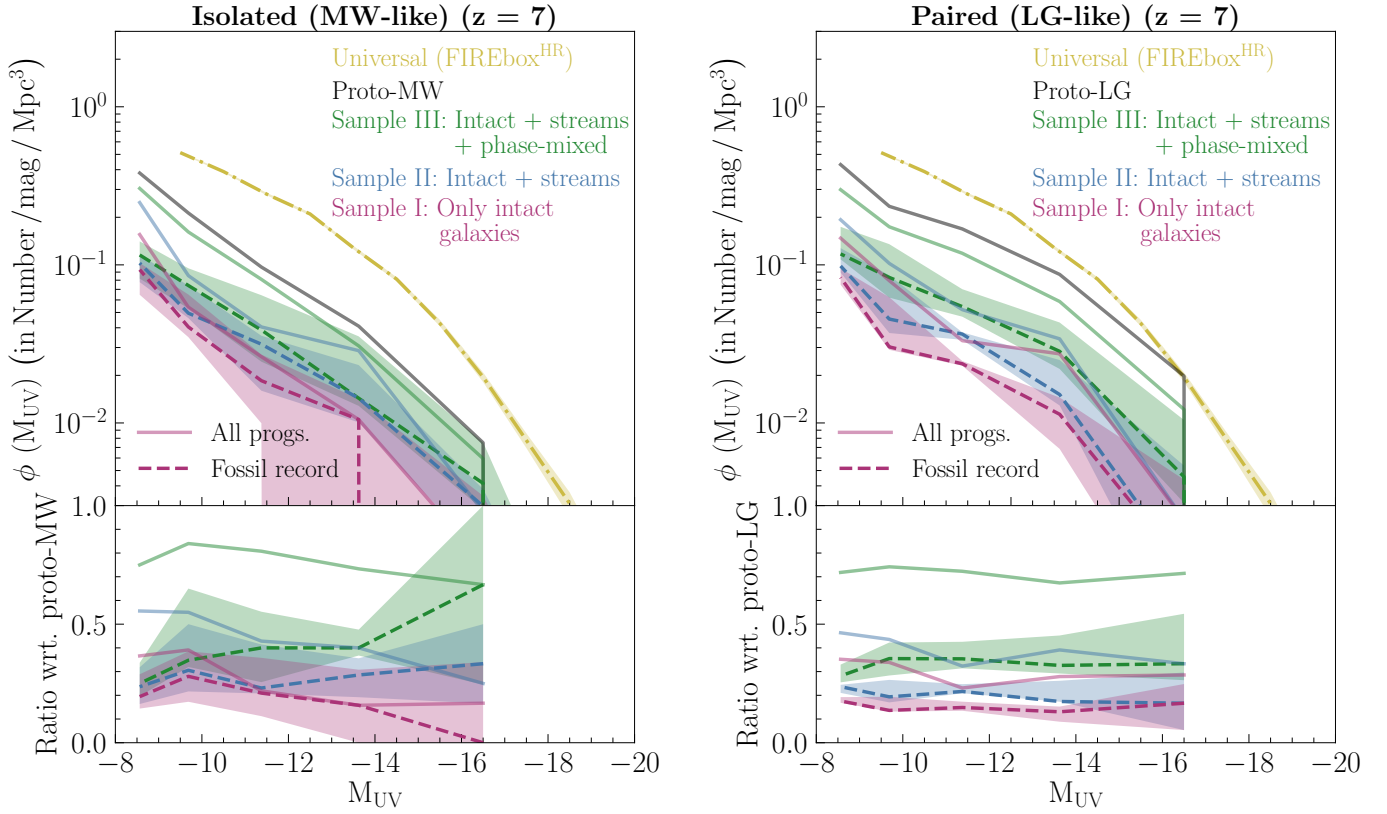


Figure 8. Differential UVLFs at $z = 7$ in the *fossil record* scenario. Similar to Figure 7, but for the differential UVLFs. The mass-dependence of the reconstruction noted in Figure 7 for the SMFs is less pronounced for the UVLFs, especially in the LG-like systems.

over the entire mass/UV magnitude range, the ratios decrease towards the low-mass/faint-end. This occurs because the *fossil record* scenario combines multiple progenitors into a single effective progenitor, which reduces the number of low-mass/faint-end progenitors. The difference is exacerbated for sample II and sample III which have better completeness at higher stellar mass. We also note that the relative shape of the SMFs/UVLFs from the three samples do not change significantly in the case of the proto-LG, particularly for the UVLFs. These results do not rule out the possibility of using disrupted galaxies along with intact galaxies in the near-far technique, but rather indicate that our inferences are sensitive to the selection criteria for *both* the present-day substructures and their high- z progenitors, along with the effects of the environment they reside in. They are also sensitive to whether we use the SMF or the UVLF. We explore this selection-dependence more quantitatively in the next section.

6. SHAPES OF THE SMFS AND UVLFs DURING THE EOR

6.1. Low-mass/faint-end slopes

Since a primary goal of the near-far technique is to determine the low-mass end slope of the SMF at $z \gtrsim 6$, we next consider the effect of including the progenitors of the disrupted galaxies on the recovery of this slope. [Gandhi et al. \(2024\)](#) showed that the low-mass-end slope can be accurately recovered from the fossil record progenitors of just the intact low-mass galaxies, without taking disrupted galaxies into account. They compute the low-mass-end slope by fitting a single power-law to the SMF across a mass range of $M_\star = 10^{4.5-6.5} M_\odot$ and report the median slope across all of the simulations, concluding that the slope inferred from the fossil record of just the intact low-mass galaxies agrees with that of the proto-MW/LG. We test this result by calculating the slopes of the reconstructed high- z SMFs/UVLFs for our three samples and the proto-MW/LG, but fit a Schechter function to incorporate massive galaxies as well ([Schechter 1976](#)). This is well-motivated by observations ([Duncan et al. 2014](#); [Bouwens et al. 2015](#); [Finkelstein et al. 2015](#); [Song et al. 2016](#); [Atek et al. 2018](#); [Ishigaki et al. 2018](#); [Bhatawdekar et al. 2019](#); [Bouwens et al.](#)

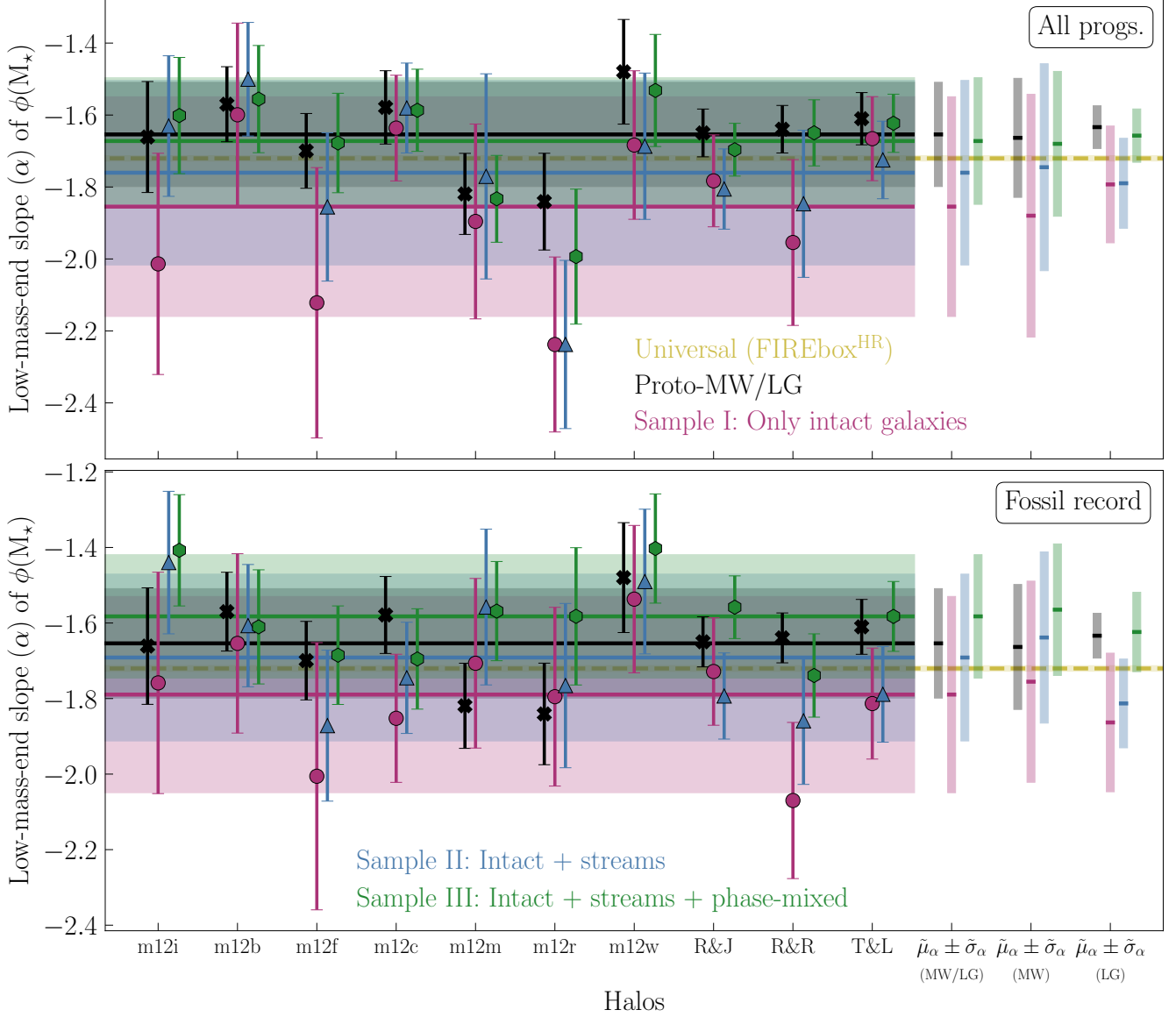


Figure 9. Adding streams and phase-mixed galaxies improves the accuracy and precision of the recovered slope of the differential SMF at low masses. We compare the low-mass slopes of the differential SMFs at $z = 7$ for samples I (magenta), II (blue), and III (green) to both the full SMF of the proto-MW/LG (black) and the universal slope estimate from FIREbox^{HR} (yellow). Errorbars indicate the standard deviation on the recovered slope for each individual halo, while the horizontal lines show the median ($\tilde{\mu}_\alpha$) of the bootstrapped sample of mean α values across all the halos, computed as described in Section 6.1 and Appendix C. The shaded regions show the median ($\tilde{\sigma}_\alpha$) of the bootstrapped sample of standard deviations in α , representing the halo-to-halo scatter. We show ($\tilde{\mu}_\alpha$) and ($\tilde{\sigma}_\alpha$) in the rightmost points for proto-MW and proto-LG by combining them as well as separately. In the *all progenitors* scenario (top), sample III consistently reproduces the low-mass-end slope of the proto-MW/LG SMF most accurately (α closest to that of the proto-MW/LG for each individual halo) and precisely (smaller errorbars for each individual halo). The median slope ($\tilde{\mu}_\alpha$) for sample III is also closest to that of the proto-MW/LG and have the smallest halo-to-halo scatter ($\tilde{\sigma}_\alpha$), followed by sample II and I. In the *fossil record* scenario (bottom), samples II and III still have consistently smaller uncertainties for each individual halo but sample I is more accurate than II and III at recovering the slope in some of the halos. The conclusions remain similar when MW and LG are separately considered.

2021, 2023; Whitler et al. 2025), but alternative models (e.g., double power-law) also exist, particularly for explaining the excess number of galaxies at the bright-end ($M_{UV} \lesssim -22$) (Bowler et al. 2015, 2017, 2020; Donnan et al. 2023; Harikane et al. 2023; Sun et al. 2023b;

Harikane et al. 2024; Donnan et al. 2024; Chemerynska et al. 2025), which is beyond the brightest set of galaxies in our samples. Instead of fitting the Schechter functions to binned SMFs/UVLFs, we perform Bayesian modeling using the PyMC package (Abril-Pla et al. 2023) to esti-

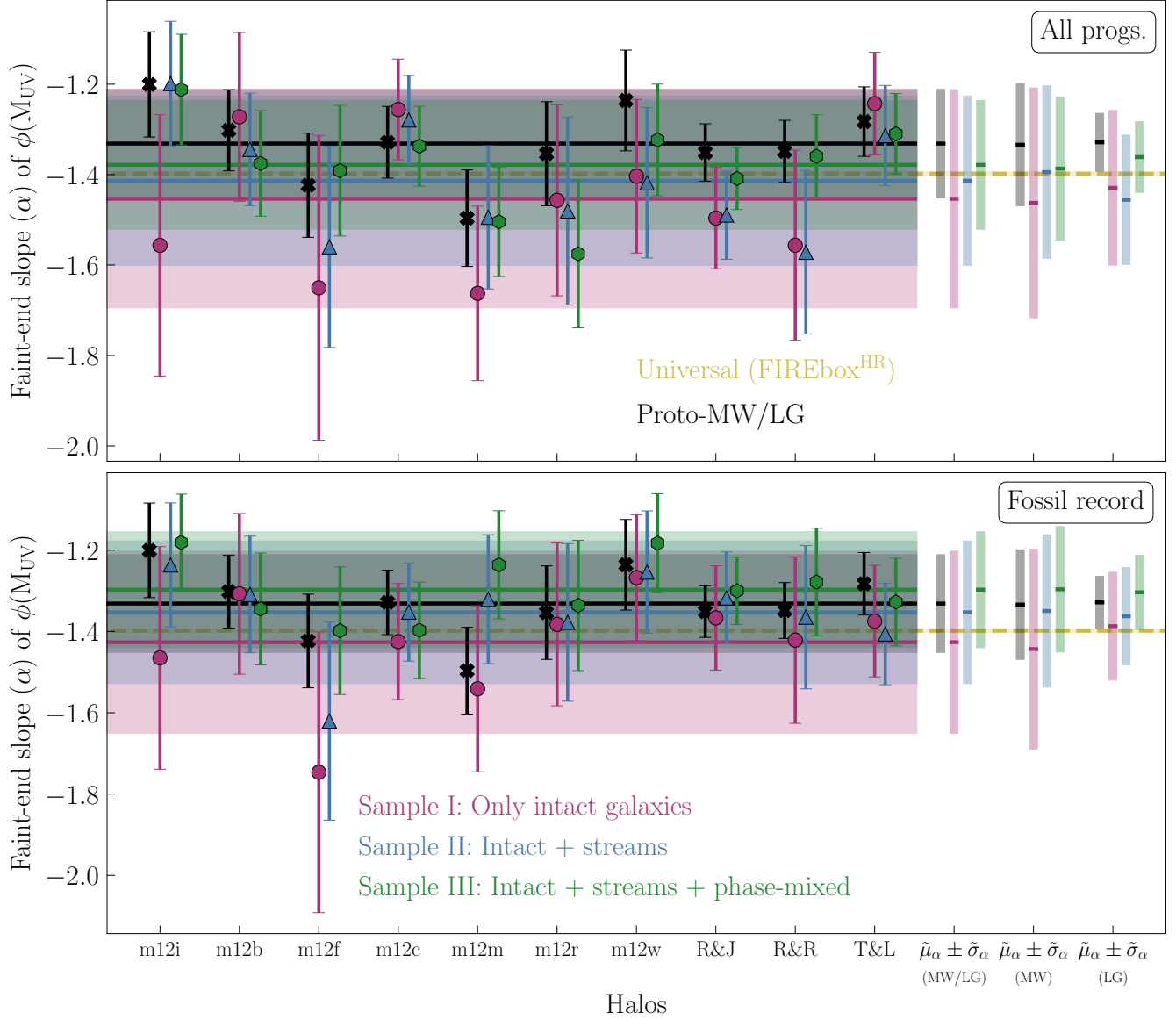


Figure 10. Adding streams and phase-mixed galaxies improves the accuracy and precision of the recovered slope of the differential UVLF at low masses Similar to Figure 9, but for the UVLFs. Overall, the UVLF slopes are shallower than the SMF slopes owing to the superlinear dependence of the UV luminosity on stellar mass (see Appendix D).

mate the posterior distribution of the slope parameter α . This approach eliminates the need for binning, which can affect the result for α , and provides a full posterior distribution for α . In the case of FIREbox^{HR} we do use the binned SMFs/UVLFs, but generate posteriors of α by fitting Schechter functions to these SMFs/UVLFs using the same approach. The details of our methodology, along with illustrative results, are described in Appendix B.

We estimate the posterior distribution of the slope α , which follows an approximately Gaussian distribution (see Appendix B), and calculate its mean and standard deviation for the three samples in each simula-

tion. Figures 9 and 10 show the slopes obtained by our fits for the SMFs and UVLFs respectively, compared to the ones from the proto-MW/LG and the universal SMFs/UVLFs at $z = 7$. The *all progenitors* scenario is shown in the top panel, while the *fossil record* scenario is in the bottom panel. The points and the error bars in each figure denote the mean and standard deviation of the posterior distribution for the low-mass/faint-end slope α for the proto-MW/LG (black), sample I (red), sample II (blue) and sample III (green) for each simulation. The mean slope of the universal SMF/UVLF is shown by the yellow line, while the shaded region shows the standard deviation. To estimate population-

level parameters, we bootstrap the individual α values, resampling from their Gaussian posteriors. This yields: (1) a central slope estimate ($\tilde{\mu}_\alpha$) - median of the bootstrapped means; (2) halo-to-halo scatter ($\tilde{\sigma}_\alpha$) - median of the bootstrapped standard deviation. We show these by the horizontal lines and the shaded regions, respectively, in Figures 9 and 10. To compare the results between proto-MW and proto-LG, we separately do the analysis for the isolated and paired galaxy suites. For clarity, we also show these estimates in the rightmost points of both the figures. We emphasize that $\tilde{\sigma}_\alpha$ is *not* the standard deviation of the bootstrapped mean α distribution, but rather an average range of possible α values across all the halos in the respective galaxy suite, defining the halo-to-halo scatter. We discuss the details of our bootstrapping method in Appendix C. Tables 2 and 3 summarizes our slope estimates. We note that the median values for the isolated and paired galaxy suites calculated separately, in general, provide similar conclusions as the median values for the 10 halos combined. However, the results for the paired galaxy suite are derived from only three simulations and thus may not be robust. Since the trend is similar we discuss our results for $\tilde{\mu}_\alpha$ and $\tilde{\sigma}_\alpha$ when all the halos are combined, unless stated otherwise.

The figures show slope estimates using progenitor galaxies with $M_\star \geq 10^5 M_\odot$ or $M_{UV} \leq -10$ (our fiducial choice). We defer a discussion of the effect of varying these cut-offs to Section 6.2. The proto-MW/LG slopes are representative of the universal slopes. There is an offset of ≈ 0.07 in the median slope values, but they agree within the median halo-to-halo scatter. This could be due to the difference in mean matter densities of the universe and the MW/LG region as discussed in Section 5.2. We will explore this aspect in future work. Since the proto-MW/LG slopes are representative of the universal values within the uncertainties, we mainly assess our results from the three samples against the proto-MW/LG for quantitative comparisons.

The low-mass end slope values of the proto-MW/LG and sample I reported here are different than the ones in Gandhi et al. (2024) by ≈ 0.1 , likely due to the difference in the slope estimation method. In addition, some of the present-day intact galaxies in Gandhi et al. (2024) are instead identified as part of our set of disrupted galaxies, so our Sample I is not identical to the set used in that paper.

We find several key features when we include the disrupted galaxy progenitors while computing the low-mass-/faint-end slopes. First, when we compare the individual low-mass/faint-end slopes from the SMFs/UVLFs of the three samples with the one from

the proto-MW/LG, we find that in most halos the slope of the proto-MW/LG SMF/UVLF is consistently better reproduced by sample III, followed by sample II and I, respectively. This is also reflected in the median values, $\tilde{\mu}_\alpha$, from our bootstrapped samples. Second, including more disrupted galaxies reduces the standard deviation of α values from sample I within each halo, by up to $\sim 30\%$ in sample II and $\sim 50\%$ in sample III. This is because the increased number of progenitors improves statistical robustness. Third, there is a scatter in the values of the slope across different halos, which implies that different accretion histories can lead to different values of the reconstructed low-mass-/faint-end slope, leading to poorly constrained values across the realizations. However, we find that this halo-to-halo scatter can be reduced by adding the disrupted galaxies, as shown by the decrease in the median $\tilde{\sigma}_\alpha$ values (rightmost points). This reduction in the scatter, in comparison to sample I, can be $\sim 20\%$ in case of sample II and $\sim 40\%$ in case of sample III. The effect of an unusual accretion history in producing a poorly constrained slope can thus be mitigated by adding the disrupted galaxies.

We reach similar conclusions when we consider the effect of adding disrupted galaxies separately to a MW-like environment and a LG-like environment. The halo-to-halo scatter in case of a LG-like environment is smaller compared to that of a MW-like environment, but the results are derived from only three simulations and may not be robust. Additionally, the low-mass-end for sample I and III is steeper than proto-LG because of the R&R paired halo (see Figure 9). Furthermore, the slopes are susceptible to how present-day and high- z galaxies are selected, which we discuss in 6.2 and hence the should be treated with caution.

Interestingly, the low-mass-end slopes of the SMFs (see Figure 9) and the faint-end slopes of the UVLFs (see Figure 10) differ. The UVLFs have a shallower slope at the faint-end (≈ -1.4) compared to the low-mass-end of the SMFs (≈ -1.7). Even though extrapolations of observed UVLFs predict a faint-end slope of ≈ -2 , results from FIREbox^{HR} also seem to be pointing to a shallower slope (see Table 2). Although this could depend on the specific distribution used to draw the progenitor masses and the slope estimation method, the difference between the SMF and UVLF slopes invites further investigation. Using the power-law nature of the SMF at the low-mass-end and the mass-to-light ratios computed from our progenitor stellar halos, we find that the shallower slope in the UVLFs, can be partially explained due to the superlinear dependence of UV luminosity with stellar mass. Since young, massive O/B stars generate most of the UV radiation and lower-

mass galaxies are less efficient at producing them, this leads to suppressed UVLFs at the faint-end. We refer the reader to Appendix D for further details and figures demonstrating our analysis. In any case, we better reproduce the proto-MW/LG SMFs/UVLFs after including present-day disrupted galaxies.

Another important difference can be discerned between the *all progenitors* scenario and the *fossil record* scenario: the latter case results in flatter slopes - by ~ 0.07 on average across the three samples when both the isolated and paired galaxy suites are combined. As discussed in Section 5.3, this effect is due to combining multiple progenitors into a single one in the *fossil record* scenario and it gets amplified for samples II and III as they have more massive substructures with multiple progenitors. The level of flattening varies depending on how both the present-day and high- z progenitor galaxies are selected.

6.2. Sensitivity to stellar mass selection criteria

In Section 5.3, we noted that the use of the fossil record seemed to increase the sensitivity of our reconstructions to the selection of galaxies used in the near-far technique. Here we explore this further, by testing how the $z = 7$ SMF slope estimates change under several variations of our selection criteria.

6.2.1. Lowering the low-mass cutoff at $z = 7$

We select progenitor galaxies with $M_\star \geq 10^{4.5} M_\odot$ instead of $M_\star \geq 10^5 M_\odot$ at $z = 7$. This increases the number of progenitor galaxies, which naturally leads to a smaller standard deviation on α in each simulation, for each of the samples. In this case, sample III produces the lowest standard deviation followed by sample II and I, respectively. However, we get steeper mean α values in the *all progenitors* scenario and flatter values in the *fossil record* scenario. This leads to a larger difference between the slopes in these two scenarios when more disrupted galaxies are added, even though the halo-to-halo scatter decreases. Our estimates using bootstrapping, shown in Table 2, demonstrate these effects. With a lower mass cutoff, Sample III provides the most accurate and precise estimate of the proto-MW/LG slope in the *all progenitors* scenario followed by sample II and I, respectively. However, in the *fossil record* scenario, sample I provides a more accurate representation of the proto-MW/LG slope than sample II and III. Thus reducing the mass cut-off incorporates more low-mass progenitors which leads to the steepening of the slope in the

all progenitors scenario. However, when combined into a single progenitor in the *fossil record* scenario, we get flatter SMFs at the low-mass end. The effect gets more drastic when we add more massive present-day substructures since we increase the amplitude more towards the high-mass-end than in the low-mass-end.

6.2.2. Increasing the present-day stellar mass cutoff for intact galaxies

We can reliably select present-day intact galaxies down to $M_\star \gtrsim 10^{4.5} M_\odot$, but disrupted galaxies only down to $M_\star \gtrsim 10^6 M_\odot$. So we test our results by also selecting intact present-day galaxies with $M_\star \gtrsim 10^6 M_\odot$. When we use progenitor galaxies with $M_\star \gtrsim 10^{4.5} M_\odot$ at high redshifts, we find similar results for the proto-MW/LG slope in the three samples, with the least halo-to-halo scatter in case of sample III in the *all progenitors* scenario. However, in the *fossil record* scenario all the three samples give a flatter slope than that of the proto-MW/LG. Upon increasing the cut-off to $M_\star \gtrsim 10^5 M_\odot$ at high redshifts, the difference between the slopes in the *all progenitors* scenario and the *fossil record* scenario reduces, as expected. In the *all progenitors* scenario the slope of sample III is closer to the proto-MW/LG slope ($\Delta\tilde{\mu}_\alpha \approx 0.006$) than sample I ($\Delta\tilde{\mu}_\alpha \approx 0.145$). However, in the *fossil record* scenario, sample I is more accurate ($\Delta\tilde{\mu}_\alpha \approx 0.007$) than sample III ($\Delta\tilde{\mu}_\alpha \approx 0.150$).

6.2.3. Removing morphologically complex and extremely massive present-day substructures

Our present-day sample includes a few morphologically-complex, massive galaxies with $M_\star \gtrsim 10^9 M_\odot$. These typically have $\sim 30 - 50$ progenitors and lead to more bias in the fossil record scenario. Hence we test our results by restricting our present-day galaxy sample to those with stellar masses between $10^6 M_\odot \lesssim M_\star \lesssim 10^9 M_\odot$. The trends in this case resemble our previous results without a high-mass cut-off.

6.2.4. Lowering the high-mass cutoff at the present day

We tried lowering the high-mass cutoff for present-day substructures to $M_\star \lesssim 10^8 M_\odot$ while keeping the low-mass cut-off at $M_\star \gtrsim 10^{4.5} M_\odot$. In this case, we find that the slopes across the three samples in the *all progenitors* scenario, although similar, are steeper than that of the proto-MW/LG by ≈ 0.23 . In the *fossil record* scenario, on the other hand, the slopes are flatter than the proto-MW/LG; with sample III being flatter (by ≈ 0.4) than sample II (by ≈ 0.3) or sample I (by ≈ 0.2). Interestingly, the halo-to-halo scatter reduces only by $\sim 3\%$ in sample II and by $\sim 15\%$ in sample III. This remains similar when selecting progenitor galaxies with $M_\star \gtrsim 10^5 M_\odot$ at high redshifts, although in this case all three samples produce

Table 2. Low-mass-end and faint-end slope estimates at $z = 7$, for isolated and paired simulations combined.

Sample	Stellar Mass Functions		UV Luminosity Functions	
Universal	-1.72 ± 0.01		-1.40 ± 0.01	
Proto-MW/LG	-1.65 ± 0.14		-1.33 ± 0.12	
	All progenitors	Fossil record	All progenitors	Fossil record
Sample I	-1.85 ± 0.30	-1.79 ± 0.26	-1.45 ± 0.24	-1.43 ± 0.22
Sample II	-1.76 ± 0.26	-1.69 ± 0.22	-1.41 ± 0.19	-1.35 ± 0.18
Sample III	-1.67 ± 0.18	-1.58 ± 0.16	-1.38 ± 0.14	-1.30 ± 0.14

NOTE—The values reported here for the three samples and the proto-MW/LG are derived from bootstrapping with replacement. The central estimates are the medians of μ_α samples, while the associated uncertainties are the medians of σ_α samples, which we define as our halo-to-halo scatter. The slopes of the universal SMF/UVLF are estimated by fitting a Schechter function and the uncertainties associated with them are statistical uncertainties.

Table 3. Low-mass-end and faint-end slope estimates at $z = 7$, for isolated and paired simulations separately.

Sample	Stellar Mass Functions				UV Luminosity Functions			
Universal	-1.72 ± 0.01				-1.40 ± 0.01			
Proto-MW	-1.66 ± 0.17				-1.33 ± 0.13			
Proto-LG	-1.63 ± 0.06				-1.33 ± 0.06			
	All progenitors		Fossil record		All progenitors		Fossil record	
	proto-MW	proto-LG	proto-MW	proto-LG	proto-MW	proto-LG	proto-MW	proto-LG
Sample I	-1.88 ± 0.34	-1.79 ± 0.16	-1.76 ± 0.27	-1.86 ± 0.18	-1.46 ± 0.25	-1.43 ± 0.17	-1.44 ± 0.25	-1.39 ± 0.13
Sample II	-1.75 ± 0.29	-1.79 ± 0.13	-1.64 ± 0.23	-1.81 ± 0.12	-1.39 ± 0.19	-1.46 ± 0.14	-1.35 ± 0.19	-1.36 ± 0.12
Sample III	-1.68 ± 0.20	-1.66 ± 0.07	-1.56 ± 0.17	-1.62 ± 0.11	-1.39 ± 0.16	-1.36 ± 0.08	-1.30 ± 0.15	-1.30 ± 0.09

NOTE—See Table 2 note.

a steeper slope than the proto-MW/LG (by $\approx 0.1 - 0.3$; sample III is more accurate than sample I).

6.2.5. Discussion of sensitivity analysis

Except when decreasing the upper mass limit, these results show that adding disrupted galaxies robustly reduces the halo-to-halo scatter, with the quantitative benefit depending on the sample selections employed. However, we can get a biased estimate of the slope in the *fossil record* approach, while still reducing the halo-to-halo scatter by adding more disrupted galaxies. The bias exists even while using just the intact galaxies, with the precise bias depending, again, on sample selection. The *all progenitors* scenario provides more accurate and precise recoveries of the underlying proto-MW/LG slope

when disrupted galaxies are added. This motivates future efforts to combine the traditional fossil record analysis with simulated merger trees, which may allow one to properly account for the multiple high redshift progenitors of individual present-day substructures, both intact and disrupted. If, instead, only the fossil record is used, it is important to test sensitivity to the selection criteria used for both present-day and high- z galaxies.

6.3. Redshift evolution

We also estimate α for the SMFs/UVLFs at $z = 6, 8$, and 9 , using our fiducial choice of $M_\star \geq 10^5 M_\odot$ or $M_{UV} \leq -10$ at high redshifts. We show the $\tilde{\mu}_\alpha$ and $\tilde{\sigma}_\alpha$ statistics as our central estimates for the low-mass/faint-end slope and the halo-to-halo scatter, re-

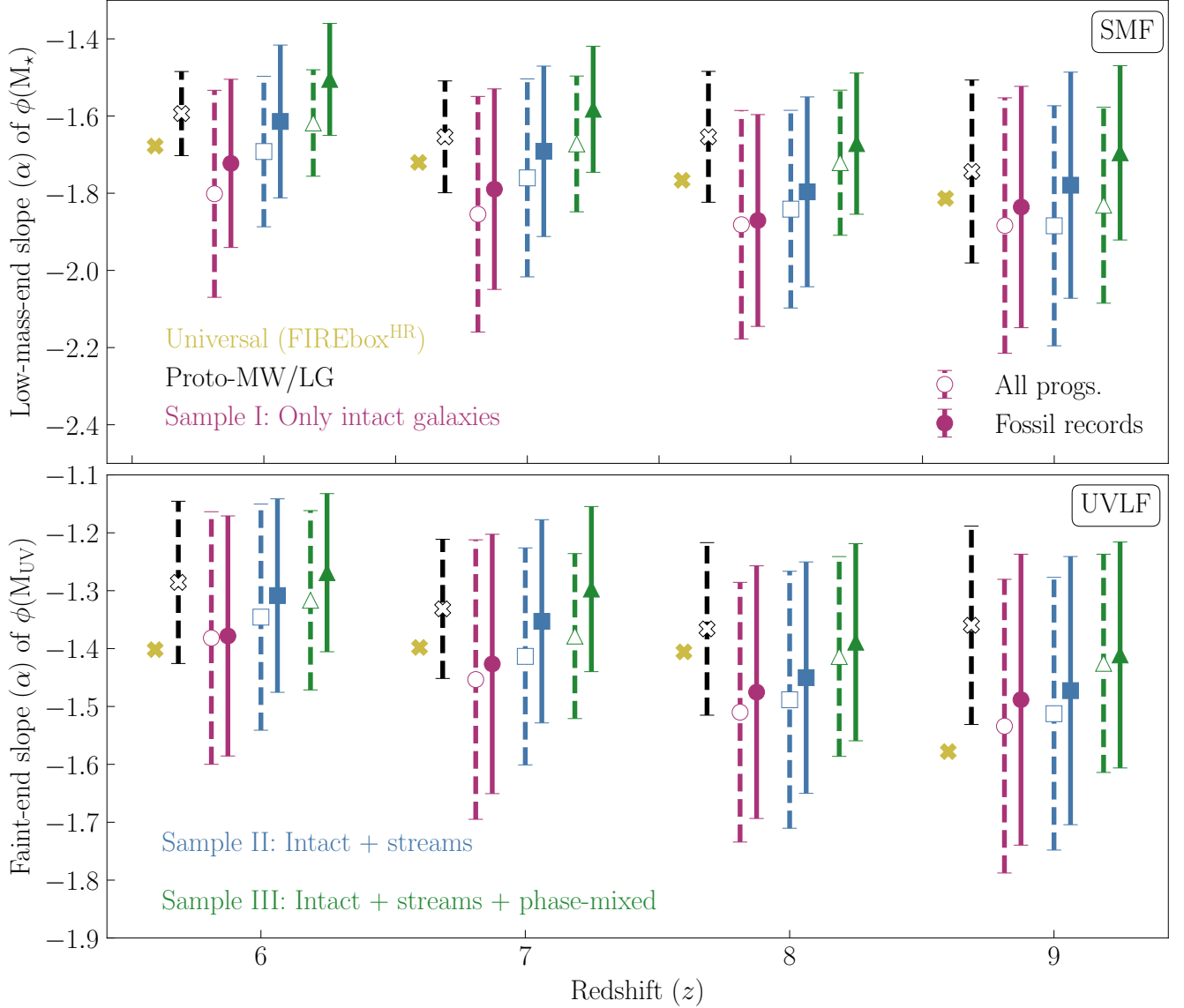


Figure 11. Redshift evolution of the low-mass/faint-end slope of the SMF/UVLF. Hierarchical assembly is reflected in the increasingly steep low-mass/faint-end slopes with redshift for all three samples, as well as the proto-MW/LG (black) and FIREbox^{HR} (yellow), for the differential SMF (top) and UVLF (bottom). In both panels, black, red, blue, and green points show $\bar{\mu}_\alpha$ values for the proto-MW/LG, sample I, II, and III respectively; errorbars show the $\bar{\sigma}_\alpha$ values as a measure of halo-to-halo scatter. Empty markers indicate slopes estimated in the *all progenitors* scenario, while filled markers denote the *fossil record* case.

spectively. The top panel of Figure 11 shows results for the low-mass-end slopes of the SMFs and the bottom panel shows the faint-end slopes for the UVLFs. The slope estimates from the *all progenitors* scenario is shown by the empty markers and dashed errorbars, while the filled markers and solid errorbars denote the slope estimates from the *fossil record* scenario. As a consequence of hierarchical structure formation in the cold dark matter picture with standard radiative and supernova feedback, the SMFs/UVLFs should generally get steeper towards higher redshifts. This is well-reflected

by the SMFs/UVLFs from FIREbox^{HR} (yellow), the proto-MW/LG (black), and all the three samples (I - red; II - blue; III - green).

In general, sample III consistently reproduces the most accurate and precise proto-MW/LG slope across the redshifts followed by sample II and I, in the *all progenitors* scenario. The trend is similar in the *fossil record* scenario, except in a few cases. The reduction in the halo-to-halo scatter with sample II and III is evident across the range of redshifts shown. In particular, at $z = 8$ and 9 , the large scatter in the slope from the progen-

itors of the intact galaxies alone is fairly well mitigated upon adding the disrupted galaxies, as shown by sample III. We caution, however, that the $z = 8$ and $z = 9$ results may be subject to the limited statistical samples of progenitor galaxies captured in our simulations at these redshifts. At $z = 9$, the faint-end slopes from our reconstructed UVLFs are shallower than the universal one. These estimates are probably poor because of the low halo abundance at these redshifts. Overall, for both the SMFs/UVLFs and for most redshifts, sample III reproduces the proto-MW/LG slope more accurately and also reduces the halo-to-halo scatter, followed by sample II and I. This is true for both the *all progenitors* and *fossil record* scenarios using our fiducial selection criteria. However, as discussed in section 6.1, the results can vary depending on our sample selection criteria, both at the present-day and at high redshifts.

7. RECONSTRUCTED COSMIC UV DENSITY

In this section, we reconstruct the cosmic UV density ρ_{UV} using the near-far technique, quantifying the contributions of different progenitor populations to the total cosmic UV density and hence their roles in reionization. We use the posterior distributions of the Schechter function parameters, α , M_{UV}^* , and Φ^* , from our Bayesian calculations (see Appendix B) to construct UVLF samples. Then we compute UV densities from the UVLFs using the following equations:

$$\rho_{UV}^{\text{re}}(> L_{UV}) = \int_{L_{UV}}^{\infty} L_{UV} \Phi^{\text{re}}(L_{UV}) dL_{UV} \quad (1)$$

$$\rho_{UV;\text{tot}}^{\text{re}} = \int_0^{\infty} L_{UV} \Phi^{\text{re}}(L_{UV}) dL_{UV} \quad (2)$$

$$\rho_{UV;\text{tot}}^{\text{FB}} = \int_0^{\infty} L_{UV} \Phi^{\text{FB}}(L_{UV}) dL_{UV} \quad (3)$$

where $\rho_{UV}^{\text{re}}(> L_{UV})$ (or equivalently $\rho_{UV}^{\text{re}}(< M_{UV})$) is the reconstructed cosmic UV density obtained from our reconstructed *fossil record* UVLFs $\Phi^{\text{re}}(L_{UV})$ due to the sources with UV luminosities L_{UV} ⁴ and brighter, $\rho_{UV;\text{tot}}^{\text{re}}$ is the total cosmic UV density from all the sources, and $\rho_{UV;\text{tot}}^{\text{FB}}$ is the total universal cosmic UV density computed by using the universal UVLF $\Phi^{\text{FB}}(L_{UV})$ from FIREbox^{HR}. Our analysis is done in the *fossil record* scenario in order to mimic real-world application of the near-far technique.

For each of the three samples in each halo, we compute the reconstructed cosmic UV density and explore their contribution to the total cosmic UV density. As shown

in Figure 12 (top panel), the proto-MW/LG (which also includes the progenitors of the present-day main host galaxy), can recover between $\sim 18\%$ and $\sim 45\%$ of the total universal cosmic UV density. We cannot fully recover the total universal cosmic UV density because of the difference in normalization between the UVLFs here and that of FIREbox^{HR}, as previously discussed in section 5. Hence, we mainly compare the contribution from the three samples and the proto-MW/LG. Sample III ($\sim 8\text{--}18\%$) is the most complete, followed by sample II ($\sim 4\text{--}16\%$), and sample I ($\sim 1\text{--}8\%$).

We also consider the relative contribution of different progenitor galaxies to the total *reconstructed* cosmic UV density within each sample at $z = 7$ (Figure 12, bottom panel). The median UV luminosity density reconstructed from sample I (red) under-predicts the contributions from galaxies brighter than $M_{UV} \sim -15$, which is the current JWST detection limit for unlensed galaxies⁵. Furthermore, the large scatter in the faint-end slope of the reconstructed UVLF from sample I also leads to sizable uncertainties in the inferred UV luminosity density and the relative importance of different galaxies in reionization. Quantitatively, as inferred from sample I, the luminosity above which galaxies produce $\sim 50\%$ of the total UV density is somewhere in the range $-16 \lesssim M_{UV} \lesssim -13.5$. Adding the progenitors of the disrupted galaxies in samples II and III (blue and green shaded regions, respectively) reduces the scatter. Sample II predicts minimum UV magnitudes in the range $-16.5 \lesssim M_{UV} \lesssim -14.5$, while sample III predicts a range of $-16.5 \lesssim M_{UV} \lesssim -15$. Furthermore, the median curves (solid curves) are more representative of the proto-MW/LG curve (black) when the disrupted galaxies are included. The median curve for sample III (green) shows slightly larger contributions from brighter galaxies than in the proto-MW/LG, because it predicts a shallower faint-end slope than that of the proto-MW/LG in the fossil record scenario (see Figure 10).

Interestingly, the results from samples II, III, and the proto-MW/LG in the bottom panel of Figure 12 indicate that contrary to the conclusion drawn from the intact galaxies alone, UV-bright progenitor galaxies ($M_{UV} \lesssim -15$) contribute significantly to the total cosmic UV density. For example, the medians for the proto-MW/LG and sample III predict that galaxies with $M_{UV} \lesssim -15$ contribute up to $\sim 65\%$ and $\sim 70\%$ of the total cosmic UV density at $z = 7$, respectively. On the other hand, in sample II this number is $\sim 55\%$, while

⁵ Measurements behind foreground lensing clusters can probe fainter galaxies down to $M_{UV} \gtrsim -13$ (Chemerynska et al. 2025), but they come with systematic uncertainties.

⁴ $M_{UV} = 51.6 - 2.5 \log_{10} (L_{UV} / [\text{erg s}^{-1} \text{ Hz}^{-1}])$ (Oke 1974)

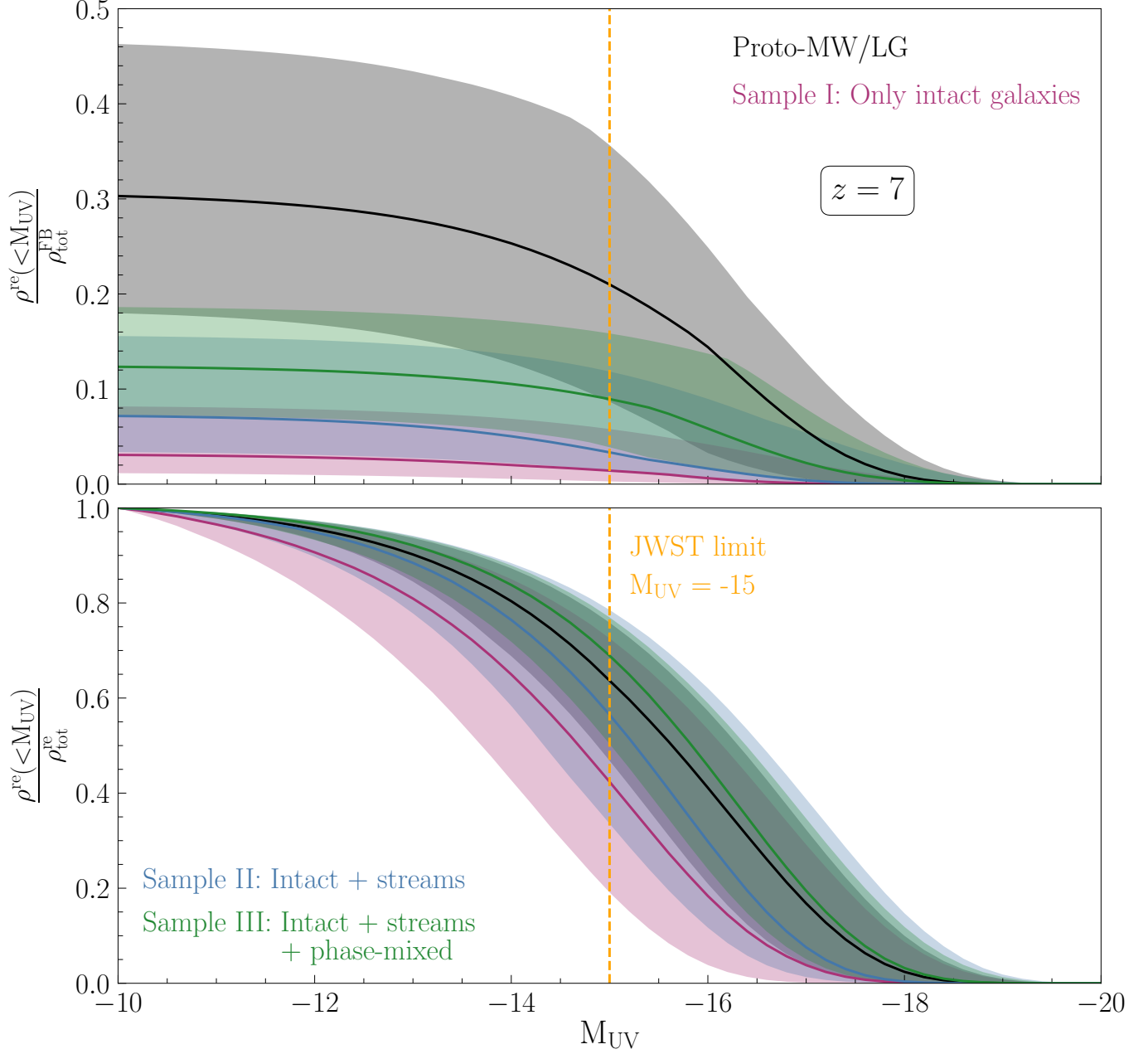


Figure 12. Reconstructed cosmic UV density at $z = 7$ using the fossil record. *Top:* contribution of different progenitor galaxies with UV magnitude brighter than M_{UV} from the fossil records of the different samples to the *universal* cosmic UV density calculated from FIREbox^{HR}. Solid lines shows the median across all of the halos; shaded regions enclose the 16th to 84th percentile range for the proto-MW/LG (black) and samples I (red), II (blue), and III (green). The orange dotted line shows JWST’s current observational threshold for unlensed galaxies at $z \sim 7$ (Adams et al. 2024). *Bottom:* relative contribution of different progenitor galaxies to the total *reconstructed* cosmic UV density within each sample. Including progenitors of disrupted galaxies reduces the uncertainty and predicts greater contributions from brighter galaxies, more representative of the proto-MW/LG. See Section 7 for further discussion.

it is only $\sim 40\%$ in sample I. Thus, the descendants of UV-bright reionization-era galaxies may end up largely in streams and phase-mixed galaxies. An accurate census of the reionization-era UV luminosity density from

the near-far technique thus requires accounting for these disrupted galaxies.

8. DETECTABILITY OF THE PRESENT-DAY SUBSTRUCTURES

Our results indicate that the present-day streams and phase-mixed galaxies are important for accurately and precisely inferring the high- z UVLF/SMF, so we next consider their detectability. It is generally more challenging to detect these substructures than intact galaxies of the same total stellar mass/luminosity. Stellar streams from low-mass galaxies are spread across their host galaxy, resulting in low surface brightnesses ($\mu \gtrsim 27$ mag arcsec $^{-2}$) and making resolved photometric measurements, needed to reconstruct their SFHs, more challenging and time-consuming than for intact low-mass galaxies. Nevertheless, numerous stellar streams have been detected in the MW (Helmi et al. 1999; Ibata et al. 2001; Majewski et al. 2003; Grillmair 2006; Newberg et al. 2009; Shipp et al. 2018; Malhan et al. 2018; Li et al. 2019; Malhan et al. 2021). Likewise, some phase-mixed galaxies have been detected in the MW, including the Gaia-Enceladus Sausage (Helmi et al. 2018; Belokurov et al. 2018; Haywood et al. 2018; Myeong et al. 2018a,b; Johnson et al. 2023), Sequoia (Barbá et al. 2019; Myeong et al. 2019), and Thamnos (Koppelman et al. 2019), although the techniques employed are different and more complex than those used for streams.

We expect important advances in identifying streams and phase-mixed galaxies with upcoming surveys. With LSST and Roman, we will be able to probe fainter surface brightness levels and detect more stellar streams, while also improving the quality of the existing sample. Roman will also resolve streams and satellites around nearby galaxies (Lancaster et al. 2022; Pearson et al. 2022b), potentially increasing our near-far sample beyond the LG. Here, we estimate surface brightness sensitivities to explore the detectability of stellar streams and intact galaxies with future LSST and Roman observations around the MW and external galaxies, respectively. We defer a study of the detectability of phase-mixed galaxies to future work, since in general they cannot be detected with photometry alone. We emphasize that here we strive only for preliminary surface brightness estimates, rather than a complete and accurate detectability test such like the one presented in Shipp et al. (2023) with DES. These approximate results may provide a useful precursor to more detailed mock catalogs for LSST and Roman, which we defer to future work.

We calculate the luminosity of each star particle in the V-band, assuming a mass-to-light ratio of $M_*/L_V = 2$ (Bullock & Johnston 2005), and use this to calculate the surface brightness of each stream and intact galaxy in both internal and external views. For the “internal view” given by the LSST of MW streams (Figure 13, top row), we put the observer at the center of the host galaxy and use the Galactocentric distance of each star particle

to calculate their respective fluxes. We use **HEALPY** to divide the sky into pixels of equal solid angle and calculate the surface brightness by dividing the total flux in each pixel by its solid angle. We divide the simulated sky into 256 equal areas and smooth the resulting surface brightness maps with a Gaussian kernel of $\sigma = 0.3$ deg, which sets the angular resolution of our final “image”. These values lead to a sufficient flux in each illuminated pixel to discern surface brightness variations without making the image too coarse. We choose a surface brightness limit of 34 magarcsec $^{-2}$ as our detectability threshold.

Among 13 simulated systems, an average of $\sim 29\%$ of streams should be detectable with visible, extended tidal tails by LSST, while an additional $\sim 29\%$ of streams should be detectable without the tidal tails as intact low-mass galaxies. Shipp et al. (2023) performed a similar analysis using DES-detectable streams in FIRE, but excluded the massive streams ($M_* \sim 5 \times 10^9, M_\odot$) included in our work. The detectability fraction for streams with extended tidal tails is comparable between the two studies, while our fraction for intact low-mass galaxies is slightly lower, likely due to differences in the stream samples considered. Although we do not show examples of intact galaxies here, we estimate detectability rates of $\sim 35\%$ for satellites and $\sim 18\%$ for field galaxies. Adding the streams detectable as intact low-mass galaxies increases the percentage to $\sim 43\%$ for satellites. These results illustrate the general promise of LSST for detecting intact and disrupted galaxies, but are only a first step towards more detailed investigations. For example, the angular resolution of LSST is actually significantly higher than in these preliminary calculations, while our V-band magnitude estimates lie in between the actual LSST u and g bands.

For analyzing the detectability of streams and satellites of external galaxies using Roman, we employ the same method to calculate the luminosities of the simulated star particles. We then place the observer at a distance of 5 Mpc from the center of the galaxy, and calculate the flux and surface brightness in 50 arcsecond angular pixels. This angular scale spans a physical size that is larger than the mean inter-particle separation in the simulations. The bottom panel of Figure 13 shows these results, for the same set of three streams and a surface brightness limit of 34 magarcsec $^{-2}$ (based on the depth reached by the High-Latitude Wide Area Survey for resolved stars). As for the internal view, the stream in the left panel should be easily detectable and the stream in the middle panel may be detected as an intact low-mass galaxy. However, although the stream in the right panel is undetectable in the internal view, it passes our surface brightness threshold in the external

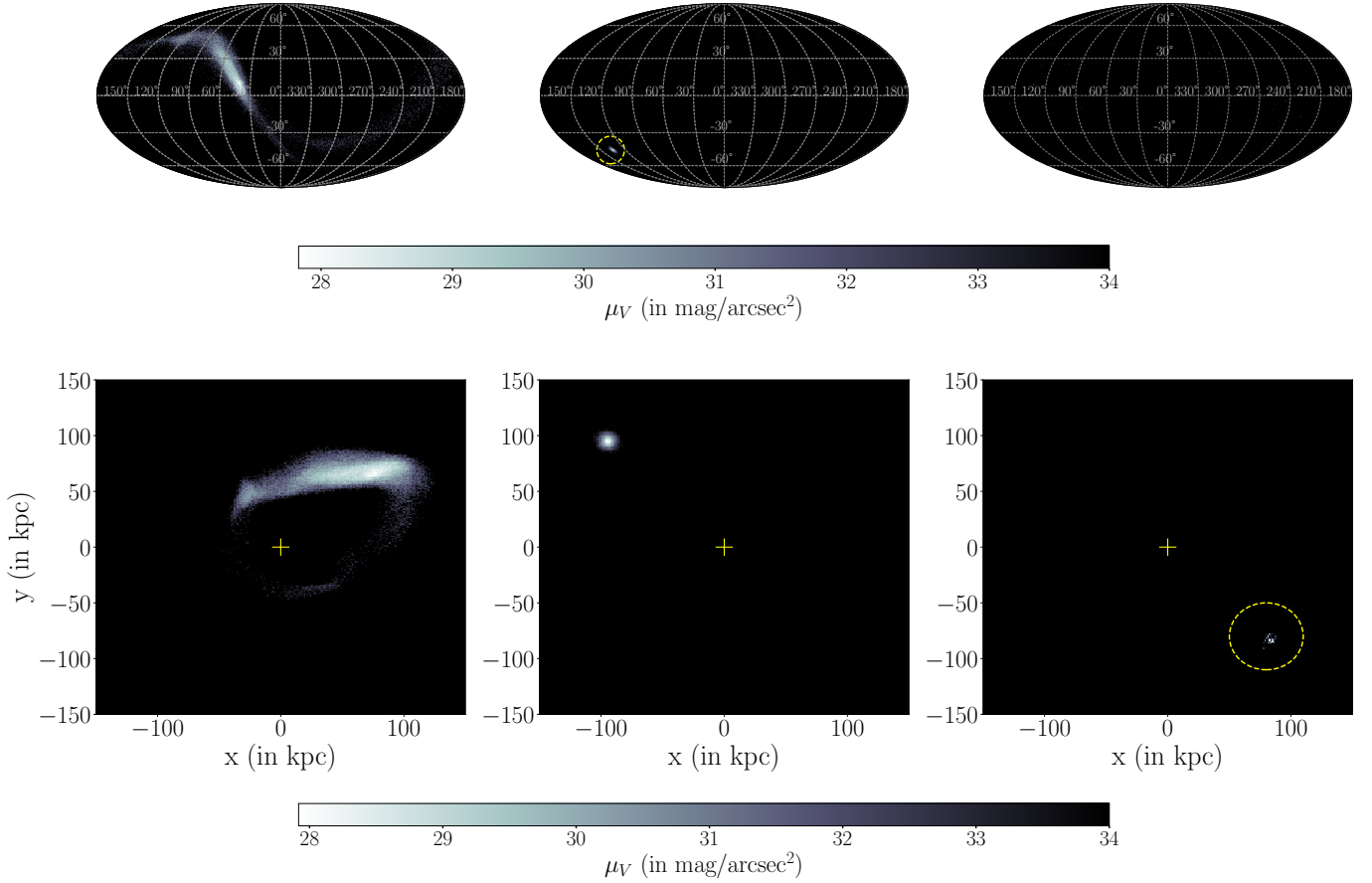


Figure 13. Detectability of streams in the MW and external galaxies. *Top row:* Surface brightness maps of three different streams as viewed from the center of the MW (“internal view”) using LSST, with a faint-end limit of $\mu_V = 34$ mag arcsec $^{-2}$. *Bottom row:* Surface brightness maps of the same three streams, but now viewed from a distance of 5 Mpc (“external view”) to the same surface brightness limit, as expected for Roman. The yellow markers correspond to the center of the host galaxy. **Left:** An easily detectable stream with long tidal tails in both internal and external views. **Center:** a stream that would be classified as an intact low-mass galaxy in both views because of its faint tidal tails. **Right:** a stream that is unlikely to be detectable in the internal view, but would be detected as a faint low-mass galaxy in the external view (inside yellow circle). This example was chosen to show a case which is almost visible from the internal view.

view and is detectable as a very faint intact galaxy. We find similar results for a distance of 20 Mpc, although the resulting images are less sharp. Overall, we find that ~ 12 % of the streams have tidal tails detectable by Roman at this surface brightness, while ~ 39 % would be classified as intact low-mass galaxies. The detectability percentages for simulated satellites and field galaxies are ~ 29 % and ~ 17 %, respectively.

These results support the notion that significant populations of stellar streams and intact galaxies may be detectable in the MW and external galaxies with the LSST and Roman, and could potentially be used to probe the low-mass/faint-end of the SMF/UVLF during the EoR. In future work, however, it will be important to refine our detectability estimates with more realistic LSST and Roman synthetic observations.

Our analysis reveals that some of the streams could be misidentified as intact low-mass galaxies, which could (incorrectly) imply that high- z progenitor galaxies are less disrupted than in our simulations. This, in turn, might be used to mistakenly infer smaller contributions of disrupted progenitors to the reionization-era stellar mass budget. Furthermore, the SFHs derived from these present-day substructures will be incomplete until deeper observations are carried out. Therefore, caution must be taken when reconstructing SFHs in the near-far technique, as these effects can influence the inferred shape of the SMFs/UVLFs. In our analysis here, note that we assign all galaxies with detectable streams in the streams category, rather than including sub-classifications related to the presence of the tidal tails.

To complete our analysis, we construct UVLFs by imposing the detectability constraints obtained from the internal view scenario using LSST. Figure 4—first shown in Section 5.1—gives the differential UVLFs at $z = 7$ reconstructed from different classification groups (solid curves) in comparison to reconstructions from their detectable counterparts alone (dashed curves). The reconstructed UVLFs appear promising, even after imposing the detectability constraints. Although the normalization of the satellite and field galaxy progenitor UVLFs is reduced, their shapes are relatively well-preserved. The progenitor UVLFs reconstructed from the streams remain almost unchanged even with just $\sim 50\%$ detectability, since the detectable streams are mostly of higher mass and each massive stream typically has numerous reionization-era progenitors (see Section 4.2), spanning much of the high- z distribution in progenitor luminosity. We show the results in the *all progenitors* scenario by utilizing the full merger histories of the present-day substructures, and not in the *fossil record* scenario. However, the difference between the shape of the UVLFs with and without detectability constraints does not change significantly in the *fossil record* scenario.

9. SUMMARY

In this work, we test whether including present-day disrupted galaxies like stellar streams and phase-mixed galaxies in the near-far technique improves the inference of the low-mass/faint-end slope of the SMF/UVLF during the EoR. Using the FIRE-2 cosmological zoom-in baryonic simulations of MW-like (isolated) and LG-like (paired) halos (Section 2), we classify present-day substructures into 4 categories: intact satellite galaxies, field galaxies, stellar streams, and phase-mixed galaxies (Section 3). The first two categories constitute our intact galaxies sample (sample I); the intact galaxies and streams comprise our sample II; adding the phase-mixed galaxies makes our sample III. Using particle tracking, we assign each star particle to its progenitor galaxy at high redshifts. We then assess the contribution of different progenitor populations to the total UV luminosity density (Sections 4–6). Finally, we estimate the proportion of substructures that will be detected by LSST (for the Milky Way) or Roman (for external galaxies) and the effect of these detection limits on the reconstructions (Section 8). Our results are summarized below.

Our results are consistent in general with Gandhi et al. (2024), with a few differences. Gandhi et al. (2024) estimated the low-mass-end slopes of the reconstructed fossil record SMFs using just intact low-mass galaxies and found close agreement with

the proto-MW/LG slopes. We reach similar conclusions but also show that the halo-to-halo scatter associated with the slope is large when using just intact galaxies. Their slope estimates are not exactly the same as our estimates for sample I, since some of their present-day low-mass galaxies are part of our streams and hence do not show up in our sample I. This also leads to differences in the shape and normalization of our respective intact low-mass galaxy fossil record SMFs, with a stronger mass dependent bias in our case. Additionally, our methods for computing the slopes are different. Furthermore, Gandhi et al. (in prep.), use an alternative set of zoom-in high-redshift FIRE-2 simulations as a benchmark for the universal SMFs/UVLFs, instead of FIREbox^{HR}. Both sets of simulations show good agreement with each other, as well as with observed high- z SMFs/UVLFs and cosmic UV luminosity density, but they may not match exactly.

The progenitors of streams and phase-mixed galaxies represent a larger fraction of the stellar mass in the proto-MW/LG at $z = 6 - 9$ than the intact galaxies commonly used for the near-far technique. Phase-mixed galaxy progenitors make the dominant contribution to the reionization-era population of low-mass galaxies in both the isolated and paired galaxy simulation suites. Thus, reionization in the proto-MW/LG may have been primarily driven by galaxies which subsequently merged/disrupted into their present-day host galaxies. In the isolated simulations, the second largest contribution comes from the stream progenitors, followed by the field galaxies. In the paired simulations, the field galaxies play a larger role than the streams, due the larger abundance of field galaxies in the paired simulation suite (likely an effect of the larger mean density and earlier formation time). In both simulation suites, the surviving satellite galaxy progenitors produce the *smallest* fraction of the $z \sim 6 - 9$ stellar mass. Ignoring the progenitors of disrupted galaxies while using the near-far technique can thus lead to less accurate conclusions regarding the low-mass-/faint-end part of the high- z SMFs/UVLFs, due to an incomplete progenitor population.

Inclusion of tidally disrupted galaxies leads to more accurate high- z reconstructed SMFs/UVLFs. We analyze the SMFs/UVLFs under two scenarios: the *all progenitors* scenario, in which we account for the stellar mass contribution from each progenitor of a present-day substructure separately; and the *fossil record* scenario, where we assume that the stellar mass from all progenitors of each substructure is contained within a single galaxy, as would be inferred observationally in the absence of merger his-

tory information. In both of these scenarios, our sample III recovers most of the amplitude ($\sim 75\%$) of the proto-MW/LG SMF/UVLF, followed by sample II ($\sim 50\%$) and I ($\sim 25\%$), respectively. This represents a significant improvement, reducing the normalization rescaling from a factor of ~ 4 in sample I to just ~ 1.5 in sample III, after including the disrupted galaxies. However, in the *fossil record* scenario, the SMFs/UVLFs become shallower in the low-mass/faint-end region, with the effect being large when more disrupted galaxies are added. Nevertheless, in most instances, the SMFs and UVLFs in the *fossil record* scenario trace the shape of the proto-MW/LG SMFs/UVLFs between $10^5 M_\odot \lesssim M_\star \lesssim 10^{6.5} M_\odot$ and $-14 \lesssim M_{\text{UV}} \lesssim -10$, respectively. The *fossil record* results differ from the all progenitors case because the latter approach combines multiple low-mass/faint progenitors into a single more massive system. This difference is compounded in the case of disrupted galaxies which are generally more massive and have multiple reionization-progenitors. We also compare our results with the differential SMFs/UVLFs of the FIREbox^{HR} simulation, as a benchmark for the universal SMF/UVLF, and find close agreement in shape. Although the proto-MW/LG slopes are in agreement with FIREbox^{HR} within our estimated uncertainties, they are in general flatter by ~ 0.07 . This could be due the difference in the mean matter density of the region around the MW/LG and the mean matter density of the universe, but a detailed exploration is needed, which is beyond the scope of this paper.

The low-mass/faint-end slope is more robustly constrained when including disrupted galaxies. We estimate the low-mass/faint-end slope of the reconstructed SMFs/UVLFs using a Bayesian approach, to test how accurately and precisely we can recover the proto-MW/LG and the universal SMF/UVLF slopes. In the *all progenitors* scenario, adding disrupted galaxies helps reproduce the low-mass/faint-end slope of the proto-MW/LG SMF/UVLF more accurately and more precisely, and these are also representative of the global-average low-mass/faint-end slopes. Sample III reproduces the proto-MW/LG slope most accurately with an offset of ≈ 0.04 , followed by sample II (≈ 0.12) and I (≈ 0.26). Furthermore, when we use only the intact galaxies, there is a large halo-to-halo scatter due to the varying accretion histories across different halos, even though the median values agree well with that from the proto-MW/LG. Sample III reduces the scatter of sample I by $\sim 40\%$, while sample II reduces it by $\sim 20\%$. Including the disrupted galaxies improves our inferences of the median slopes and also reduces the halo-to-halo scat-

ter. In the *fossil record* scenario, we find similar results when we select high- z progenitors with $M_\star \geq 10^5 M_\odot$ and $M_{\text{UV}} \leq -10$. However, even though the halo-to-halo scatter decreases, we can get a biased estimate of the slope from the fossil record depending on our choice of both the present-day and high- z galaxy samples. The estimated slopes vary depending on the stellar mass range of the galaxies selected, both at the present-day and during the EoR, even in the case of using only intact galaxies. In the future, it may be possible to augment the observed SFHs with average merger histories from cosmological simulations, providing yet more accurate and precise estimates of the low-mass/faint-end slope. Nevertheless, including the streams and phase-mixed galaxies improves the statistical constraints on the proto-MW/LG SMFs/UVLFs, reducing the halo-to-halo scatter (~ 0.13 in case of proto-MW/LG) from ~ 0.24 (sample I) to ~ 0.20 (sample II) and ~ 0.15 (sample III). A similar trend holds when we separate into MW-like and LG-like environment. This improvement is therefore crucial, since we have only one MW and LG with which to apply the near-far technique.

Adding disrupted galaxies implies a larger contribution of bright star-forming galaxies to the total cosmic UV density in our simulations. Using the reconstructed UVLFs in the *fossil record* scenario, we compute the cosmic UV density at $z = 7$ and calculate the contribution of different progenitor galaxies to the total universal cosmic UV density. We recover only a small fraction of the UV luminosity density budget when including only intact galaxy progenitors, and especially under-predict the amount of UV radiation from galaxies brighter than $M_{\text{UV}} \lesssim -15$. Adding disrupted galaxies improves the recovery at these luminosities, and reduces the halo-to-halo scatter in the recovered UV luminosity density. The improved reconstructions at JWST-accessible luminosities with $M_{\text{UV}} \lesssim -15$ will allow more robust comparisons with direct UV luminosity density estimates from JWST UVLFs.

Applying the near-far technique to streams is feasible with upcoming surveys. We estimate that 58% of our simulated stellar streams, down to $10^6 M_\odot$ in stellar mass, would be detectable using the LSST: $\sim 29\%$ of streams with detectable tidal tails, and the others without a prominent tail like intact low-mass galaxies. Roman will enable searches for streams in external galaxies; in this case, we still expect to detect a significant number of streams, but a larger fraction will be classified as intact low-mass galaxies. These preliminary results illustrate the general promise of including stream detections in the near-far technique.

1 AK would like to thank Zachary Langford for valu-
 2 able discussions on Bayesian modeling. We would also
 3 like to thank Arpit Arora, Guochao Sun, and Nora
 4 Shipp for insightful suggestions, discussions, and com-
 5 ments that have improved the manuscript. We acknowl-
 6 edge use of the *marmalade* cluster at the University of
 7 Pennsylvania, on which the analysis of this work was
 8 done. AK and RES acknowledge support from NSF
 9 grants AST-2007232 and AST-2307787. RES further ac-
 10 knowledges support from NASA awards 22-ROMAN22-
 11 0055, 22-ROMAN22-0011, and 22-ROMAN22-0013, Si-
 12 mons Foundation grant 1018462, and Sloan Founda-
 13 tion award FG-2023-20669. AK and AL acknowl-
 14 edge support from the Charles E. Kaufman founda-
 15 tion through grant KA2022-129518. MBK acknowledges
 16 support from NSF grants AST-1910346, AST-2108962,
 17 and AST-2408247; NASA grant 80NSSC22K0827; HST-
 18 GO-16686, HST-AR-17028, HST-AR-17043, JWST-
 19 GO-03788, and JWST-AR-06278 from the Space Tele-
 20 scope Science Institute, which is operated by AURA,
 21 Inc., under NASA contract NAS5-26555; and from the
 22 Samuel T. and Fern Yanagisawa Regents Professorship
 23 in Astronomy at UT Austin.

Software: *astropy* (Astropy Collaboration et al. 2013, 2018, 2022), *matplotlib* (Hunter 2007), *IPython* (Perez & Granger 2007), *NumPy* (Van Der Walt et al. 2011; Harris et al. 2020), *SciPy* (Virtanen et al. 2020), *GizmoAnalysis* (Wetzel & Garrison-Kimmel 2020a), *HaloAnalysis* (Wetzel & Garrison-Kimmel 2020b; Wetzel et al. 2016), *PyMC* (Abril-Pla et al. 2023), *healpy* (Zonca et al. 2019), *scikit-learn* (Pedregosa et al. 2011), NASA’s Astrophysics Data System (ADS), arXiv

APPENDIX

A. CUMULATIVE STELLAR MASS FUNCTIONS AND UV LUMINOSITY FUNCTIONS

Figures A1 and A2 show the cumulative versions of the SMFs and the UVLFs at $z = 7$, respectively. Similar to Figures 7 and 8, we show the median SMFs/UVLFs in the *all progenitors* scenario (solid curves), and in the *fossil record* scenario (dashed curves). The shaded regions show the halo-to-halo scatter at 68% confidence. The conclusions are similar to what we found from the differential functions. Adding the disrupted galaxies improves the normalization and the shape, particularly in the *all progenitors* scenario. The cumulative SMFs in the ‘all progenitors’ scenario flattens near $10^4 M_\odot$ due to the mass resolution limits of our simulations. This is more prominent in the isolated suite, since their mass resolution is two times coarser than in the paired simulations. In the *fossil record* scenario, the flattening in the cumulative SMFs is mainly due to combining multiple progenitors into a single one, which in turn reduces the number of low-mass galaxies. The flattening effect is less apparent for the cumulative UVLFs because we avoid the extreme faint end of the luminosity functions which are susceptible to numerical resolution artifacts.

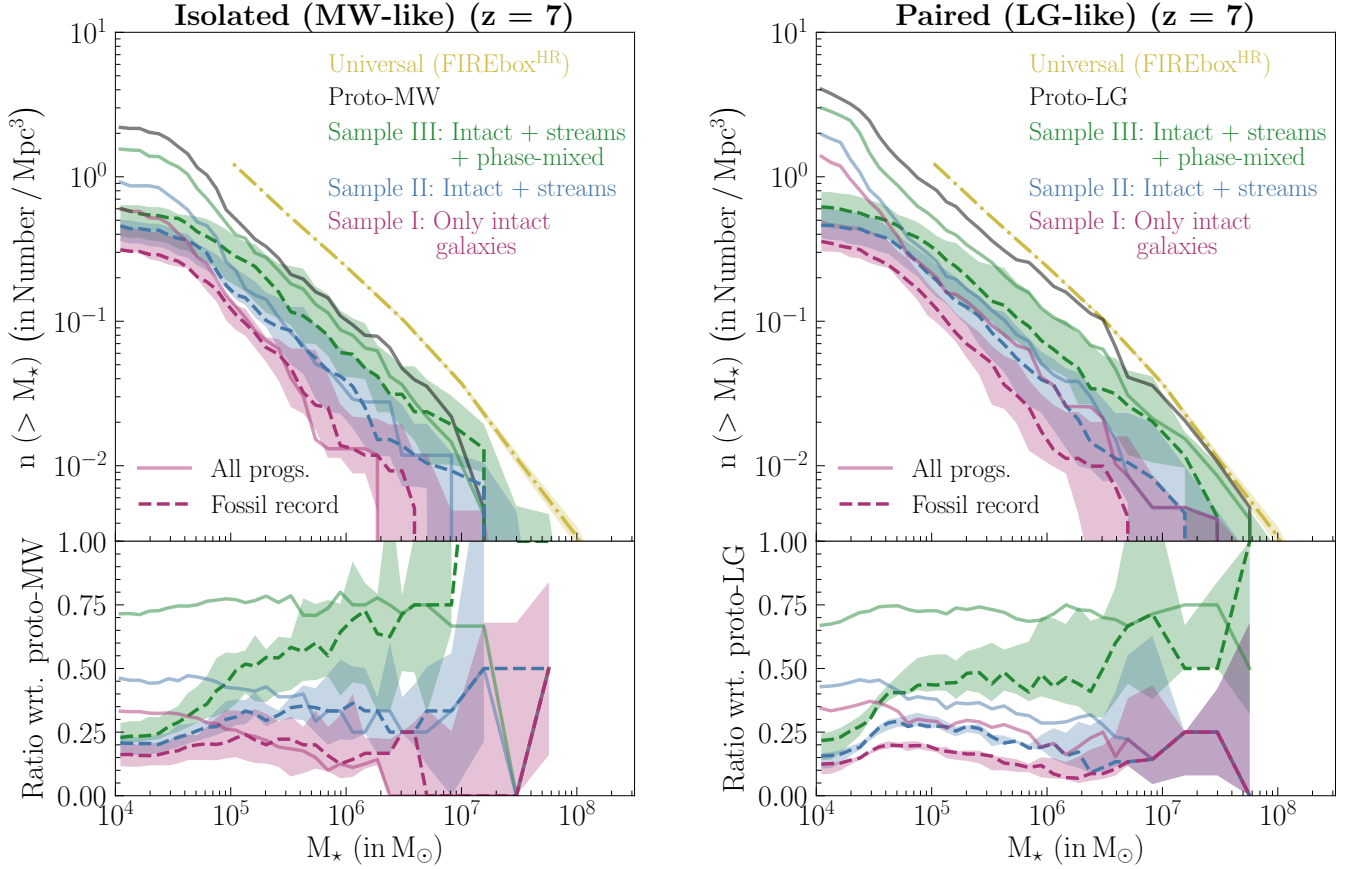


Figure A1. Cumulative SMFs at $z = 7$. Same as Figure 7, but for the cumulative version of the SMFs. The key takeaways are similar to those from Figure 7.

B. SLOPE ESTIMATES FROM BAYESIAN MODELING

In section 6.1, we use Bayesian models of Schechter functions to estimate the low-mass/faint-end slopes of the SMFs/UVLFs. This approach avoids binning, which is important because otherwise the slope estimates depend somewhat on the choice of binning scheme. The Bayesian approach described here produces more robust slope determinations.

Using progenitor galaxies with $M_* \geq 10^5 M_\odot$ and with $M_{UV} \leq -10$, we estimate the posterior distributions for the α parameter, assuming Schechter functions of the form:

$$\Phi(\log M) = 10^{(\alpha+1)(\log M - \log M^*)} \exp\left(-10^{\log M - \log M^*}\right) \ln 10 \quad (\text{B1})$$

$$\Phi(M_{UV}) = \frac{\ln 10}{25} 10^{0.4(\alpha+1)(M_{UV}^* - M_{UV})} \exp\left(-10^{0.4(\alpha+1)(M_{UV}^* - M_{UV})}\right) \quad (\text{B2})$$

where M is the stellar mass, M_{UV} is the rest-frame UV magnitude, M^* and M_{UV}^* are the characteristic stellar mass and UV magnitude, respectively. These describe the shape of the SMFs and UVLFs, while the normalization is set according to:

$$\Phi_{\text{norm}}(\log M) = \frac{\Phi(\log M)}{\int_{\log M_{\min}}^{\log M_{\max}} d(\log M) \Phi(\log M)}, \quad (\text{B3})$$

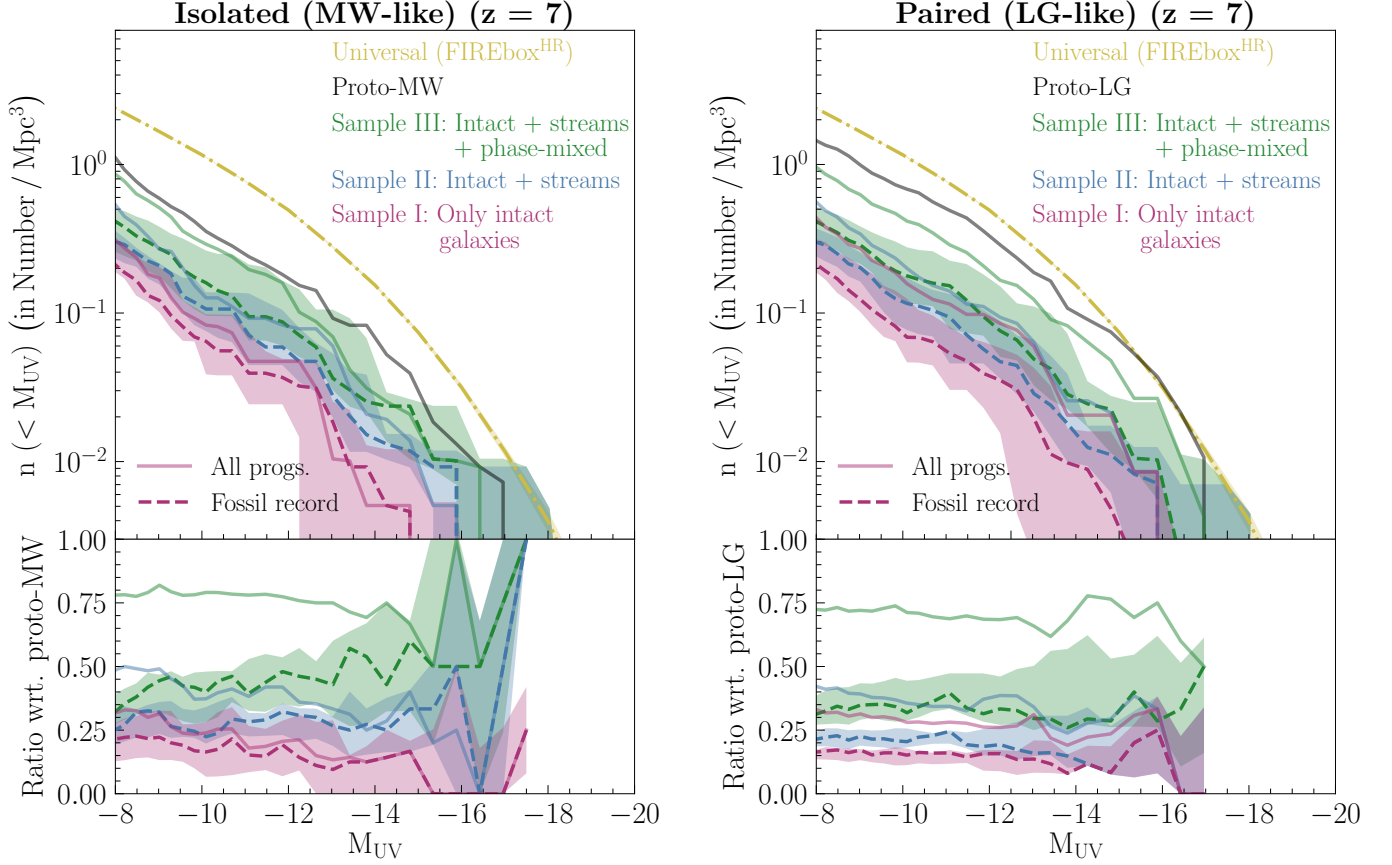


Figure A2. Cumulative UVLFs at $z = 7$. Same as Figure 8, but for the cumulative version of the UVLFs. The key takeaways are similar to those from Figure 8.

and:

$$\Phi_{norm}(M_{UV}) = \frac{\Phi(M_{UV})}{\int_{M_{UV, \text{bright}}}^{M_{UV, \text{faint}}} dM_{UV} \Phi(M_{UV})}. \quad (\text{B4})$$

Our results are insensitive to the high-mass/bright-end limit of the integrations as long as our highest-mass/brightest progenitor galaxy lies within this limit. The low-mass/faint-end limit does affect our results significantly, as discussed in 6.1. We count the number of progenitors n in a given sample to estimate the expected number N using a Poisson distribution:

$$P(n|N) = \frac{N^n \exp(-N)}{n!}. \quad (\text{B5})$$

We incorporate this as a multiplicative factor in our likelihood function, accounting for the Poisson scatter in the number of progenitor galaxies. Hence, our log-likelihood function can be written as,

$$\ln \mathcal{L} = \sum \log(\Phi_{norm}) + n \log N - N - \log(n!) \quad (\text{B6})$$

for both the SMFs and the UVLFs.

We adopt truncated normal distributions for the priors on each parameter. For the SMFs we assume:

$$\begin{aligned} \alpha &\sim \psi(\mu = -1.8, \sigma = 0.4, a = -3.0, b = -1.0) \\ \log M^* &\sim \psi(\mu = 7.0, \sigma = 1.0, a = 4.0, b = 11.0) \\ N &\sim \psi(\mu = 100, \sigma = 100, a = 1, b = 2000), \end{aligned} \quad (\text{B7})$$

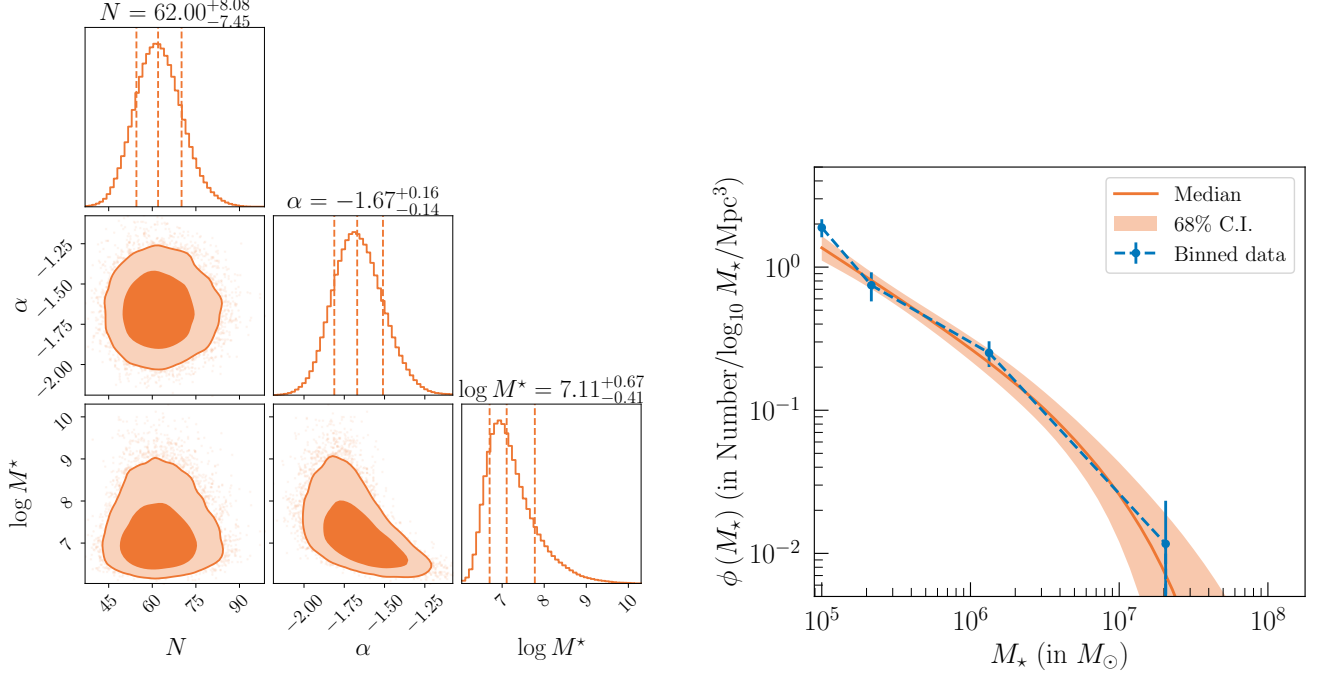


Figure B3. Bayesian modeling of Schechter stellar mass functions for the proto-MW of the *m12i* halo at $z = 7$. *Left* - The posterior distributions of our parameters α , N , and $\log M^*$, and their covariances showing 1σ and 2σ contour intervals. The dotted lines denote the median and 68% CI of the distributions. *Right* - The blue points denote the binned SMF at $z = 7$, while the orange solid line and shaded regions give the median and 68% CI among the samples of Schechter stellar mass functions drawn from the parameter posteriors. The binned SMF is well represented by our Schechter function models.

while for the UVLFs we use:

$$\begin{aligned}
 \alpha &\sim \psi(\mu = -1.5, \sigma = 0.4, a = -3.0, b = -1.0) \\
 M_{\text{UV}}^* &\sim \psi(\mu = -15.0, \sigma = 2.0, a = -25.0, b = -2.0) \\
 N &\sim \psi(\mu = 100, \sigma = 100, a = 1, b = 2000).
 \end{aligned} \tag{B8}$$

Here μ , σ , a , and b denote the mean, standard deviation, lower and upper limits of the truncated normal distributions, respectively. These priors are partially motivated by values obtained in literature and also allow reasonably large range to explore the parameter space. Using the PyMC package, we then estimate the posterior distributions from the above priors and likelihood function.

To test whether our method can reproduce the binned SMF, we generate samples of Schechter functions $\Phi(M)$ using the parameter posteriors. We determine the characteristic normalization, Φ^* , of the Schechter functions according to:

$$\Phi^* = \frac{N}{V \times \int_{\log M_{\min}}^{\log M_{\max}} d(\log M) \Phi(\log M)}, \tag{B9}$$

where V is the convex hull volume used in our calculations. Finally, the normalization from equation B9 is multiplied by the Schechter function shape of equation B1 to return a Schechter stellar mass function sample. We employ similar technique to generate samples from the Schechter UV luminosity function.

The left panel of Figure B3 shows an example corner plot, illustrating the posterior distributions from our analysis. The example distributions are for the proto-MW SMF in the *m12i* halo at $z = 7$. The α and N posteriors appear Gaussian, while that of $\log M^*$ deviates from Gaussianity. The expected number of progenitor galaxies N is also in

close agreement with the actual number of progenitor galaxies for the m12i proto-MW with $M_\star \geq 10^5 M_\odot$. The right panel of Figure B3 shows how well our method reproduces the SMF in this case, where the solid orange line gives the median of our Schechter stellar mass function samples, while the shaded region denotes the 68% CI. The blue points show the binned SMF, with Poisson error bars. The Schechter function models agree with the binned measurements, as expected.

To derive the slopes for the universal SMFs/UVLFs from FIREBox^{HR}, we use a normalized Schechter function with parameters α , $\Phi^\star$, and $M^\star$ or $M_{UV}^\star$, as our likelihood function and fit to the binned SMFs/UVLFs. This is different than what we do for our core suite as discussed above, but we also get posteriors of the parameters using this Bayesian approach.

C. BOOTSTRAPPING METHODOLOGY

In section 6.1, we used bootstrapping to compute a central estimate of the low-mass/faint-end slope and the scatter across the 10 simulated halos. First, we use the mean and standard deviation of the posterior distribution of the α parameter to draw samples from a Gaussian distribution with this mean and standard deviation. The assumption of Gaussianity is supported by the posterior distribution of α shown in the left panel of Figure B3. Next, we use bootstrapping with replacement to select 10 halos at random such that we get 10 values of α , and calculate their mean (μ_α) and standard deviation (σ_α). Then we repeat these two steps 10000 times to get samples of μ_α and σ_α . Finally, we compute the median values of these samples, $\tilde{\mu}_\alpha$ and $\tilde{\sigma}_\alpha$, and define them as our central estimate of the low-mass/faint-end slope and the associated halo-to-halo scatter across the 10 halos. We report these values in Table 2. Figure C4 shows the distribution of our μ_α and σ_α samples for the proto-MW/LG Schechter mass function, with the vertical blue line indicating the median values $\tilde{\mu}_\alpha$ and $\tilde{\sigma}_\alpha$, respectively.

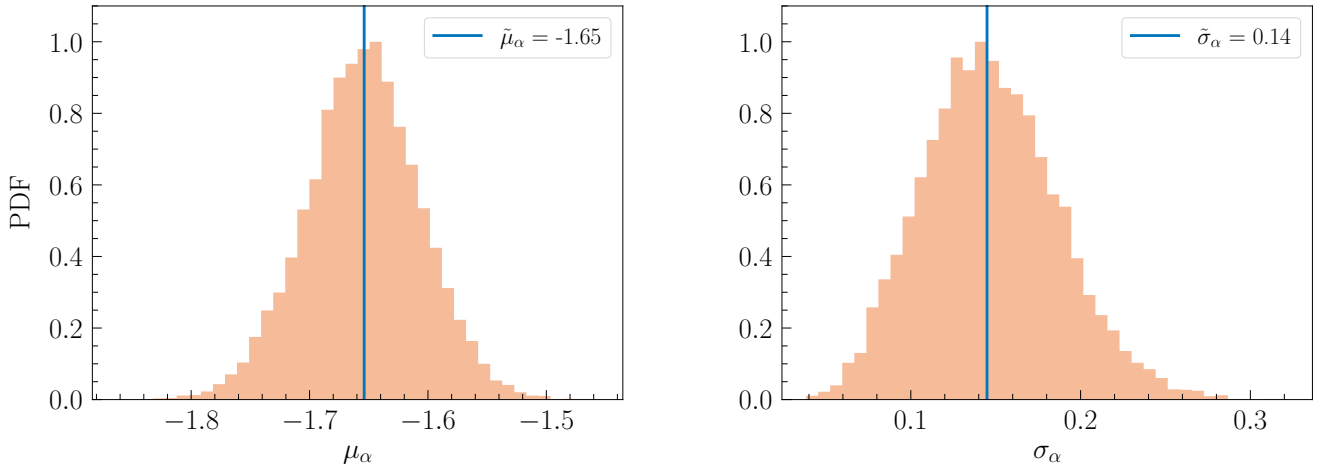


Figure C4. Distribution of μ_α and σ_α using bootstrapping. The histograms show the probability density function of μ_α (left) and σ_α (right) obtained from bootstrapping with replacement for the proto-MW/LG Schechter mass function. The vertical lines denote the median values of these respective distributions which we use as our central estimate and halo-to-halo scatter of the low-mass/faint-end slope respectively.

D. UV LUMINOSITY - STELLAR MASS RELATION

To investigate the shallower slope of the faint-end part of the UVLFs compared to that of the low-mass-end of the SMFs, we first analyze the relation between the UV luminosities and the stellar masses of the progenitor galaxies. We take all of the luminous galaxies from our halo catalogs at $z = 7$, and combine the results across all of our simulations. We then bin in stellar mass and compute the median and 68% CI in each bin. To fit the relation, we use a power-law of the form:-

$$L_{UV} = L_0 (M_\star / M_0)^\beta \quad (\text{D10})$$

Our best-fit power-law index is $\beta = 1.22$. From the differential SMF, $\Phi(M_\star)$, we can derive the differential UVLF, $\Phi(M_{\text{UV}})$, using the following equations:-

$$\Phi(L_{\text{UV}}) = \Phi(M_\star) \frac{dM_\star}{dL_{\text{UV}}}, \quad (\text{D11})$$

and

$$\Phi(M_{\text{UV}}) = \Phi(L_{\text{UV}}) \frac{dL_{\text{UV}}}{dM_{\text{UV}}}, \quad (\text{D12})$$

where $M_{\text{UV}} \propto -2.5 \log_{10} L_{\text{UV}}$.

Since the Schechter mass function $\Phi(M_\star)$ follows a power-law of the form $\Phi(M_\star) \propto M_\star^{\alpha_{M_\star}}$ at the low-mass-end, we use the $L_{\text{UV}} - M_\star$ relation to derive the form of $\Phi(L_{\text{UV}})$. Using equations D10, D11, and D12, we find that $\Phi(M_{\text{UV}})$ also follows a power-law of the form:-

$$\Phi(M_{\text{UV}}) \propto 10^{-0.4 M_{\text{UV}} (1 + \alpha_{M_\star}) / \beta} \quad (\text{D13})$$

The Schechter function at the faint-end is $\Phi(M_{\text{UV}}) \propto 10^{-0.4 M_{\text{UV}} (1 + \alpha_{\text{UV}})}$, and so we can write $\alpha_{\text{UV}} = \frac{1 + \alpha_{M_\star}}{\beta} - 1$. Using the low-mass-end slope of the proto-MW/LG differential SMF from table 2 ($\alpha_{M_\star} = -1.65$), we get $\alpha_{M_{\text{UV}}} = -1.53$ for $\Phi(M_{\text{UV}})$, which is a bit shallower, but still differs from our main results (see table 2). Hence, the shallower faint-end slopes of the UVLFs compared that of the low-mass-end of the SMF can be partially attributed to the superlinear dependence of the UV luminosity on the stellar mass. This is because the UV radiation is mostly generated by young stars and the smaller galaxies are not as efficient at forming them, which leads to a suppression at the faint-end. We show the relation and its corresponding linear fit in Figure D5. The gray points show the entire set of luminous galaxies from our halo catalogs, the blue points with errorbars show the median L_{UV} and 68% CI in each stellar mass bin, and the orange line is the linear fit to the relation.

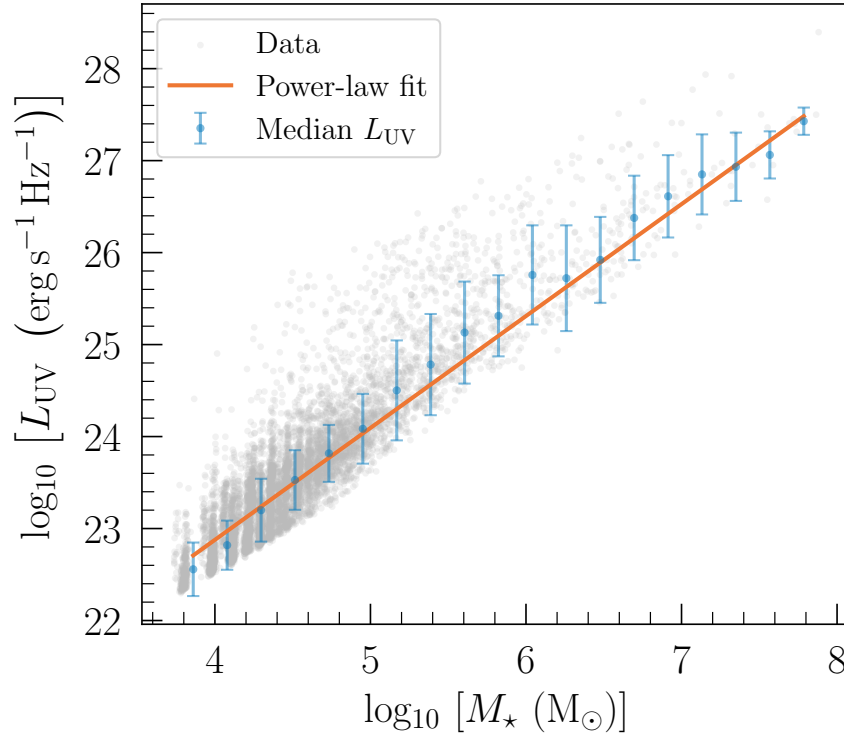


Figure D5. Superlinear dependence of UV luminosity on the stellar mass at $z = 7$. The grey points show the UV luminosity and stellar mass of each simulated luminous halo at $z = 7$. We bin the points according to their stellar mass. The blue points shows the median values for the UV luminosities in each bin, with the errorbars showing the standard deviations in the respective bins. The orange line shows the best-fit power-law relation between L_{UV} and $M_\star$. A best-fit power-law index of 1.22 shows a superlinear dependence of the UV luminosity on stellar mass, which in turns leads to a shallower faint-end slope for the UVLF compared to the low-mass-end of the SMF.

REFERENCES

- Abril-Pla, S., Andreani, V., Carroll, C., et al. 2023, *PeerJ Computer Science*, 9, 1516, doi: <https://doi.org/10.7717/peerj-cs.1516>
- Adams, N. J., Conselice, C. J., Austin, D., et al. 2024, *ApJ*, 965, 169, doi: [10.3847/1538-4357/ad2a7b](https://doi.org/10.3847/1538-4357/ad2a7b)
- Aganze, C., Pearson, S., Starkenburg, T., et al. 2024, *ApJ*, 962, 151, doi: [10.3847/1538-4357/ad159c](https://doi.org/10.3847/1538-4357/ad159c)
- Ansar, S., Pearson, S., Sanderson, R. E., et al. 2025, *ApJ*, 978, 37, doi: [10.3847/1538-4357/ad8b45](https://doi.org/10.3847/1538-4357/ad8b45)
- Arora, A., Sanderson, R. E., Panithanpaisal, N., et al. 2022, *The Astrophysical Journal*, 939, 2
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, 558, A33, doi: [10.1051/0004-6361/201322068](https://doi.org/10.1051/0004-6361/201322068)
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, 156, 123, doi: [10.3847/1538-3881/aabc4f](https://doi.org/10.3847/1538-3881/aabc4f)
- Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, *ApJ*, 935, 167, doi: [10.3847/1538-4357/ac7c74](https://doi.org/10.3847/1538-4357/ac7c74)
- Atek, H., Richard, J., Kneib, J.-P., & Schaerer, D. 2018, *MNRAS*, 479, 5184, doi: [10.1093/mnras/sty1820](https://doi.org/10.1093/mnras/sty1820)
- Atek, H., Richard, J., Jauzac, M., et al. 2015, *ApJ*, 814, 69, doi: [10.1088/0004-637X/814/1/69](https://doi.org/10.1088/0004-637X/814/1/69)
- Atek, H., Labbé, I., Furtak, L. J., et al. 2024, *Nature*, 626, 975, doi: [10.1038/s41586-024-07043-6](https://doi.org/10.1038/s41586-024-07043-6)
- Balbinot, E., Yanny, B., Li, T. S., et al. 2016, *ApJ*, 820, 58, doi: [10.3847/0004-637X/820/1/58](https://doi.org/10.3847/0004-637X/820/1/58)
- Barbá, R. H., Minniti, D., Geisler, D., et al. 2019, *ApJL*, 870, L24, doi: [10.3847/2041-8213/aaf811](https://doi.org/10.3847/2041-8213/aaf811)
- Barkana, R., & Loeb, A. 2001, *PhR*, 349, 125, doi: [10.1016/S0370-1573\(01\)00019-9](https://doi.org/10.1016/S0370-1573(01)00019-9)
- Barry, M., Wetzel, A., Chapman, S., et al. 2023, *MNRAS*, 523, 428, doi: [10.1093/mnras/stad1395](https://doi.org/10.1093/mnras/stad1395)
- Becker, G. D., & Bolton, J. S. 2013, *MNRAS*, 436, 1023, doi: [10.1093/mnras/stt1610](https://doi.org/10.1093/mnras/stt1610)
- Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013a, *ApJ*, 762, 109, doi: [10.1088/0004-637X/762/2/109](https://doi.org/10.1088/0004-637X/762/2/109)
- Behroozi, P. S., Wechsler, R. H., Wu, H.-Y., et al. 2013b, *ApJ*, 763, 18, doi: [10.1088/0004-637X/763/1/18](https://doi.org/10.1088/0004-637X/763/1/18)
- Belokurov, V., Erkal, D., Evans, N. W., Koposov, S. E., & Deason, A. J. 2018, *MNRAS*, 478, 611, doi: [10.1093/mnras/sty982](https://doi.org/10.1093/mnras/sty982)
- Bhatawdekar, R., Conselice, C. J., Margalef-Bentabol, B., & Duncan, K. 2019, *MNRAS*, 486, 3805, doi: [10.1093/mnras/stz866](https://doi.org/10.1093/mnras/stz866)
- Bouwens, R. J., Illingworth, G., Ellis, R. S., et al. 2022a, *ApJ*, 931, 81, doi: [10.3847/1538-4357/ac618c](https://doi.org/10.3847/1538-4357/ac618c)
- Bouwens, R. J., Illingworth, G. D., van Dokkum, P. G., et al. 2022b, *ApJ*, 927, 81, doi: [10.3847/1538-4357/ac4791](https://doi.org/10.3847/1538-4357/ac4791)
- Bouwens, R. J., Oesch, P. A., Illingworth, G. D., Ellis, R. S., & Stefanon, M. 2017, *ApJ*, 843, 129, doi: [10.3847/1538-4357/aa70a4](https://doi.org/10.3847/1538-4357/aa70a4)
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, *ApJ*, 803, 34, doi: [10.1088/0004-637X/803/1/34](https://doi.org/10.1088/0004-637X/803/1/34)
- Bouwens, R. J., Oesch, P. A., Stefanon, M., et al. 2021, *AJ*, 162, 47, doi: [10.3847/1538-3881/abf83e](https://doi.org/10.3847/1538-3881/abf83e)
- Bouwens, R. J., Stefanon, M., Brammer, G., et al. 2023, *MNRAS*, 523, 1036, doi: [10.1093/mnras/stad1145](https://doi.org/10.1093/mnras/stad1145)
- Bowler, R. A. A., Dunlop, J. S., McLure, R. J., & McLeod, D. J. 2017, *MNRAS*, 466, 3612, doi: [10.1093/mnras/stw3296](https://doi.org/10.1093/mnras/stw3296)
- Bowler, R. A. A., Jarvis, M. J., Dunlop, J. S., et al. 2020, *MNRAS*, 493, 2059, doi: [10.1093/mnras/staa313](https://doi.org/10.1093/mnras/staa313)
- Bowler, R. A. A., Dunlop, J. S., McLure, R. J., et al. 2015, *MNRAS*, 452, 1817, doi: [10.1093/mnras/stv1403](https://doi.org/10.1093/mnras/stv1403)
- Boylan-Kolchin, M., Weisz, D. R., Bullock, J. S., & Cooper, M. C. 2016, *MNRAS*, 462, L51, doi: [10.1093/mnrasl/slw121](https://doi.org/10.1093/mnrasl/slw121)
- Boylan-Kolchin, M., Weisz, D. R., Johnson, B. D., et al. 2015, *MNRAS*, 453, 1503, doi: [10.1093/mnras/stv1736](https://doi.org/10.1093/mnras/stv1736)
- Bullock, J. S., & Boylan-Kolchin, M. 2017, *ARA&A*, 55, 343, doi: [10.1146/annurev-astro-091916-055313](https://doi.org/10.1146/annurev-astro-091916-055313)
- Bullock, J. S., & Johnston, K. V. 2005, *ApJ*, 635, 931, doi: [10.1086/497422](https://doi.org/10.1086/497422)
- Chemerynska, I., Atek, H., Furtak, L. J., et al. 2025, arXiv e-prints, arXiv:2509.24881, <https://arxiv.org/abs/2509.24881>
- Cunningham, E. C., Sanderson, R. E., Johnston, K. V., et al. 2022, *ApJ*, 934, 172, doi: [10.3847/1538-4357/ac78ea](https://doi.org/10.3847/1538-4357/ac78ea)
- Dey, A., Najita, J., Filion, C., et al. 2023, arXiv e-prints, arXiv:2306.12302, doi: [10.48550/arXiv.2306.12302](https://doi.org/10.48550/arXiv.2306.12302)
- Donnan, C. T., McLeod, D. J., Dunlop, J. S., et al. 2023, *MNRAS*, 518, 6011, doi: [10.1093/mnras/stac3472](https://doi.org/10.1093/mnras/stac3472)
- Donnan, C. T., McLure, R. J., Dunlop, J. S., et al. 2024, *MNRAS*, 533, 3222, doi: [10.1093/mnras/stae2037](https://doi.org/10.1093/mnras/stae2037)
- Duncan, K., Conselice, C. J., Mortlock, A., et al. 2014, *MNRAS*, 444, 2960, doi: [10.1093/mnras/stu1622](https://doi.org/10.1093/mnras/stu1622)
- Faucher-Giguère, C.-A., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008, *ApJ*, 688, 85, doi: [10.1086/592289](https://doi.org/10.1086/592289)
- Faucher-Giguère, C.-A., Lidz, A., Zaldarriaga, M., & Hernquist, L. 2009, *ApJ*, 703, 1416, doi: [10.1088/0004-637X/703/2/1416](https://doi.org/10.1088/0004-637X/703/2/1416)
- Feldmann, R., Quataert, E., Faucher-Giguère, C.-A., et al. 2023, *MNRAS*, 522, 3831, doi: [10.1093/mnras/stad1205](https://doi.org/10.1093/mnras/stad1205)
- Feldmann, R., Boylan-Kolchin, M., Bullock, J. S., et al. 2025, *MNRAS*, 536, 988, doi: [10.1093/mnras/stae2633](https://doi.org/10.1093/mnras/stae2633)
- Finkelstein, S. L., Ryan, Russell E., J., Papovich, C., et al. 2015, *ApJ*, 810, 71, doi: [10.1088/0004-637X/810/1/71](https://doi.org/10.1088/0004-637X/810/1/71)

- Finkelstein, S. L., D'Aloisio, A., Paardekooper, J.-P., et al. 2019, *ApJ*, 879, 36, doi: [10.3847/1538-4357/ab1ea8](https://doi.org/10.3847/1538-4357/ab1ea8)
- Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., et al. 2023, *ApJL*, 946, L13, doi: [10.3847/2041-8213/acade4](https://doi.org/10.3847/2041-8213/acade4)
- Finkelstein, S. L., Leung, G. C. K., Bagley, M. B., et al. 2024, *ApJL*, 969, L2, doi: [10.3847/2041-8213/ad4495](https://doi.org/10.3847/2041-8213/ad4495)
- Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, *PhR*, 433, 181, doi: [10.1016/j.physrep.2006.08.002](https://doi.org/10.1016/j.physrep.2006.08.002)
- Gandhi, P. J., Wetzell, A., Boylan-Kolchin, M., et al. 2024, *MNRAS*, doi: [10.1093/mnras/stae1584](https://doi.org/10.1093/mnras/stae1584)
- Garrison-Kimmel, S., Wetzell, A., Bullock, J. S., et al. 2017, *MNRAS*, 471, 1709, doi: [10.1093/mnras/stx1710](https://doi.org/10.1093/mnras/stx1710)
- Garrison-Kimmel, S., Hopkins, P. F., Wetzell, A., et al. 2019a, *MNRAS*, 487, 1380, doi: [10.1093/mnras/stz1317](https://doi.org/10.1093/mnras/stz1317)
- Garrison-Kimmel, S., Wetzell, A., Hopkins, P. F., et al. 2019b, *MNRAS*, 489, 4574, doi: [10.1093/mnras/stz2507](https://doi.org/10.1093/mnras/stz2507)
- Grillmair, C. J. 2006, *ApJL*, 645, L37, doi: [10.1086/505863](https://doi.org/10.1086/505863)
- Guzmán, R. 2022, Analysis of Resolved Remnants of Accreted galaxies as a Key Instrument for Halo Surveys. <https://www.cosmos.esa.int/documents/7423467/7423486/ESA-F2-ARRAKIHS-Phase-2-PUBLIC-v0.9.2.pdf/61b363d7-2a06-1196-5c40-c85aa90c2113?t=1667557422996>
- Hahn, O., & Abel, T. 2011, *MNRAS*, 415, 2101, doi: [10.1111/j.1365-2966.2011.18820.x](https://doi.org/10.1111/j.1365-2966.2011.18820.x)
- Harikane, Y., Nakajima, K., Ouchi, M., et al. 2024, *ApJ*, 960, 56, doi: [10.3847/1538-4357/ad0b7e](https://doi.org/10.3847/1538-4357/ad0b7e)
- Harikane, Y., Ouchi, M., Oguri, M., et al. 2023, *ApJS*, 265, 5, doi: [10.3847/1538-4365/acaaa9](https://doi.org/10.3847/1538-4365/acaaa9)
- Harikane, Y., Inoue, A. K., Ellis, R. S., et al. 2025, *ApJ*, 980, 138, doi: [10.3847/1538-4357/ad9b2c](https://doi.org/10.3847/1538-4357/ad9b2c)
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, 585, 357, doi: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2)
- Harshan, A., Bradač, M., Abraham, R., et al. 2024, *MNRAS*, 532, 1112, doi: [10.1093/mnras/stae1574](https://doi.org/10.1093/mnras/stae1574)
- Haywood, M., Di Matteo, P., Lehnert, M. D., et al. 2018, *ApJ*, 863, 113, doi: [10.3847/1538-4357/aad235](https://doi.org/10.3847/1538-4357/aad235)
- Helmi, A., Babusiaux, C., Koppelman, H. H., et al. 2018, *Nature*, 563, 85, doi: [10.1038/s41586-018-0625-x](https://doi.org/10.1038/s41586-018-0625-x)
- Helmi, A., White, S. D. M., de Zeeuw, P. T., & Zhao, H. 1999, *Nature*, 402, 53, doi: [10.1038/46980](https://doi.org/10.1038/46980)
- Hopkins, P. F. 2015, *MNRAS*, 450, 53, doi: [10.1093/mnras/stv195](https://doi.org/10.1093/mnras/stv195)
- Hopkins, P. F., Wetzell, A., Kereš, D., et al. 2018a, *MNRAS*, 480, 800, doi: [10.1093/mnras/sty1690](https://doi.org/10.1093/mnras/sty1690)
- . 2018b, *MNRAS*, 477, 1578, doi: [10.1093/mnras/sty674](https://doi.org/10.1093/mnras/sty674)
- Horta, D., Cunningham, E. C., Sanderson, R. E., et al. 2023, *ApJ*, 943, 158, doi: [10.3847/1538-4357/acae87](https://doi.org/10.3847/1538-4357/acae87)
- Horta, D., Cunningham, E. C., Sanderson, R., et al. 2024, *MNRAS*, 527, 9810, doi: [10.1093/mnras/stad3834](https://doi.org/10.1093/mnras/stad3834)
- Hunter, J. D. 2007, *Computing in Science and Engineering*, 9, 90, doi: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55)
- Ibata, R., Irwin, M., Lewis, G. F., & Stolte, A. 2001, *ApJL*, 547, L133, doi: [10.1086/318894](https://doi.org/10.1086/318894)
- Ishigaki, M., Kawamata, R., Ouchi, M., et al. 2018, *ApJ*, 854, 73, doi: [10.3847/1538-4357/aaa544](https://doi.org/10.3847/1538-4357/aaa544)
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, 873, 111, doi: [10.3847/1538-4357/ab042c](https://doi.org/10.3847/1538-4357/ab042c)
- Johnson, J. W., Conroy, C., Johnson, B. D., et al. 2023, *MNRAS*, 526, 5084, doi: [10.1093/mnras/stad2985](https://doi.org/10.1093/mnras/stad2985)
- Johnston, K. V., Bullock, J. S., Sharma, S., et al. 2008, *ApJ*, 689, 936, doi: [10.1086/592228](https://doi.org/10.1086/592228)
- Koppelman, H. H., Helmi, A., Massari, D., Price-Whelan, A. M., & Starkenburg, T. K. 2019, *A&A*, 631, L9, doi: [10.1051/0004-6361/201936738](https://doi.org/10.1051/0004-6361/201936738)
- Kuhlen, M., & Faucher-Giguère, C.-A. 2012, *MNRAS*, 423, 862, doi: [10.1111/j.1365-2966.2012.20924.x](https://doi.org/10.1111/j.1365-2966.2012.20924.x)
- Laine, S., Martinez-Delgado, D., Trujillo, I., et al. 2018, *arXiv e-prints*, arXiv:1812.04897, doi: [10.48550/arXiv.1812.04897](https://doi.org/10.48550/arXiv.1812.04897)
- Lancaster, L., Pearson, S., Williams, B. F., et al. 2022, *AJ*, 164, 142, doi: [10.3847/1538-3881/ac8a95](https://doi.org/10.3847/1538-3881/ac8a95)
- Leitherer, C., Schaerer, D., Goldader, J. D., et al. 1999, *ApJS*, 123, 3, doi: [10.1086/313233](https://doi.org/10.1086/313233)
- Leung, G. C. K., Bagley, M. B., Finkelstein, S. L., et al. 2023, *ApJL*, 954, L46, doi: [10.3847/2041-8213/acf365](https://doi.org/10.3847/2041-8213/acf365)
- Li, T. S., Koposov, S. E., Zucker, D. B., et al. 2019, *MNRAS*, 490, 3508, doi: [10.1093/mnras/stz2731](https://doi.org/10.1093/mnras/stz2731)
- Livermore, R. C., Finkelstein, S. L., & Lotz, J. M. 2017, *ApJ*, 835, 113, doi: [10.3847/1538-4357/835/2/113](https://doi.org/10.3847/1538-4357/835/2/113)
- Ma, X., Hopkins, P. F., Wetzell, A. R., et al. 2017, *MNRAS*, 467, 2430, doi: [10.1093/mnras/stx273](https://doi.org/10.1093/mnras/stx273)
- Ma, X., Hopkins, P. F., Garrison-Kimmel, S., et al. 2018a, *MNRAS*, 478, 1694, doi: [10.1093/mnras/sty1024](https://doi.org/10.1093/mnras/sty1024)
- Ma, X., Hopkins, P. F., Boylan-Kolchin, M., et al. 2018b, *MNRAS*, 477, 219, doi: [10.1093/mnras/sty684](https://doi.org/10.1093/mnras/sty684)
- Ma, X., Hayward, C. C., Casey, C. M., et al. 2019, *MNRAS*, 487, 1844, doi: [10.1093/mnras/stz1324](https://doi.org/10.1093/mnras/stz1324)
- Madau, P., Haardt, F., & Rees, M. J. 1999, *ApJ*, 514, 648, doi: [10.1086/306975](https://doi.org/10.1086/306975)
- Majewski, S. R., Skrutskie, M. F., Weinberg, M. D., & Ostheimer, J. C. 2003, *ApJ*, 599, 1082, doi: [10.1086/379504](https://doi.org/10.1086/379504)
- Malhan, K., Ibata, R. A., & Martin, N. F. 2018, *MNRAS*, 481, 3442, doi: [10.1093/mnras/sty2474](https://doi.org/10.1093/mnras/sty2474)
- Malhan, K., Yuan, Z., Ibata, R. A., et al. 2021, *ApJ*, 920, 51, doi: [10.3847/1538-4357/ac1675](https://doi.org/10.3847/1538-4357/ac1675)
- Martin, N. F., Ibata, R. A., Starkenburg, E., et al. 2022, *MNRAS*, 516, 5331, doi: [10.1093/mnras/stac2426](https://doi.org/10.1093/mnras/stac2426)

- Mascia, S., Pentericci, L., Calabrò, A., et al. 2024, *A&A*, 685, A3, doi: [10.1051/0004-6361/202347884](https://doi.org/10.1051/0004-6361/202347884)
- McQuinn, K. B. W., B. Newman, M. J., Savino, A., et al. 2024, *ApJ*, 961, 16, doi: [10.3847/1538-4357/ad1105](https://doi.org/10.3847/1538-4357/ad1105)
- Myeong, G. C., Evans, N. W., Belokurov, V., Sanders, J. L., & Koposov, S. E. 2018a, *ApJL*, 856, L26, doi: [10.3847/2041-8213/aab613](https://doi.org/10.3847/2041-8213/aab613)
- . 2018b, *ApJL*, 863, L28, doi: [10.3847/2041-8213/aad7f7](https://doi.org/10.3847/2041-8213/aad7f7)
- Myeong, G. C., Vasiliev, E., Iorio, G., Evans, N. W., & Belokurov, V. 2019, *MNRAS*, 488, 1235, doi: [10.1093/mnras/stz1770](https://doi.org/10.1093/mnras/stz1770)
- Navarro-Carrera, R., Rinaldi, P., Caputi, K. I., et al. 2023, *arXiv e-prints*, arXiv:2305.16141, doi: [10.48550/arXiv.2305.16141](https://doi.org/10.48550/arXiv.2305.16141)
- Necib, L., Lisanti, M., Garrison-Kimmel, S., et al. 2019, *ApJ*, 883, 27, doi: [10.3847/1538-4357/ab3afc](https://doi.org/10.3847/1538-4357/ab3afc)
- Necib, L., Ostdiek, B., Lisanti, M., et al. 2020, *Nature Astronomy*, 4, 1078, doi: [10.1038/s41550-020-1131-2](https://doi.org/10.1038/s41550-020-1131-2)
- Newberg, H. J., Yanny, B., & Willett, B. A. 2009, *ApJL*, 700, L61, doi: [10.1088/0004-637X/700/2/L61](https://doi.org/10.1088/0004-637X/700/2/L61)
- Newberg, H. J., Yanny, B., Rockosi, C., et al. 2002, *ApJ*, 569, 245, doi: [10.1086/338983](https://doi.org/10.1086/338983)
- Oke, J. B. 1974, *ApJS*, 27, 21, doi: [10.1086/190287](https://doi.org/10.1086/190287)
- Panithanpaisal, N., Sanderson, R. E., Wetzel, A., et al. 2021, *ApJ*, 920, 10, doi: [10.3847/1538-4357/ac1109](https://doi.org/10.3847/1538-4357/ac1109)
- Pearson, S., Clark, S. E., Demirjian, A. J., et al. 2022a, *ApJ*, 926, 166, doi: [10.3847/1538-4357/ac4496](https://doi.org/10.3847/1538-4357/ac4496)
- . 2022b, *ApJ*, 926, 166, doi: [10.3847/1538-4357/ac4496](https://doi.org/10.3847/1538-4357/ac4496)
- Pearson, S., Price-Whelan, A. M., Hogg, D. W., et al. 2022c, *ApJ*, 941, 19, doi: [10.3847/1538-4357/ac9bfb](https://doi.org/10.3847/1538-4357/ac9bfb)
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *Journal of Machine Learning Research*, 12, 2825, doi: [10.48550/arXiv.1201.0490](https://doi.org/10.48550/arXiv.1201.0490)
- Perez, F., & Granger, B. E. 2007, *Computing in Science and Engineering*, 9, 21, doi: [10.1109/MCSE.2007.53](https://doi.org/10.1109/MCSE.2007.53)
- Pérez-González, P. G., Costantin, L., Langeroodi, D., et al. 2023, *ApJL*, 951, L1, doi: [10.3847/2041-8213/acd9d0](https://doi.org/10.3847/2041-8213/acd9d0)
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, 641, A6, doi: [10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910)
- Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, *ApJL*, 802, L19, doi: [10.1088/2041-8205/802/2/L19](https://doi.org/10.1088/2041-8205/802/2/L19)
- Rose, J. C., Torrey, P., Villaescusa-Navarro, F., et al. 2025, *ApJ*, 982, 68, doi: [10.3847/1538-4357/adb8e5](https://doi.org/10.3847/1538-4357/adb8e5)
- Sales, L. V., Wetzel, A., & Fattahi, A. 2022, *Nature Astronomy*, 6, 897, doi: [10.1038/s41550-022-01689-w](https://doi.org/10.1038/s41550-022-01689-w)
- Samuel, J., Wetzel, A., Chapman, S., et al. 2021, *MNRAS*, 504, 1379, doi: [10.1093/mnras/stab955](https://doi.org/10.1093/mnras/stab955)
- Samuel, J., Wetzel, A., Tollerud, E., et al. 2020, *MNRAS*, 491, 1471, doi: [10.1093/mnras/stz3054](https://doi.org/10.1093/mnras/stz3054)
- Sanderson, R. E., Wetzel, A., Loebman, S., et al. 2020, *ApJS*, 246, 6, doi: [10.3847/1538-4365/ab5b9d](https://doi.org/10.3847/1538-4365/ab5b9d)
- Santistevan, I. B., Wetzel, A., El-Badry, K., et al. 2020, *MNRAS*, 497, 747, doi: [10.1093/mnras/staa1923](https://doi.org/10.1093/mnras/staa1923)
- Savino, A., Weisz, D. R., Skillman, E. D., et al. 2023, *ApJ*, 956, 86, doi: [10.3847/1538-4357/acf46f](https://doi.org/10.3847/1538-4357/acf46f)
- Schechter, P. 1976, *ApJ*, 203, 297, doi: [10.1086/154079](https://doi.org/10.1086/154079)
- Shapiro, P. R., Giroux, M. L., & Babul, A. 1994, *ApJ*, 427, 25, doi: [10.1086/174120](https://doi.org/10.1086/174120)
- Shipp, N., Drlica-Wagner, A., Balbinot, E., et al. 2018, *ApJ*, 862, 114, doi: [10.3847/1538-4357/aacdab](https://doi.org/10.3847/1538-4357/aacdab)
- Shipp, N., Panithanpaisal, N., Necib, L., et al. 2023, *ApJ*, 949, 44, doi: [10.3847/1538-4357/acc582](https://doi.org/10.3847/1538-4357/acc582)
- Simmonds, C., Tacchella, S., Hainline, K., et al. 2024, *MNRAS*, 527, 6139, doi: [10.1093/mnras/stad3605](https://doi.org/10.1093/mnras/stad3605)
- Sipple, J., & Lidz, A. 2024, *ApJ*, 961, 50, doi: [10.3847/1538-4357/ad06a7](https://doi.org/10.3847/1538-4357/ad06a7)
- Sipple, J., Lidz, A., Grin, D., & Sun, G. 2025, *MNRAS*, 538, 1830, doi: [10.1093/mnras/staf340](https://doi.org/10.1093/mnras/staf340)
- Skillman, E. D., Monelli, M., Weisz, D. R., et al. 2017, *ApJ*, 837, 102, doi: [10.3847/1538-4357/aa60c5](https://doi.org/10.3847/1538-4357/aa60c5)
- Song, M., Finkelstein, S. L., Ashby, M. L. N., et al. 2016, *ApJ*, 825, 5, doi: [10.3847/0004-637X/825/1/5](https://doi.org/10.3847/0004-637X/825/1/5)
- Spergel, D., Gehrels, N., Breckinridge, J., et al. 2013, *arXiv e-prints*, arXiv:1305.5422, doi: [10.48550/arXiv.1305.5422](https://doi.org/10.48550/arXiv.1305.5422)
- Sun, G., Faucher-Giguère, C.-A., Hayward, C. C., & Shen, X. 2023a, *MNRAS*, 526, 2665, doi: [10.1093/mnras/stad2902](https://doi.org/10.1093/mnras/stad2902)
- Sun, G., Faucher-Giguère, C.-A., Hayward, C. C., et al. 2023b, *ApJL*, 955, L35, doi: [10.3847/2041-8213/acf85a](https://doi.org/10.3847/2041-8213/acf85a)
- Sun, G., & Furlanetto, S. R. 2016, *MNRAS*, 460, 417, doi: [10.1093/mnras/stw980](https://doi.org/10.1093/mnras/stw980)
- Van Der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *Computing in Science and Engineering*, 13, 22, doi: [10.1109/MCSE.2011.37](https://doi.org/10.1109/MCSE.2011.37)
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nature Methods*, 17, 261, doi: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2)
- Weisz, D. R., & Boylan-Kolchin, M. 2017, *MNRAS*, 469, L83, doi: [10.1093/mnrasl/slx043](https://doi.org/10.1093/mnrasl/slx043)
- Weisz, D. R., Dolphin, A. E., Skillman, E. D., et al. 2014a, *ApJ*, 789, 148, doi: [10.1088/0004-637X/789/2/148](https://doi.org/10.1088/0004-637X/789/2/148)
- . 2014b, *ApJ*, 789, 147, doi: [10.1088/0004-637X/789/2/147](https://doi.org/10.1088/0004-637X/789/2/147)
- Weisz, D. R., Johnson, B. D., & Conroy, C. 2014c, *ApJL*, 794, L3, doi: [10.1088/2041-8205/794/1/L3](https://doi.org/10.1088/2041-8205/794/1/L3)
- Wetzel, A., & Garrison-Kimmel, S. 2020a, *GizmoAnalysis: Read and analyze Gizmo simulations*, *Astrophysics Source Code Library*, record ascl:2002.015

- . 2020b, HaloAnalysis: Read and analyze halo catalogs and merger trees, Astrophysics Source Code Library, record ascl:2002.014
- Wetzel, A., Hayward, C. C., Sanderson, R. E., et al. 2023, ApJS, 265, 44, doi: [10.3847/1538-4365/acb99a](https://doi.org/10.3847/1538-4365/acb99a)
- Wetzel, A., Samuel, J., Gandhi, P. J., et al. 2025, arXiv e-prints, arXiv:2508.06608, doi: [10.48550/arXiv.2508.06608](https://doi.org/10.48550/arXiv.2508.06608)
- Wetzel, A. R., Hopkins, P. F., Kim, J.-h., et al. 2016, ApJL, 827, L23, doi: [10.3847/2041-8205/827/2/L23](https://doi.org/10.3847/2041-8205/827/2/L23)
- Whitler, L., Stark, D. P., Topping, M. W., et al. 2025, arXiv e-prints, arXiv:2501.00984, doi: [10.48550/arXiv.2501.00984](https://doi.org/10.48550/arXiv.2501.00984)
- Wise, J. H., Demchenko, V. G., Halicek, M. T., et al. 2014, MNRAS, 442, 2560, doi: [10.1093/mnras/stu979](https://doi.org/10.1093/mnras/stu979)
- Wu, Z., & Kravtsov, A. 2024, The Open Journal of Astrophysics, 7, 56, doi: [10.33232/001c.121193](https://doi.org/10.33232/001c.121193)
- Zonca, A., Singer, L., Lenz, D., et al. 2019, The Journal of Open Source Software, 4, 1298, doi: [10.21105/joss.01298](https://doi.org/10.21105/joss.01298)