





























Protoclusters as drivers of stellar mass growth in the early Universe, a case study: Taralay – a massive protocluster at $z \sim 4.57$

Priti Staab ^{1,★} Brian C. Lemaux ^{1,2,★} Ben Forrest ¹ Ekta Shah ¹ Olga Cucciati ³ Lori Lubin ¹
 Roy R. Gal ⁴ Denise Hung ^{2,4} Lu Shen ^{5,6} Finn Giddings ⁴ Yana Khusanova ^{7,8}
 Giovanni Zamorani ³ Sandro Bardelli ³ Letizia Pasqua Cassara ⁹ Paolo Cassata ¹⁰
 Yi-Kuan Chiang ¹¹ Yoshinobu Fudamoto ¹² Shuma Fukushima ¹³ Bianca Garilli ¹⁴
 Mauro Giavalisco ¹⁵ Carlotta Gruppioni ³ Lucia Guaita ¹⁶ Gayathri Gururajan ^{3,17}
 Nimish Hathi ¹⁸ Daichi Kashino ^{19,20} Nick Scoville ²¹ Margherita Talia ^{3,17} Daniela Vergani ³
 and Elena Zucca ³

Affiliations are listed at the end of the paper

Accepted 2024 January 22. Received 2024 January 21; in original form 2023 September 30

ABSTRACT

Simulations predict that the galaxy populations inhabiting protoclusters may contribute considerably to the total amount of stellar mass growth of galaxies in the early universe. In this study, we test these predictions observationally, using the Taralay protocluster (formerly PCI J1001+0220) at $z \sim 4.57$ in the COSMOS field. With the Charting Cluster Construction with VUDS and ORELSE (C3VO) survey, we spectroscopically confirmed 44 galaxies within the adopted redshift range of the protocluster ($4.48 < z < 4.64$) and incorporate an additional 18 galaxies from ancillary spectroscopic surveys. Using a density mapping technique, we estimate the total mass of Taralay to be $\sim 1.7 \times 10^{15} M_{\odot}$, sufficient to form a massive cluster by the present day. By comparing the star formation rate density (SFRD) within the protocluster (SFRD_{pc}) to that of the coeval field ($\text{SFRD}_{\text{field}}$), we find that SFRD_{pc} surpasses the $\text{SFRD}_{\text{field}}$ by $\Delta \log (\text{SFRD}/M_{\odot} \text{yr}^{-1} \text{Mpc}^{-3}) = 1.08 \pm 0.32$ (or $\sim 12 \times$). The observed contribution fraction of protoclusters to the cosmic SFRD adopting Taralay as a proxy for typical protoclusters is $33.5 \text{ per cent}^{+8.0}_{-4.3} \text{ per cent}$, a value $\sim 2\sigma$ higher than the predictions from simulations. Taralay contains three peaks that are 5σ above the average density at these redshifts. Their SFRD is ~ 0.5 dex higher than the value derived for the overall protocluster. We show that 68 per cent of all star formation in the protocluster takes place within these peaks, and that the innermost regions of the peaks encase ~ 50 per cent of the total star formation in the protocluster. This study strongly suggests that protoclusters drive stellar mass growth in the early universe and that this growth may proceed in an inside-out manner.

Key words: galaxies:active – galaxies: clusters: individual: PCI J1001+0220 – galaxies: evolution – galaxies: high-redshift – galaxies: star formation – cosmology:large-scale structure of Universe.

1 INTRODUCTION

The density of the environment that galaxies live in plays an important role in influencing their evolution. In general, the studies of samples of galaxies across cosmic time have shown that the galaxies in the earlier stages of the universe tend to exhibit heightened levels of star formation activity with respect to their lower-redshift counterparts. This activity peaks around $z \sim 2$ and precipitously drops at higher redshifts (Madau & Dickinson 2014 and references therein). However, such studies primarily focus on field galaxies, i.e. those residing in typical environments in the universe. Whether the galaxies in overdense environment follow the same trend is uncertain.

In the local universe, the galaxies in the overdense environment show suppressed star formation activity compared to their field

counterparts. Though the Butcher–Oemler effect (Butcher & Oemler 1984) exists, the fraction of optically blue galaxies increasing in overdense environment at higher redshifts, the galaxies in the overdense environment at $z < 1$ still seem deficient in star formation activity compared to their field counterparts (e.g. Wagner et al. 2017, Hamadouche et al. 2023). In the epoch of $1 < z < 2$, some galaxies in high-density environments display higher star formation activity (see Alberts & Noble 2022 and references therein), though the general trend seems to be that galaxies in high-density environments at these redshifts show suppressed star formation activity relative to their field counterparts (e.g. Gómez et al. 2003; Balogh et al. 2004; von der Linden et al. 2010; Muzzin et al. 2012; Nantais et al. 2017; Lemaux et al. 2019; Tomczak et al. 2019; Chartab et al. 2020; Old et al. 2020). Additionally, the presence of an overabundance of massive quiescent galaxies in overdense environments at these redshifts (e.g. Davidzon et al. 2016, Tomczak et al. 2017) implies that rapid stellar mass (SM) growth occurred at the early stages of cluster assembly

* E-mail: ppatil@ucdavis.edu (PS); brian.lemaux@noirlab.edu (BL)

either through some combination of *in situ* star formation processes and *ex situ* galaxy-galaxy merging activity. It is therefore necessary to observe overdensities at higher redshift to place constraints on the assembly history and evolution of low- and intermediate-redshift cluster populations.

Discovering numerous overdense environments at $z > 2$, though challenging, is the first step towards understanding the behaviour of the galaxies that inhabit them. Over the past two decades, there has been a considerable advancement in the breadth and depth of observations capable of detecting protoclusters – the progenitors of galaxy clusters – that have, in turn, lent themselves to the discovery of 100s of such structures. These nascent galaxy clusters span large areas in the sky (> 10 arcmin; e.g. Chiang, Overzier & Gebhardt 2013, Muldrew, Hatch & Cooke 2015, Contini et al. 2016, Alberts & Noble 2022 and references therein) and have density contrasts relative to the field approximately an order of magnitude less than mature clusters (e.g. Lemaux et al. 2019, 2022, Mei et al. 2023). They are typically defined as a structure that will eventually collapse into a virialized galaxy cluster of mass $\geq 10^{14} M_{\odot}$ at $z \geq 0$ (Overzier 2016), though, in practice, such a definition can be difficult to impose observationally. Detecting these structures requires specialized methods that differ from those used to identify galaxy clusters at lower redshifts. This is because the protoclusters typically cannot be detected by looking for presence of overdensity of redder galaxies and/or a hot medium, methods that are commonly used for finding low- to intermediate- z clusters.

One popular technique for detecting protoclusters involves using rare, more easily observed galaxy populations and/or active galactic nuclei (AGNs) as tracers of massive structures and then probing their surroundings. Tracers include quasars (e.g. Matsuda et al. 2011, Bañados et al. 2013, Adams et al. 2015) although sometimes targeting quasars did not find overdense environments (Overzier 2016 and references therein), radio galaxies (e.g. Venemans et al. 2007; Miley & De Breuck 2008; Wylezalek et al. 2013; Orsi et al. 2016; Shen et al. 2021; Huang et al. 2022), sub-mm galaxies (e.g. Blain et al. 2004, Dannerbauer et al. 2014, Pavesi et al. 2018, Calvi, Castignani & Dannerbauer 2023), ultra-massive galaxies (UMGs; McConachie et al. 2022, Ito et al. 2023, McConachie et al. in prep.) strong Ly α emitters (LAEs; $\lambda 1216\text{\AA}$) (e.g. Ouchi et al. 2005, Higuchi et al. 2019, Fuller et al. 2020, Hu et al. 2021, Yonekura et al. 2022), Ly α blobs (e.g. Li et al. 2022, Ramakrishnan et al. 2023), and strong H α ($\lambda 6563\text{\AA}$) emitters (e.g. Hatch et al. 2011, Cooke et al. 2014, Koyama et al. 2021) or other strong emitters of rest-frame optical lines (e.g. Forrest et al. 2017). All of these tracers are more easily detected than typical star forming galaxies at these redshifts (e.g. Shapley et al. 2003, Le Fèvre et al. 2019).

Until recently, the number of known protoclusters discovered by field spectroscopic surveys of typical star forming galaxies at $z > 2$ was limited (Overzier 2016; Overzier & Kashikawa 2019), but surveys that target typical star forming galaxies are increasing this number (e.g. Steidel et al. 2005; Toshikawa et al. 2012; Le Fèvre et al. 2015; 2016; Shi et al. 2019; 2020; 2020; Shen et al. 2022; Uchiyama et al. 2022; Forrest et al. 2023), providing an opportunity to study how SM buildup proceeds in these structures in the early universe. Such wide and deep galaxy surveys that target normal star-forming galaxies using deep rest-frame ultraviolet (UV) spectra over large cosmic volumes are crucial in order to obtain a more representative sample of protoclusters.

The ongoing discoveries of protoclusters through various surveys are revealing that protoclusters are host to diverse stellar populations. The galaxies in the overdense environment at $z > 2$ seem to display

substantial star formation. A recent study by Lemaux et al. (2022) of ~ 7000 spectroscopically confirmed galaxies found a weak but significant trend of star formation rate (SFR) increasing with denser environment for galaxies at $2 < z < 5$. This trend is a reversal of what is observed at $z < 1.5$ (e.g. Tomczak et al. 2019, Old et al. 2020) and points to accelerated *in situ* SM growth at higher redshift. Large sub-millimeter surveys are also being used to discover protoclusters whose galaxy populations are undergoing incredible amounts of star formation activity, with aggregate SFRs in excess of $10\,000 M_{\odot} \text{ yr}^{-1}$ (see Alberts & Noble 2022 and references therein) more than equivalent volumes in the coeval field (e.g. Greenslade et al. 2018; Miller et al. 2018). However, some protoclusters at high redshifts appear to contain an overabundance of redder or quiescent galaxies (e.g. Lemaux et al. 2014; 2018; Long et al. 2020; Shen et al. 2021; Shi et al. 2021; McConachie et al. 2022; Ito et al. 2023), which implies that both enhancement and suppression of star formation activity is occurring in high-density environments at these redshifts. This diversity underscores the need for the study of a larger sample of protoclusters at higher redshifts in order to understand the earlier stages of cluster formation and galaxy evolution.

Simulations suggest that there is increased star formation activity within higher density environments at higher redshift. In particular, the fraction of SFR per unit volume from protoclusters increases with increasing redshift (Chiang et al. 2017, though see also Muldrew, Hatch & Cooke 2018). According to the finding of Chiang et al. (2017), protoclusters contribute, e.g. ~ 20 per cent of the cosmic star formation rate density (SFRD) at $z \sim 4.5$, despite occupying only 4 per cent of the comoving volume of the universe at this epoch. Their contribution to the cosmic SFRD is predicted to increase to 50 per cent at $z \sim 10$. The higher the redshift of the protocluster, on average, the higher its predicted contribution to the cosmic SFR density. These predictions portray protoclusters as important drivers of SM growth in the early universe and emphasize the need to both probe star formation activity in observational data and understand what drives it.

In order to deepen our understanding of how the surrounding environment influences the process of star formation in protocluster members, it is essential to pursue two complementary approaches: comprehensive studies of individual systems and analysis of large protocluster samples. In this study, we focus on the former approach, and present a detailed study of a massive protocluster located in the COSMOS field at $z \sim 4.57$ (Lemaux et al. 2018), here dubbed ‘Taralay’¹, in order to study the star formation activity of member galaxies in detail. Taralay was the inaugural target of the Charting Cluster Construction with Observations of Redshift Evolution in Large-Scale Environments (ORELSE) and VUDS survey (C3VO; Lemaux et al. 2022). Due to being the highest redshift protocluster² in the C3VO sample Taralay was chosen as the focus of this study. It is the excellent spectroscopic coverage that we obtained for this structure along with the deep and wide panchromatic photometry in the COSMOS field that allows us to accurately measure the SFRD of the galaxies in the protocluster. Here we provide one of the first observational tests of the prediction that the volume-averaged star formation activity is more vigorous at higher redshifts in denser

¹Taralay is a fusion of words Tara and Aalay that mean a star and a house respectively in Marathi, the native language of the first author. Taralay means house of stars.

²Though we are designating Taralay a protocluster in this work, its size, extent, complex structure, and mass may indicate that it is, in fact, a proto-supercluster. Additional structure (e.g. Kakimoto et al. 2023) has also been discovered around Taralay on relatively large scales, and future work will be needed to disambiguate its true nature.

environments than in the field. In the future, this analysis will be expanded to an ensemble of structures detected in the C3VO survey (detailed in Section 2.2).

This paper is organized as follows. In Section 2, we describe the data, Section 3 explains the details of analysis to map out the protocluster, Section 4 describes the characteristics of the protoclusters and its peaks, Section 5 describes the method to obtain the SFRD for the protocluster and coeval field, Section 6 describes the results, Section 7 is a discussion, and lastly, Section 8 presents a summary and a discussion of future directions. We adopt a Λ CDM cosmology with $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_\Lambda = 0.73$, and $\Omega_M = 0.27$. For convenience, SFRs are represented in units of $h_{70}^{-2} M_\star \text{ yr}^{-1}$ throughout the paper where $h_{70} \equiv H_0/70 \text{ km}^{-1} \text{ s Mpc}$.

2 OBSERVATIONS

In this section, we describe the photometry and spectroscopy used in this study specifically in the subsection of the COSMOS field spanning from RA and Dec range of $150.1^\circ < \text{RA} < 150.48^\circ$, $2.21^\circ < \text{Dec} < 2.5^\circ$. This subsection of the COSMOS field was chosen to encompass the entirety of the Taralay protocluster as mapped out by our C3VO observations.

2.1 Photometric data

The Cosmic Evolution Survey (COSMOS) field (Scoville et al. 2007) is a large extragalactic survey that covers an area of about 2 deg^2 on the sky. It is one of the largest and most comprehensive multiwavelength surveys ever conducted. Relevant for this study, the survey includes data from a wide range of telescopes and instruments, including the *Hubble Space Telescope*, the *Spitzer Space Telescope*, the *Galaxy Evolution Explorer* (GALEX) and ground-based telescopes such as the Subaru Telescope, the Canada France Hawai'i Telescope, and the Visible and Infrared Survey Telescope for Astronomy (VISTA; see Capak et al. 2007, Laigle et al. 2016, Weaver et al. 2022 and references therein). These observations cover UV, optical, as well as near-infrared band-passes. Additional observations from the Very Large Array, the Chandra X-Ray Observatory, XMM-Newton, the Herschel Space Observatory cover the radio, X-ray and far-infrared wavelengths. Finally, the COSMOS field will be partially covered by the *James Webb Space Telescope* (JWST) as a part of the COSMOS-Web program (Casey et al. 2023). This makes this field an extremely valuable resource for studying the properties and evolution of galaxies over a wide range of cosmic epochs.

In this study, we utilize three different public photometric catalogs for the COSMOS field, each of which provide progressively deeper data. These catalogs are Capak+07 (Capak et al. 2007; hereafter C07), COSMOS2015 (Laigle et al. 2016; hereafter C15), and COSMOS2020 (Weaver et al. 2022; hereafter C20). Our primary source of photometry is the C20 catalog, which offers the deepest data in the COSMOS field for over 1.7 million objects. For this catalog, we adopt the CLASSIC version (see below for more details). The choices used for masking and deblending are different for each catalog, and adopting a combined set of catalogs allows us to mitigate the effect of these choices. Our secondary source of photometry is the C15 catalog containing more than half million objects followed by C07 that contains over 1.7 million objects. The basics of each catalog are summarized in Table 1. Below we briefly explain the properties of photometric data important for our science.

As mentioned above, we primarily use photometry from C20 as it contains a wealth of data from X-ray to Radio. Of these we use optical/near-IR from Subaru Hyper Suprime-Cam (Subaru-HSC;

Table 1. Properties of photometric catalogs used, such as the number of filters, the wavelength range and depth of data.

Catalog	No. of filters	Wavelength range	Depth of data ^a
COSMOS2020	40	1600–80 000 Å	25.7 in IRAC1
COSMOS2015	33	1600–80 000 Å	24.8 in IRAC1
Capak+2007	14	1600–45 000 Å	24.8 in IRAC1

^a 3σ depth of data in 3 arcsec aperture is directly reported for COSMOS2020 from Weaver et al. (2022) and COSMOS2015 from Laigle et al. (2016). The Capak+2007 IRAC1 depth reported here is 3σ in 3 arcsec aperture that we converted from the 5σ value reported in Capak et al. (2007).

Miyazaki et al. 2018), VISTA InfraRed CAMera (VISTA-VIRCAM; Sutherland et al. 2015), and *Spitzer* (Werner et al. 2004, Sanders et al. 2007) surveys compared to the older catalogs mentioned above. The multiwavelength data ranges from far-UV from *GALEX* (Martin et al. 2005) to Infrared Array Camera (IRAC) channel 4 from *Spitzer* (near-IR) yielding rest-frame wavelength coverage redward of 4000 Å and Balmer breaks at $z \sim 4.57$. Such coverage is crucial for reliably recovering physical properties of galaxies such as SM and SFRs (e.g. Ryan et al. 2014, Faisst et al. 2020).

The C20 photometry is processed in two separate catalogs, CLASSIC and FARMER, which use different methods. As mentioned earlier, we adopt the CLASSIC version for our study as it is consistent with the approach used by other catalogs in the COSMOS field. For this study, we corrected C20 catalog for Milky Way extinction as well as aperture correction using the code provided along with the catalog. We correct the C15 catalog for aperture correction, systematic offsets and Milky Way extinction using formulae 9 and 10 in Laigle et al. (2016). The C07 photometry contains optical and near-IR data in the COSMOS field. We exclude the IRAC channels 3 and 4 mag in C07 from our analysis for the reasons described in Lemaux et al. (2018). The issue seems to be limited to C07 photometry and we do use the data in these bands from C15 and C20. To correct the C07 photometry for zero-point offsets, we subtract the zero-point correction given in table 13 of Capak et al. (2007) from the reported magnitudes in corresponding bands. C20, C15, and C07, all three catalogs contain photometric redshifts in the redshift range of $0 < z < 6$.

2.2 Spectroscopic data

The spectroscopic data used in this study is taken from a variety of different surveys. We start with the VIMOS Ultra-Deep Survey (VUDS; Le Fèvre et al. 2015) as it was through this survey that Lemaux et al. (2018) discovered the target of this work, the protocluster Taralay. Next, we describe the Charting Cluster Construction with VUDS and ORELSE (C3VO) survey that re-targeted this structure. Lastly, we describe ancillary spectroscopic data from the DEIMOS 10k Spectroscopic Survey (Hasinger et al. 2018) and the zCOSMOS Spectroscopic Survey that consists of zCOSMOS-Bright survey (Lilly et al. 2007, 2009) and zCOSMOS-Deep survey (Lilly+ in prep., Diener et al. 2013, 2015).

2.2.1 The VIMOS Ultra-Deep Survey

The VUDS (Le Fèvre et al. 2015) is a large spectroscopic redshift survey designed to study galaxies beyond $z \simeq 2$. With 640 h of observing time on VIMOS spectrograph of Very Large Telescope (VLT), this survey targeted $\sim 10\,000$ faint ($i_{AB} \sim 25$) galaxies over $2 < z \leq 6$ in COSMOS, the Extended Chandra Deep Field South and the 02 h field of the VIMOS VLT Deep Survey 02 h field, spanning a total

of 1 deg² coverage. The VIMOS spectroscopic observations covered wavelength range of 3650–9350 Å spectroscopically confirming LAEs and Lyman break galaxies (LBGs) that do not exhibit Ly α in emission, over $2 < z < 6$, which resulted in a sample of galaxies roughly representative of the star-forming galaxy population at the epochs over the luminosity range of $0.3L_{UV}^* \leq L_{UV} \leq 3L_{UV}^*$ (Lemaux et al. 2022). Such a sample made it possible to study a range of overdense environments at different times in the history of the universe.

Reliability flags were assigned to galaxies with spectroscopic redshifts (z_{spec}) that describe the confidence level of the assigned spectroscopic redshift being correct (Le Fèvre et al. 2015). These flags range from X0 to X4 with higher values of the second digit indicating greater confidence in z_{spec} . X varies from 0, 1, 2, or 3, where 0 denotes targeted galaxies, 1 denotes a type-1 AGN, 2 denotes a serendipitous detection that is separated in (projected) location from the target and 3 denotes a serendipitous detection that has the same apparent location as the target. Additionally, a flag of X9 is assigned to a galaxy whose spectrum shows a single spectral feature. While the X1 flag represents 41 per cent probability of the assigned z_{spec} being correct, flags X2/X9 and X3/X4 indicate ~ 70 per cent and ~ 99.3 per cent probability of the assigned z_{spec} being correct, respectively. We refer the reader to Lemaux et al. (2022) for more details. We adopt flags X2, X3, X4, and X9 from the VUDS observations as secure³ flags.

We utilize a total of 709 objects from the VUDS survey over our adopted sky region: $150.1^\circ < \text{RA} < 150.48^\circ$, $2.21^\circ < \text{Dec} < 2.5^\circ$ and redshift range of $0 < z < 6$. 453 of these entries have secure flags.

The VUDS survey was monumental for the discovery of ~ 50 spectroscopic overdensities (e.g. Cucciati et al. 2014; Lemaux et al. 2014; 2018; 2018; Shen et al. 2021; Forrest et al. 2023; Shah et al. 2023). Discovered by Lemaux et al. (2018), Taralay protocluster is the highest redshift structure identified through the VUDS survey.

2.2.2 The C3VO survey

The Charting Cluster Construction with VUDS and ORELSE (C3VO) survey was devised as an extension to higher redshift of the ORELSE survey (Lubin et al. 2009), aimed at studying a statistical sample of groups and clusters at $z \sim 1$. The C3VO survey is dedicated to mapping out the structures previously discovered through VUDS at $2 < z < 5$ in a manner consistent with that which ORELSE mapped out intermediate redshift structures. The survey was devised to both perform detailed studies of individual protoclusters and their populations and to statistically connect progenitor protoclusters to their descendent clusters. The main objective of C3VO is to better understand the relationship between SM, star formation, AGNs activity, and local environmental density and in turn to probe the evolution of galaxies in large-scale structures across cosmic times from $z = 5$ to $z = 0.6$.

C3VO is an ongoing campaign targeting protoclusters with the DEep Imaging Multi-Object Spectrograph (DEIMOS; Faber et al. 2003) and the Multi-Object Spectrometer For Infra-Red Exploration

(MOSFIRE; McLean et al. 2008) on the Keck Telescopes, Simultaneous Color Wide-field Infrared Multi-object Spectrograph (Konishi et al. 2012), Multi-Object InfraRed Camera and Spectrograph (Ichikawa et al. 2006) on the Subaru Telescope, and Wide Field Camera 3 (Kimble et al. 2008) F160W imaging and G141 grism on the *Hubble Space Telescope*. We focus on the Keck observations of the C3VO survey that are designed to further map out the six most prominently detected protocluster environments in VUDS at $2 < z < 5$, including Taralay at $z \sim 4.57$, and to target similar types of galaxies that do not have a spectroscopic redshift from rest-frame UV surveys in the field.

The Taralay protocluster at redshift $z \sim 4.57$ was targeted with six masks with DEIMOS in order to acquire rest-frame UV spectra of prospective member galaxies. The highest priority targets were limited to $i_{AB} \leq 25.3$ in order to obtain continuum redshifts, corresponding to L_{FUV}^* at $z \sim 4.5$, where L^* is the characteristic luminosity (Bouwens et al. 2007). We also included fainter objects by extending the limit to $i_{AB} \approx 26.7$ to acquire redshifts from Ly α emission. The observing details for each mask can be found in Table 2.

For this campaign, the targets were selected using photometric redshifts (z_{phot}) from C15 catalog. The targeting priorities and the selection criteria used for the first two masks are described in Lemaux et al. (2022). The same targeting priorities and the selection criteria were largely used for the four remaining masks as well. In addition to the original targeting, our last two DEIMOS masks, dongA1 and dongA2, included targets from the UV-selected Atacama Large Millimeter Array (ALMA) Large Program to Investigate C⁺ at Early Times survey (ALPINE-ALMA, Faisst et al. 2020; Béthermin et al. 2020; Le Fèvre et al. 2020). The ALMA observations of these targets revealed close companions with [CII] 158 μm emission (see Le Fèvre et al. 2020 and Ginolfi et al. 2020). These companions lacked rest-frame UV spectral information and were targeted by DEIMOS in an attempt to recover that information.

With the total integration time of approximately 28 h for all masks (the time per mask is given in Table 2), we obtained 204 secure redshifts. 44 of these galaxies are in the fiducial redshift range of Taralay, $4.48 < z < 4.64$ (see Section 3.4), making them potential Taralay protocluster members.⁴ These galaxies have ≥ 95 per cent probability of the assigned z_{spec} being correct. The C3VO-DEIMOS data were reduced following the method described in Lemaux et al. (2022) using a modified version of spec2D (Cooper et al. 2012, Newman et al. 2013). A modified version of the ZPSEC tool was used to assign redshifts by two independent users with a flagging code that is similar to that of the DEEP2 redshift survey (Davis et al. 2003, Newman et al. 2013). Secure flags –1, 3–, and 4 were assigned where the flag –1 denotes a star whereas flags 3 and 4 denote ≥ 95 per cent probability of the assigned z_{spec} being correct for a galaxy. For this study, we adopt flags 3 and 4 as being secure extragalactic redshifts.

Having the secure redshifts for 44 galaxies makes this structure one of the most extensively studied at such high redshifts. Based on the additional redshifts from the C3VO campaign, we adopt the

³Secure flags are the flags that represent the probability of the assigned z_{spec} being correct to higher than 70 per cent. For the VUDS, DEIMOS10k and zCOSMOS surveys the secure flags are X2, X3, X4, X9. For the C3VO survey, secure flags are 3 and 4, which represent ≥ 95 per cent probability of the assigned z_{spec} being correct.

⁴The term potential member is used here due to these galaxies satisfying two out of three criteria adopted in this work to define true membership of the Taralay protocluster. The two criteria are: (1) the RA, Dec range of $150.1^\circ < \text{RA} < 150.48^\circ$, $2.21^\circ < \text{Dec} < 2.5^\circ$ and (2) redshift range of $4.48 < z < 4.64$. This potential member sample will be further refined by the third criterion used to define true membership, an overdensity (σ_3) cut (see Section 3.4). The total number of galaxies that satisfy the above two criteria as well as the number of galaxies that ultimately qualify as true protocluster members are given in Table 3.

Table 2. Observing details of the C3VO campaign targeting Taralay protocluster with DEIMOS on Keck. The details include mask details, total integration time, seeing, grating, order blocking filter, and central wavelength.

Mask	Center of the mask	PA	Total integration time	Seeing ^a	Grating	Order blocking filter	λ_c
dongN1	10:01:22.9, 2:21:41.6	90.0	3 h 30 min	$\sim 0.6\text{--}0.8$ arcsec	600 l mm ⁻¹	GG400	6500 Å
dongS1	10:01:02.3, 2:17:20.0	90.0	4 h 35 min	$\sim 0.6\text{--}0.8$ arcsec	600 l mm ⁻¹	GG400	6500 Å
dongD1	10:01:27.4, 2:19:27.0	50.0	5 h 20 min	$\sim 0.7\text{--}1.3$ arcsec	600 l mm ⁻¹	GG455	7200 Å
dongD2	10:01:05.4, 2:22:49.9	50.0	6 h	$\sim 0.7\text{--}1$ arcsec	600 l mm ⁻¹	GG455	7200 Å
dongA1	10:01:29.0, 2:21:10.0	25.0	4 h 59 min	$\sim 0.6\text{--}0.8$ arcsec	600 l mm ⁻¹	GG455	7200 Å
dongA2	10:00:44.6, 2:16:08.7	78.0	4 h	$\sim 0.45\text{--}0.55$ arcsec	600 l mm ⁻¹	GG455	7200 Å

^a No meaningful cloud coverage for the duration of observations for all masks.

redshift range of $4.48 < z < 4.64$ for the Taralay protocluster, one of the criteria that we use to define this structure. Overall, we observed 716 objects with C3VO-DEIMOS for this study over our adopted sky region and in the redshift range of $0 < z < 6$, including serendipitous detections, of which 204 have secure extragalactic redshifts.

2.2.3 Ancillary spectroscopic data

Some of the ancillary spectroscopic data for this project comes from the DEIMOS 10k Spectroscopic Survey (Hasinger et al. 2018). For this survey, 10 718 objects were observed with DEIMOS in the COSMOS field over the redshift range $0 < z < 6$. Two or more spectral features were observed for 6617 objects, while 1798 objects have spectra with a single spectral feature that was consistent with the photometric redshift. The magnitude distribution of objects targeted in this survey peaks at $I_{AB} \sim 23$. We utilized a total of 1161 entries from the DEIMOS10k catalog over our adopted sky region, with redshifts ranging from $0 < z < 6$ for this study. 840 of these entries have secure quality flags. Some of these objects show broad lines in their spectrum. This information is represented by adding 10 to their quality flags, i.e. 11–14, 19.

The majority of the ancillary spectroscopic data is obtained from the zCOSMOS Spectroscopic Survey, a large VLT/VIMOS redshift survey in the COSMOS field, with the vast majority of the galaxies in the redshift range $0 < z < 3$. This survey consists of zCOSMOS-Bright survey (Lilly et al. 2007, 2009) and zCOSMOS-Deep survey (Lilly+ in prep., Diener et al. 2013, 2015). The zCOSMOS-Bright survey spans 1.7 deg^2 COSMOS ACS field, targeting 20 000 galaxies over $0.1 \leq z \leq 1.2$ range with $I_{AB} < 22.5$. The zCOSMOS-Deep survey spans central 1 deg^2 and targets 10 000 galaxies over $1.4 \leq z \leq 3$ that were selected through colour-selection criteria. The flagging system for zCOSMOS survey is effectively the same as that of VUDS survey. We used an updated version of the zCOSMOS catalog provided by one of the authors (DK) that changed the assigned redshift for a small number of entries (~ 1 per cent of the secure spectral redshifts) and reevaluated the assigned flags. Some of the entries that were previously assigned flags X2 and lower or X9 received a demoted flag. This catalog was previously used in Kashino et al. (2022) and will be described more fully in an upcoming paper. We utilize a total of 3637 entries in our fiducial region of interest, with 2088 of these entries having secure flags (i.e. X2, X3, X4, and X9). In addition, we include a small number of galaxies from Casey et al. (2015) and Chiang et al. (2015).

The VUDS and zCOSMOS flagging system has well-measured reliabilities and form the basis of our statistical framework. To establish whether the confidence intervals for DEIMOS10k flags are the same as those of VUDS and zCOSMOS surveys, we compare their flagging system. This check is important because in our statistical framework that combines spectroscopic and photometric redshifts (explained

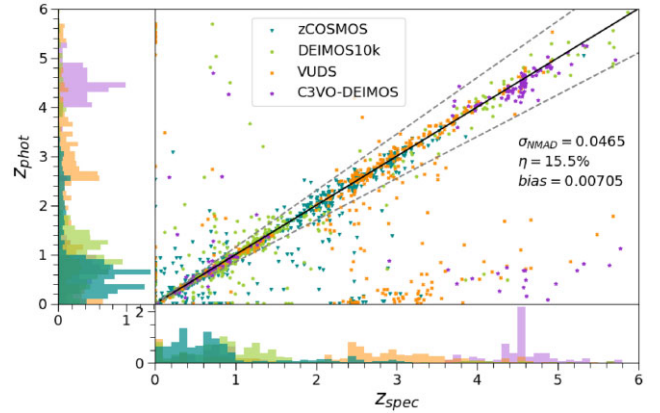


Figure 1. Comparison of photometric redshifts, z_{phot} from COSMOS2020, COSMOS2015 catalogs and Capak+2007 photometry with spectroscopic redshifts obtained from C3VO, VUDS, DEIMOS10k, and zCOSMOS surveys. The galaxies for which this comparison is performed include serendipitous C3VO detections and have either quality flags of 3, 4 (C3VO) or X2–X4 and X9 (VUDS, DEIMOS10k, and zCOSMOS) ensuring high confidence level in the assigned z_{spec} . See Section 2.2 for the discussion about quality flags. The z_{phot} and z_{spec} histograms are normalized such that the total area under the histogram sums up to 1; so the height of each bar in the histogram corresponds to the probability density rather than the count of redshifts in that bin. The σ_{NMAD} , catastrophic outlier rate ($\eta = |z_{\text{spec}} - z_{\text{phot}}| / (1 + z_{\text{spec}}) > 0.15$), and bias which is defined as the median of the difference between z_{phot} and z_{spec} are shown in the main panel.

in Section 3.2), the spectroscopic redshifts are handled according to their quality flags. To make this check, we examine a sample of 70 DEIMOS10k targets with flag=X2/X9, which were observed in other surveys and assigned quality flag=X3/X4 therein. This process involves assessing the catastrophic outlier rate, which is calculated as $|z_{\text{other}} - z_{\text{DEIMOS10k}}| / (1 + z_{\text{other}}) > 5\sigma_{\text{NMAD}}$ where NMAD is the normalized median absolute deviation of $z_{\text{other}} - z_{\text{DEIMOS10k}} / (1 + z_{\text{other}})$. Out of 70 objects, we find that 15 have a catastrophic outlier rate greater than $5\sigma_{\text{NMAD}}$, resulting in a reliability of 79.6 per cent. We repeat this process for 751 DEIMOS10k targets with flag=X3/X4, comparing them with spectroscopic redshifts from another survey also with flag=X3/X4. In this case, we identify 38 objects with a catastrophic outlier rate greater than $5\sigma_{\text{NMAD}}$, corresponding to a reliability of 94.9 per cent. These results broadly align with the flagging system employed by the VUDS and zCOSMOS surveys, indicating that the DEIMOS10k survey shares the same reliability.

In Fig. 1, we show a comparison of the z_{spec} from the surveys mentioned above and z_{phot} from C20, C15, and C07 that are utilized in this study. The galaxies included for this comparison are brighter in IRAC channel 1 and/or 2 than the completeness limits given in Weaver et al. (2022; i.e. 25.7 and/or 25.6 in the (3.6) and (4.5)

bands, respectively] in order to mitigate the effects of Malmquist bias (Malmquist 1922, 1925). This cut is also made for our entire spectroscopic and photometric sample.

When combining data from these surveys, entries need to be resolved. This process of identifying and resolving duplicates is explained in Section 3.1. The number of spectroscopically confirmed unique galaxies before making an IRAC channel 1 and/or IRAC channel 2 cut, which fall within the range of the protocluster redshift range, $z_{pc} = 4.48 < z < 4.64$ (see Section 3.4), and the spectroscopically confirmed unique galaxies that fall within the field redshift range $z_{field} = 4.2 < z < 4.48, 4.64 < z < 4.93$ (see Section 3.4) are listed in Table 3. The unique entries in z_{pc} range are shown in both panels of Fig. 2. In addition, the left-hand panel of this figure also shows the DEIMOS masks given in Table 2 and used for targeting the Taralay protocluster in the C3VO campaign. The right-hand panel of the figure additionally shows unique entries in z_{field} range.

Applying the conditions outlined in 3.4 on these unique entries gives us the final protocluster and the field sample given in Table 3. All galaxies, except the ones from C3VO survey, have quality flags of X2, X3, X4, or X9 indicating the high confidence in their redshift measurements. The galaxies included from C3VO survey have quality flags of 3 and 4, indicating more than 95 per cent confidence in the assigned spectroscopic redshift. As such, the flagging systems of C3VO, VUDS, DEIMOS10k, and zCOSMOS are all compatible with each other in terms of the likelihood of a spectroscopic redshift of a given flag to be correct. This homogeneity is crucial when we utilize the spectroscopic information in the various Monte Carlo processes in our analysis (see next section). Overall we utilize a catalog with total 6629 spectral entries. After undergoing the IRAC cut described earlier ($[3.6] < 25.7$ and/or $[4.5] < 25.6$), there are 115 and 3169 unique secure spectral redshifts over the redshift ranges of $4.2 < z < 4.93$ and $0 < z < 6$, respectively, drawn from all surveys.

3 METHODS

The primary objective of this study is to examine the rate of *in situ* SM growth in the Taralay protocluster at $z \sim 4.57$ compared to that of the field. We use the spectral energy distribution (SED) fitting method (explained in Section 5.1) to determine the physical properties such as SM, SFR for all galaxies. However, before we can apply the SED fitting, it is necessary to assemble a catalog of high redshift galaxies with z_{spec} and find their photometric counterparts. We explain this procedure in Section 3.1. Though we have extensive spectroscopy, the galaxies with z_{spec} may not be representative of the full underlying galaxy population and the galaxies without spectral redshifts far outnumber those with z_{spec} . To utilize all the data available, we create a framework, explained in Section 3.2 that statistically combines z_{spec} and z_{phot} . We then use the Voronoi Tessellation Monte Carlo (VMC; Lemaux et al. 2017; Tomczak et al. 2017; 2018; Cucciati et al. 2018; Hung et al. 2020; 2022) map technique that relies on this statistical framework to reconstruct the density field in Section 3.3. The reconstruction of the density field allows for an estimation of the local overdensity, mapping out the structure and getting the volume and mass of Taralay. We then use the overdensity information to update the criteria for a galaxy to qualify as the protocluster or field galaxy in Section 3.4.

3.1 Catalog matching procedure

To identify the photometric counterparts for the spectroscopic data, an angular separation within which a match can be found is

determined based on the sky error.⁵ To choose the matching radius appropriate given the unknown sky error associated with all the catalogs, we start by matching with a radius of 1 arcsec and look at the distribution of the angular separation of the closest match between galaxies with z_{spec} and the photometric counterparts they matched to. A local minimum was observed at 0.3 arcsec, which strongly indicates that the matches at distance of < 0.3 arcsec are likely genuine and the matches that are at > 0.3 arcsec distances are likely contaminated with impurities.

Starting with the spectroscopic data from C3VO, VUDS, DEIMOS10k, and zCOSMOS surveys within our adopted sky region, we first look for photometric counterparts in C20 which contains the deepest data to date that covers the entirety of the COSMOS field. If there are no photometric counterparts within a circle of 0.3 arcsec radius for a z_{spec} entry, the search is expanded to include the C15 and C07 catalogs.

Upon conducting this search for the 6229 total entries in our spectral catalog over a redshift range of $0 < z < 6$, we found that C20 has at least one counterparts for 89.4 per cent of the total z_{spec} entries, with 7.1 per cent of the total z_{spec} entries matching to two photometric sources and approximately 0.4 per cent of entries matching to more than two photometric sources. C15 counterparts were found for the 52.93 per cent of the remaining 614 z_{spec} entries (5.22 per cent of the total z_{spec} entries), with 2.60 per cent entries matching to two photometric sources and 0.16 per cent of entries matching to more than two photometric sources. Finally, of the 280 z_{spec} entries that had no counterparts in C20/C15, C07 photometry counterparts were identified for 51.78 per cent of the remaining z_{spec} entries (2.32 per cent of the total z_{spec} entries), 0.71 per cent of entries matching to two photometric sources. No photometric counterparts were found for 135 (2.17 per cent of the total z_{spec}) entries.

In situations where there are multiple z_{spec} entries that correspond to a single z_{phot} counterpart, a set of rules was established to choose the most likely counterpart and avoid duplicates. These rules are as follows:

- (i) If the z_{spec} entries have different redshift quality flags, the galaxy with the most secure redshift quality flag is given priority.
- (ii) If the z_{spec} entries come from different spectroscopic surveys but have the same redshift quality flags, then the priority order is as follows: C3VO-DEIMOS entries, then VUDS entries, followed by DEIMOS10k entries, and lastly zCOSMOS entries. For flags X2, X9, the priority order is VUDS observations, then DEIMOS10k observations, followed by zCOSMOS survey observations, and lastly C3VO-DEIMOS observations, as lower quality flags from C3VO were considered unreliable.
- (iii) zCOSMOS-Deep was prioritized over zCOSMOS-Bright when there were only zCOSMOS entries and the flags and z_{spec} were effectively identical.

We also looked at some cases by eye, where both the matches had identical z_{spec} values and very similar quality flags, to confirm that these rules result in a match that is the most sensible in each case. By following these rules, the appropriate match is selected and duplicates are resolved when multiple spectroscopic redshifts are associated with a single photometric counterpart.

⁵The sky error represents the potential discrepancy between the measured coordinates of an astronomical source and its true celestial coordinates. The sky error can arise from various factors, including instrumental limitations, atmospheric effects, and the accuracy of the astrometric calibration.

Table 3. The table shows the total number of secure spectroscopic redshifts (z_{spec}) from the four surveys that satisfy the various criteria we impose throughout this study.

Survey	Total entries in z_{pc}^a	Unique entries in z_{pc}^b	PC members ^c	Total entries in z_{field}^d	Unique entries in z_{field}^e	Field galaxies ^f
C3VO-DEIMOS	44	44	34	46	46	41
VUDS	16	11	7	16	15	13
DEIMOS10k	15	7	2	24	15	13
zCOSMOS	22	0	0	27	0	0

^a The total number of secure z_{spec} galaxies that fall in the protocluster redshift range, $z_{pc} = 4.48 < z < 4.64$. ^b The unique secure z_{spec} entries from each survey in z_{pc} range. ^c The number of secure z_{spec} galaxies from each survey that satisfy the final criterion for being a protocluster member where the final criterion is residence in a region that is $\sigma_\delta \geq 2$ overdense than the field and also contains a $\sigma_\delta \geq 5$ peak. ^d The total number of secure z_{spec} galaxies that fall in the field redshift range, $z_{\text{field}} = 4.2 < z < 4.48, 4.64 < z < 4.93$. ^e The unique secure z_{spec} entries from each survey in z_{field} range. ^f The final number of secure z_{spec} field galaxies. We exclude any galaxy as a field galaxy if it falls in a protocluster-like structure i.e. a region that is $\sigma_\delta \geq 2$ overdense than the field and also contains a $\sigma_\delta \geq 5$ peak.

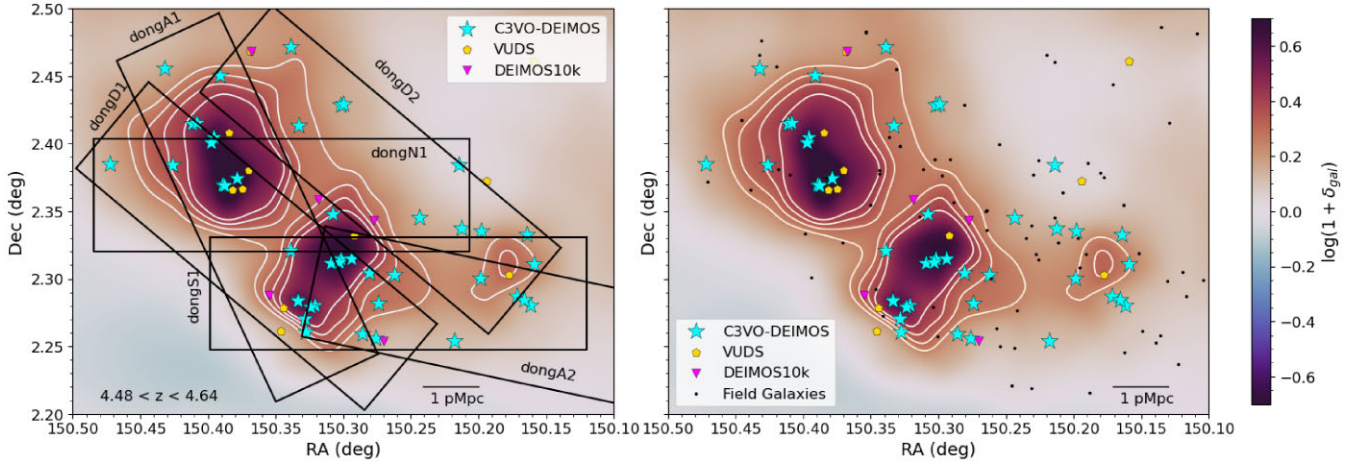


Figure 2. Both the panels of this figure show a sky plot for the Taralay protocluster at $z \sim 4.57$. The cyan stars, yellow pentagons and fuchsia inverted triangles are spectroscopically confirmed potential members (see Section 2.2.2 for the definition of potential members) of Taralay observed as a part of C3VO survey, VUDS survey and DEIMOS10k survey, respectively. The fiducial redshift range for the structure, $4.48 < z < 4.64$ is indicated at the bottom left corner. The black bar indicating 1 proper Mpc towards the right corner roughly corresponds to 0.04° . The color-scale varies with overdensity $\log(1 + \delta_{\text{gal}})$ such that the darker the region, the denser is the environment at those sky coordinates. In the left-hand panel, the black rectangles laid on the protocluster are six masks that we observed this structure with, detailed in Table 2. The right-hand panel of this figure shows unique field galaxies from C3VO, VUDS, DEIMOS10k surveys in the redshift range of $4.2 < z < 4.48$ and $4.64 < z < 4.93$. We note that the outermost contour that roughly marks the shape of Taralay in both panels of this figure does not correspond to the $\sigma_\delta \geq 2$ boundary of Taralay in the right-hand panel of Fig. 3 due to the difference in methods with which these two figures are constructed.

We conducted a comparison test between our selection method and an alternative method developed by Forrest et al. (2023) to verify the accuracy of our choice of a 0.3 arcsec radius for identifying the correct z_{phot} counterpart for each z_{spec} entry. This alternative method involves finding the nearest match for each z_{spec} entry, updating the coordinates of the z_{spec} entries by calculating the median positional offsets between the spectroscopic survey and photometric catalog, and determining new z_{phot} counterparts using the updated coordinates. This is necessary to account for differences in astrometry between the catalogs.

After resolving duplicates, we found that our selection method agreed with the alternative method for 99.3 per cent of the z_{spec} entries. For the remaining ~ 0.7 per cent of z_{spec} entries, the alternative method provided better matches outside the 0.3 arcsec radius. For those ~ 0.7 per cent z_{spec} entries, the matches resulting from the alternative method were adopted.

In addition to the spectroscopic data described in Section 2.2, we take a comprehensive approach of incorporating all of the available photometric data for our study. This includes 36 232 entries from the C20 catalog within our adopted sky region that survived the

IRAC cut. By incorporating this additional data, we can leverage the extensive photometric information contained in the C20 catalog to complement and enrich our analysis. The completed master catalog contains different types of galaxies, including those with spectroscopic redshifts (z_{spec}) from various surveys, their photometric counterparts with associated z_{phot} values, and galaxies with photometric information.

3.2 Statistical framework with Monte Carlo

As stated earlier, we utilize all the available data, objects with z_{spec} as well as z_{phot} for this analysis since the galaxies with z_{spec} may not be representative of the full underlying galaxy population. To establish our framework that statistically combines z_{spec} and z_{phot} , we refer to the statistical model described in appendix A of Lemaux et al. (2022) and perform Monte Carlo on redshifts for 100 iterations. We begin with the output catalog from Section 3.1 with combined spectroscopic and photometric data. The statistical framework described below is necessary to map overdense structures using the density

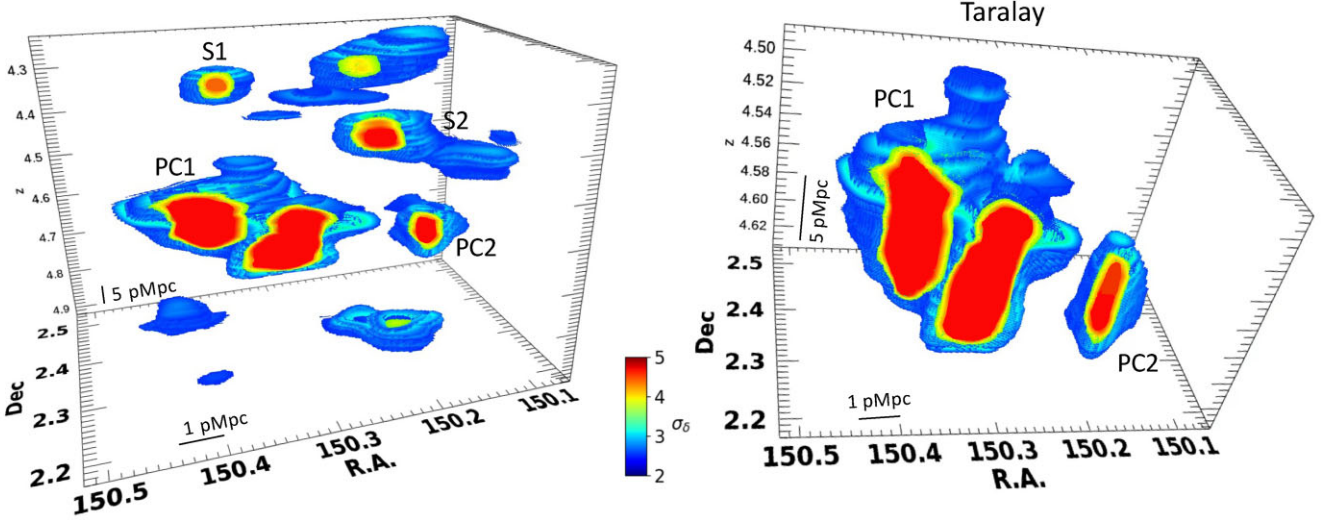


Figure 3. Reconstructed galaxy overdensity using the VMC mapping technique. The panel on the left shows the Taralay protocluster and other overdensities within the volume that contains the coeval field sample. The blue regions are bounded by overdensity isopleths of $\sigma_\delta \geq 2$. The right-hand panel of the figure shows the 3D structure of only the Taralay protocluster. The red regions have an overdensity of $\sigma_\delta \geq 5$ relative to the field, whereas the blue regions set the $\sigma_\delta \geq 2$ boundary. The different colours and their corresponding overdensity values are shown by the colour scale that represents the galaxy overdensity. Scale bars representing proper distances of varying size are shown in each panel.

mapping technique (Section 3.3), to perform SED fitting (Section 5.1) and obtain the SFRD of the Taralay protocluster (Section 5.2).

For each Monte Carlo realization of a galaxy, we sample a value, ξ , from likelihood probability density function based on the spectral quality flag, with the likelihood representing the chance that the spectroscopic redshift is correct. The likelihood values and their associated uncertainties for quality flags are given in appendix A of Lemaux et al. (2022). Next, we sample a value, χ , from the uniform distribution between 0 and 1. If $\chi < \xi$, z_{spec} is retained as the redshift for that galaxy in that iteration. If $\chi \geq \xi$, then we assign a z_{phot} drawn from the redshift probability density function (zPDF) constructed as an asymmetric Gaussian using the median and photometric redshift confidence intervals given in the catalog of the photometric counterpart. The authors of the photometric catalogs have derived the confidence intervals by performing SED fitting using the code LePhare (Arnouts et al. 2002, Ilbert et al. 2006).

For photometric objects, with no measured z_{spec} , we sample an asymmetric Gaussian distribution based on the zPDF to obtain redshifts instead of relying on the z_{phot} assigned to them. The final master catalog encompasses galaxies with z_{spec} from different surveys, their photometric counterparts with z_{phot} , and galaxies with z_{phot} information but no spectroscopy, all satisfying $4.2 < z < 4.93$ redshift range as that is our redshift range of interest (see Section 3.4).

3.3 Density mapping and the size of Taralay

To map out Taralay and measure the underlying density field, we use the VMC map technique. The VMC method is a statistical approach that employs Voronoi tessellation to estimate the density of galaxies in a given region of the sky but does so over a large number of Monte Carlo realizations of the input data. Voronoi tessellation divides space into polygons around each galaxy, with each polygon containing all the points in space that are closer to that galaxy than any other. The area of each polygon is inversely proportional to the local density of galaxies around the corresponding galaxy. The VMC technique is explained in great detail in Lemaux et al. (2017, 2018, 2019, 2022), Tomczak et al. (2017, 2019), Shen et al. (2017, 2018, 2019, 2021),

Cucciati et al. (2018), Pelliccia et al. (2019), Hung et al. (2020), and Hung et al in prep. We adopt the version outlined in Lemaux et al. (2022) for this work. As a result of this process, we get a 3D cube with a local overdensity at every single voxel, which is defined as

$$\log(1 + \delta_{\text{gal}}) = \log\left(1 + \frac{(\Sigma - \tilde{\Sigma})}{\tilde{\Sigma}}\right) \quad (1)$$

where δ_{gal} is the local galaxy overdensity. We can then assign $\log(1 + \delta_{\text{gal}})$ values to galaxies by tethering a galaxy to the voxel which contains the RA, Dec, and redshift of the galaxy.

We fit the distribution of $\log(1 + \delta_{\text{gal}})$ with a Gaussian for each redshift slice of depth 7.5 pMpc and obtain its μ and σ . The $\sigma(z)$ and $\mu(z)$ are then fitted as a function of redshift with a fifth-order polynomial from which we obtain μ_δ and σ_δ , where

$$\sigma_\delta = \frac{\log(1 + \delta_{\text{gal}}) - \mu(z)}{\sigma(z)} \quad (2)$$

Both $\mu(z)$ and $\sigma(z)$ are in units of $\log(1 + \delta_{\text{gal}})$. Over $4.2 < z < 4.93$, $\mu(z)$ is ~ 0 and $\sigma(z) \sim 0.1$. This methodology is explained in more detail in Cucciati et al. (2018). For the rest of this paper, we adopt σ_δ to describe the overdensity.

By plotting the regions above a certain σ_δ , we can visualize the density field and map out the protocluster as well as any additional structures that may exist along the line of sight (LoS) as shown in Fig. 3. The left-hand panel in this figure shows all overdense structures with $\sigma_\delta \geq 2$ along the LoS. The right-hand panel of the figure shows only the protocluster Taralay. The full distribution of the σ_δ values across the redshift range of interest is shown in Fig. 4. The resulting underlying galaxy density field is converted into a matter density field using a bias factor (see Appendix A1 for more discussion). For this study, we use a bias factor of $b = 3.6$, which is based on previous works (Chiang et al. 2013, Durkalec et al. 2018).

3.4 Definition of field and protocluster sample

Here, we outline the full criteria used to define our protocluster and the coeval field. To ensure a fair and unbiased comparison of the

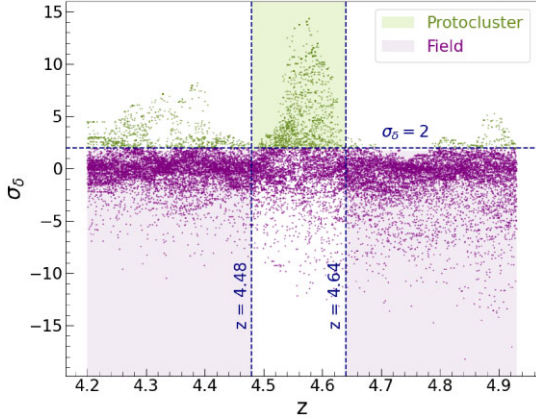


Figure 4. This figure shows the spread of σ_δ (overdensity) values associated with each galaxy that forms the final master catalog across all MC iterations (see Section 3.2). The green dots represent galaxies with $\sigma_\delta \geq 2$, while the purple dots represent galaxies with $\sigma_\delta \leq 2$. The green region, bounded by $4.48 < z < 4.64$ and $\sigma_\delta \geq 2$, represents galaxies that qualify as protocluster members. The two purple shaded regions show galaxies that qualify as coeval field galaxies. For more discussion, see Section 3.4.

SFRD between the protocluster and the coeval field, we define the boundaries of the region for analysis as a rectangular area around all six masks used on DEIMOS to target Taralay as a part of the C3VO campaign. This approach helps us mitigate any unforeseen selection biases that could affect the comparison. These boundaries are the aforementioned sky region of interest: $150.1^\circ < \text{RA} < 150.48^\circ$ and $2.21^\circ < \text{Dec} < 2.5^\circ$.

We define the Taralay protocluster as a region with $\sigma_\delta \geq 2$ that also contains a $\sigma_\delta \geq 5$ peak within the redshift range of $4.48 < z < 4.64$. At the time of discovery, Lemaux et al. (2018) established the redshift range of Taralay to be $4.53 < z < 4.6$. We extend this original redshift range to include all galaxies in the full extent of the $\sigma_\delta \geq 2$ voxels associated with Taralay. The left-hand panel of Fig. 3 shows two other structures along the LoS that also satisfy the conditions for being a protocluster, i.e. $\sigma_\delta \geq 2$ region with a $\sigma_\delta \geq 5$ peak. These are denoted S1 and S2. Both of these structures are outside the redshift range $4.48 < z < 4.64$ and so do not comprise part of Taralay. The properties of these two protocluster candidates are briefly discussed in Section 4.

For the coeval field, we aim to select field galaxies such that the sample has a similar average redshift to the systemic redshift of the protocluster. To achieve this, we establish a redshift range that encompasses a temporal window of ± 100 Myr around the systemic redshift of the protocluster. The reason for this choice is the expectation that, due to the relatively short time span, the properties of galaxies in the field sample will be comparable to those of member galaxies within the protocluster, as their evolution may not have diverged significantly. All galaxies in the redshift range $4.2 < z < 4.48$ and $4.64 < z < 4.93$ are considered field galaxies except for those in S1 and S2. We exclude these two structures from our definition of the coeval field because they resemble protoclusters and may confuse our analysis. The conditions satisfied by galaxies to qualify as protocluster members, protocluster peak members or coeval field members are summarized in Table 4. The properties of S1 and S2 are given in Table 5.

The galaxies that make up the coeval field sample are shown in Fig. 4. The galaxies in the purple regions, bounded by $4.2 < z < 4.48$, $4.64 < z < 4.93$ and $\sigma_\delta = 2$ make up most of the field galaxy sample. The field galaxies that are part of the $\sigma_\delta \geq 2$ overdensities

Table 4. Location and overdensity conditions satisfied by galaxies in order to qualify as protocluster members, protocluster peak members, or coeval field members.

Region	Condition
Protocluster peak	$150.1^\circ < \text{RA} < 150.48^\circ$, $2.21^\circ < \text{Dec} < 2.5^\circ$ $4.48 < z < 4.64$ $\sigma_\delta \geq 5$
Protocluster	$150.1^\circ < \text{RA} < 150.48^\circ$, $2.21^\circ < \text{Dec} < 2.5^\circ$ $4.48 < z < 4.64$ $\sigma_\delta \geq 2$ region with a $\sigma_\delta \geq 5$ peak
Field	$150.1^\circ < \text{RA} < 150.48^\circ$, $2.21^\circ < \text{Dec} < 2.5^\circ$ $4.2 < z < 4.48$ or $4.64 < z < 4.93$ All galaxies except the ones in S1 and S2

as well as the galaxies that are neither field galaxies or protocluster members i.e. the galaxies in S1 and S2, are shown by green points on either side of the green region. We see a clear structure emerging in the green region within $4.48 < z < 4.64$ and $\sigma_\delta \geq 2$. The galaxies in this region are the protocluster galaxies. The galaxies in $4.48 < z < 4.64$ with $\sigma_\delta \leq 2$ are excluded from our field given their close proximity to the protocluster.

4 PROPERTIES OF TARALAY

In the initial discovery paper (Lemaux et al. 2018), Taralay had nine members with secure spectroscopic redshifts. With the new data obtained through the C3VO campaign, combined with the VUDS data that led to the discovery and the data from DEIMOS10k and zCOSMOS, we are able to reestablish the morphology, extent, and the internal structure of this protocluster. We found that the Taralay has two substructures, shown in the right-hand panel of Fig. 3, that do not connect by a density isopleth of $\sigma_\delta \geq 2$. We refer to the bigger structure as PC1, which is very roughly similar in location and extent to Taralay at the time of discovery. However, the smaller structure that we refer to as PC2, was not detected in the original work. PC1 hosts two $\sigma_\delta \geq 5$ overdense peaks PC1.P1 and PC1.P2, while PC2 has one $\sigma_\delta \geq 5$ overdense peak PC2.P.

To investigate the properties of Taralay, the method that we use to characterize the protocluster and its $\sigma_\delta \geq 5$ peaks is identical to that in Forrest et al. (2023), which is, in turn, identical to the method used in Cucciati et al. (2018) and Shen et al. (2021). The mass and volume of the Taralay protocluster along with some other properties are summarized in Table 6. The properties of $\sigma_\delta \geq 5$ peaks of the protocluster are given in Table 7. We also list some of the properties of two potential foreground protoclusters S1 and S2 (see Section 3.4) in Table 5.

The apparent comoving volume of the Taralay protocluster obtained by summing the volume of all voxels (see Section 3.3) within the $\sigma_\delta \geq 2$ envelope is $\sim 33\,695 \text{ cMpc}^3$. This apparent volume is artificially increased due to the redshift elongation that originates from the uncertainties in the photometric redshifts and induced motion. To correct the apparent volume, we consider the anisotropic interpretation of the different dimensions. The transverse dimensions are distinct from the LoS dimension, requiring us to factor this discrepancy when calculating the characteristic radii of the structure in each dimension. To correct for this effect, we use the same approach as Cucciati et al. (2018) defining an effective radius that depends on the density and position of each galaxy as well as the barycenter of the overdensity in question. This effective radius is defined for all three dimensions, i.e. x, y, and z. To calculate the elongation, we take a ratio of the effective radius in the z (LoS)

Table 5. Properties of the two potential foreground protoclusters in the area surrounding Taralay.

ID	RA _{peak} (deg)	Dec _{peak} (deg)	z _{peak}	<δ _{gal} >	E _{z/xy}	V (cMpc ³)	M _{tot} (10 ¹⁴ M _⊙)	V _{corr} (cMpc ³)	<δ _{gal} > _{corr}
S1	150.351892	2.478142	4.327	0.99	6.02	2587	1.20	429	24.08
S2	150.195099	2.306873	4.397	0.77	4.29	11 003	4.89	2564	15.16

Table 6. Properties of the Taralay protocluster.

ID	RA _{peak} (deg)	Dec _{peak} (deg)	z _{peak}	<δ _{gal} >	R _x (cMpc)	R _y (cMpc)	R _z (cMpc)	E _{z/xy}	V (cMpc ³)	M _{tot} (10 ¹⁴ M _⊙)	V _{corr} (cMpc ³)	<δ _{gal} > _{corr}
PC1	150.352213	2.354023	4.567	1.53	5.85	7.19	16.40	2.51	29 622	15.49	11 784	9.28
PC2	150.177104	2.301292	4.592	1.07	2.05	2.77	11.93	4.94	4133	1.95	836	19.51

Table 7. Properties of the 5σ overdense peaks of Taralay protocluster at $z \sim 4.57$.

ID	RA _{peak} (deg)	Dec _{peak} (deg)	z _{peak}	<δ _{gal} >	R _x (cMpc)	R _y (cMpc)	R _z (cMpc)	E _{z/xy}	V (cMpc ³)	M _{tot} (10 ¹⁴ M _⊙)	V _{corr} (cMpc ³)	<δ _{gal} > _{corr}
PC1_P1	150.385321	2.380454	4.567	3.37	2.27	2.86	13.24	5.16	4902	3.76	950	35.28
PC1_P2	150.315664	2.300566	4.578	2.95	2.51	3.00	13.41	4.86	4631	3.32	952	30.64
PC2_P	150.178475	2.298961	4.593	1.98	1.04	2.05	10.34	6.69	1063	0.64	159	35.93

Table 8. Formulae used to calculate the total mass, barycentric position, effective radius, elongation correction factor, corrected volume, and corrected average galaxy overdensity of each peak.

Quantity	Formula
M _{tot}	$\rho_m V(1 + \delta_{gal}/b)$
X _{peak}	$\sum_i (\delta_{gal, X_i} X_i) / \sum_i (\delta_{gal, X_i})$
R _X	$\sqrt{\sum_i (\delta_{gal, X_i} (X_i - X_{peak})^2) / \sum_i (\delta_{gal, X_i})}$
E _{z/xy}	$2R_z / (R_x + R_y)$
V _{corr}	$V/E_{z/xy}$
<δ _{gal} > _{corr}	$b[M_{tot}/(V_{corr}\rho_m) - 1]$

direction with the mean of effective radii in x and y directions. The corrected volume, then, is the apparent volume divided by elongation. Table 8 lists the formulae used to calculate these quantities, where ρ_m is the comoving matter density, V is the apparent volume and δ_m is the mass overdensity in the region under consideration. RA_{peak}, Dec_{peak}, z_{peak} are the barycentric position in RA, Dec and z, respectively, with R_x, R_y, R_z being the effective radii in the three dimensions. E_{z/xy} is the elongation correction factor, V_{corr} is corrected volume for elongation effect and <δ_{gal}>_{corr} is the corrected average galaxy overdensity.

The comoving volume of the Taralay protocluster, corrected for elongation, is $\sim 12\,620$ cMpc³, the value we use to calculate the SFRD of the protocluster. We obtain the upper and lower uncertainty in the volume by using density thresholds of $\sigma_\delta \geq 2.2$ and $\sigma_\delta \geq 1.8$, respectively, and calculate the resultant elongation-corrected volume, which results in a final value of $\sim 12620_{-956}^{+1042}$ cMpc³.

The mass of the Taralay is calculated using a formula given in Table 8 for M_{tot} with a bias factor of $b = 3.6$ (see Section 3.3). Due to our ignorance on the precise value of the bias factor appropriate for our tracer population, we additionally vary the bias factor between 4.5 and 3.12, values that are obtained from Einasto et al. (2023) and Ata et al. (2021), respectively, and propagate this uncertainty into the mass uncertainty. The uncertainty in mass due to the variation in the bias factor is added in quadrature with the uncertainty in mass coming from varying the density threshold (as described above). We

estimate the mass of Taralay to be $1.74_{-0.77}^{+1.36} \times 10^{15}$ M_⊙. This value is ~ 6 times higher ($\sim 2\sigma$) than the value reported in Lemaux et al. (2018) at the time of discovery. This difference is likely due to the $\sim 5\times$ increase in the number of spectroscopic members, the larger adopted redshift extent, the higher spectral redshift fraction overall which decreases the dilution from photometric redshifts (see Hung et al. in prep.), and the discovery of the substructure PC2.

4.1 Dynamical versus overdensity mass

We compare the mass obtained from the overdensity with the mass obtained from the estimated LoS velocity dispersion σ_v . We do this comparison for the protocluster, two $\sigma_\delta \geq 4.5$ regions in PC1 (we will call them PC1_R1 and PC1_R2) and one $\sigma_\delta \geq 2.8$ region in PC2 (we will call it PC2_R). These values are a compromise between a sufficiently large sample size to measure σ_v and a region that is sufficiently small such that gravitational interactions between galaxies are reasonably likely. The LoS velocity dispersion σ_v is estimated using the gapper method (Beers, Flynn & Gebhardt 1990 and references therein) with jackknifed confidence intervals for PC1_R1, PC1_R2 and PC2_R regions. For the full protocluster, we use biweight method also with the jackknifed confidence intervals to estimate σ_v . PC1_R1, PC1_R2 and PC2_R regions have sample size of $n \sim 10$ and the entire protocluster has a sample size of $n \sim 40$ with secure spectroscopic redshifts. The gapper method is used for the individual peaks as the smaller sample size makes this the preferred method (Beers et al. 1990), while the sample size is sufficiently large to adopt the biweight estimator in the case of the full protocluster. Fig. 5 shows the velocity histograms and the fit to these histograms to estimate σ_v .

The dynamical mass which refers to the mass enclosed in R₂₀₀, the radius within which the density is 200 times the critical density, is calculated from the σ_v using the following formula

$$M_{200} = \frac{3}{G} \sigma_v^2 R_{200}, \quad (3)$$

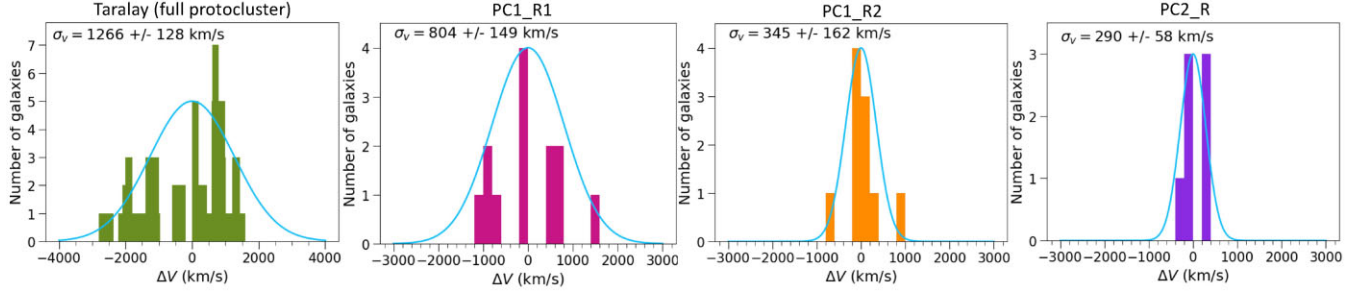


Figure 5. Velocity histograms for Taralay, two $\sigma_\delta \geq 4.5$ regions in PC1 named PC1_R1 and PC1_R2, and one $\sigma_\delta \geq 2.8$ region in PC2 named PC2_R along with the fits to these histograms to estimate velocity dispersion σ_v .

Table 9. The LoS velocity dispersion (σ_v), the resulting dynamical masses and masses from the overdensity for the Taralay protocluster, two $\sigma_\delta \geq 4.5$ regions in PC1 and one $\sigma_\delta \geq 2.8$ region in PC2 along with the number of redshifts (n) that were fitted in each region to obtain σ_v .

Region	n	σ_v (km s^{-1})	M_{200} [$\log(M_{200}/M_\odot)$]	M_δ [$\log(M_\delta/M_\odot)$]
Taralay	43	1266 ± 128	$14.71^{+0.12}_{-0.14}$	$15.24^{+0.25}_{-0.25}$
PC1_R1	13	804 ± 149	$14.11^{+0.22}_{-0.27}$	$14.61^{+0.25}_{-0.26}$
PC1_R2	11	345 ± 162	$13.01^{+0.5}_{-0.83}$	$14.54^{+0.23}_{-0.26}$
PC2_R	7	290 ± 58	$12.78^{+0.24}_{-0.29}$	$14.15^{+0.25}_{-0.25}$

where

$$R_{200} = \frac{\sqrt{3}\sigma_v}{10H(z)}. \quad (4)$$

(Carlberg et al. 1997). The LoS velocity dispersion of $1266 \pm 128 \text{ km s}^{-1}$ for the Taralay protocluster results in the M_{200} of

$$\log(M_{200}/M_\odot)_{z \sim 4.57} = 14.71^{+0.12}_{-0.14}$$

which agrees within the errors with the dynamical mass estimated by Lemaux et al. (2018). The σ_v and M_{200} for PC1_R1, PC1_R2 and PC2_R are given in Table 9. We obtain the error bars for M_{200} by taking into account the error on σ_v . The mass from the overdensity method is obtained through density mapping method (Section 3.3) and the error bars for the mass from overdensity are obtained by varying σ_δ and the bias factor. To this, we added a systematic uncertainty of 0.25 dex based on masses estimated through overdensity reconstruction based on comparison to simulation (Hung et al. in prep). For the protocluster, σ_δ is varied to 1.8 and 2.2, for PC1_R1 and PC1_R2 the σ_δ is varied to 4.3 and 4.7, and for PC2_R the σ_δ is varied to 2.6 and 3. For all regions, the bias factor is varied from 3.12 to 4.5. The comparison between the overdensity masses and the dynamical masses from Fig. 6 shows that the dynamical masses have an average deficit of 2.5σ (range of 1.5σ to 4σ) compared to the overdensity masses. The source of this consistent deficit of the dynamical masses relative to the overdensity will be explored in simulations in future work.

5 GALAXY PROPERTIES OF TARALAY

Now that we have established the morphology and internal structure of the protocluster as well as some of its characteristics, we can investigate the galaxy properties such as the SFR. Understanding the rate at which stars form within galaxies is essential to understand their evolution and behaviour. Various indicators can be employed to estimate the SFR, all of which involve analysing the emitted

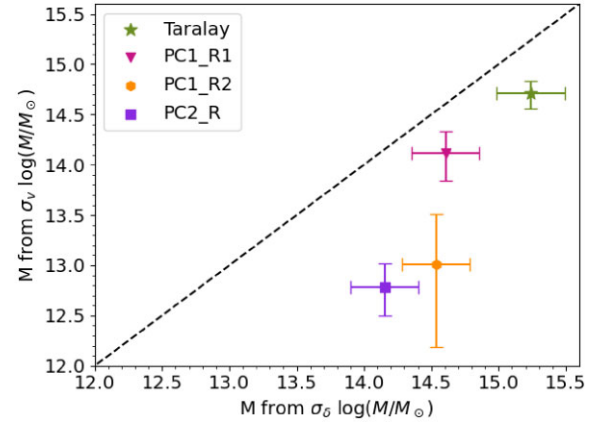


Figure 6. The comparison of mass estimated from the LoS velocity dispersion to the mass estimated from overdensity for the Taralay protocluster, PC1_R1, PC1_R2 and PC2_R regions (see Section 4).

light at different wavelengths. These indicators include UV luminosity, infrared (IR) luminosity, UV and IR luminosity, as well as the strength of recombination lines or their proxies. While these indicators are generally accessible for samples at lower redshifts, estimating the SFR becomes progressively more challenging in the high redshift universe. Acquiring the required recombination line data for a large set of galaxies can be an overwhelming task due to a variety of different factors and atmospheric transmission issues like absorption and scattering. Moreover, specialized equipment such as the Atacama Large Millimeter/sub-millimeter Array (ALMA) and *JWST* are necessary not only to mitigate transmission issues but also to obtain data far enough in the infrared to allow for a more reliable picture of star forming activity in the early universe.

To navigate these challenges, we use a more accessible method for estimating the SFR of high redshift galaxies. Model-based SED fitting provides an alternative approach to estimate the SFR using not just the spectroscopic data but also the available multiwavelength photometric data. The SED fitting process is performed to determine the SFR of each galaxy in order to estimate the SFR per volume per environment i.e. the SFRD in the protocluster versus the coeval field. We explain this process below.

5.1 SED fitting

SED fitting is a powerful tool in astrophysics that involves modelling the SED of celestial objects ideally across a broad range of wavelengths. A model SED is constructed by combining various components or sources of emission that are expected to contribute to

the observed spectrum. Some of the important contributors at high redshift are stellar emission, nebular emission, thermal emission from dust, self absorption, and extinction. These components are varied to find the best fit to the observed data points.

To perform our SED fitting, we chose the Code Investigating GALaxy Emission (CIGALE; Boquien et al. 2019) software, which uses models that describe the various components of a galaxy, such as the stellar population, dust, and gas. These models are constructed based on physical principles and observations of objects in our local universe. The user can select the models and parameters to be included in the fitting, such as the star formation history (SFH), metallicity, dust attenuation, and emission lines. CIGALE compares the observed photometric data (i.e. flux densities measured at various wavelengths) with the model predictions and finds a set of values for a range of parameters that best match the data.

For our SED fitting using CIGALE, we select a set of models to describe the various components of galaxies. The SFH is modelled with sfhdelayed, which stands for a delayed SFH: $\text{SFR} \propto t \times \exp(-t/\tau)$, where t represents time passed since the onset of star formation and τ is the time when the SFR peaks. Our choice of this SFH is based on the study of Thomas et al. (2017) which found this SFH to be an appropriate choice for large samples of galaxies at high redshifts.

We model the spectra for composite stellar population with the libraries of single stellar populations (SSPs) from Bruzual & Charlot (2003) (module bc03). We choose the bc03 module over, e.g. that of Maraston (2005; module m2005) because observations (van Dokkum 2008) suggest that the Bruzual & Charlot (2003) initial mass function (IMF) is more appropriate for higher redshift galaxies than an, e.g. Salpeter (Salpeter 1955) IMF, and the former is supported in the bc03 module but not supported in the m2005 module.

The dust in a galaxy absorbs UV and near-infrared (NIR) radiation and re-emits it in the mid- and far-IR. Two curves associated with this process are extinction curve that only depends on the dust grain size and attenuation curve, which depends on the dust grain size as well as the geometry, i.e. where the dust grains are relative to the source of radiation and the observer. These curves are taken into account through two different modules, the Charlot & Fall (2000) based ‘dustatt_modified_CF00’ module and the Calzetti et al. (2000) based ‘dustatt_modified_starburst’ module. We choose the dustatt_modified_starburst module as it offers more flexibility in terms of slope of the curve and the presence of 217.5 nm bump. This module also includes Small and Large Magellanic Cloud extinction curves of Pei (1992) along with the Milky Way curve of Cardelli, Clayton & Mathis (1989) with O’Donnell (1994) update. More discussion about how various models and the choices of free parameters affect the SED fitting is in Appendix B. We compare SFR and SM fitted with CIGALE and LePhare in Appendix C (see Fig. C1).

By selecting these models and adjusting their parameters, we fit the galaxies and model their SED to recover parameters like SM and SFRs. The parameter values for each of the modules in this fitting are given in Table 10 although see more discussion on the choice of these parameters in Appendix A2. Their detailed description can be found in Boquien et al. (2019). We discuss the effect of lack of far-infrared (FIR) data on the estimated SFR in Appendix A3.

5.2 From SFR to SFRD

After using Monte Carlo to generate 100 realizations of our master spectroscopic plus photometric catalog (see Section 3.2), in each realization, any galaxy may fall into one of three redshift bins: the

Table 10. SED fitting parameters of CIGALE modules sfhdelayed, bc03, and dustatt_modified_starburst.

Modules	Parameter values
tau_main (Myr) ^a	100–30 000
age_main (Myr) ^b	50–1400
tau_burst (Myr) ^c	100, 300
f_burst ^d	0, 0.0001, 0.0005, 0.001, 0.005, 0.01
imf	1 ^e
metallicity (Z_{\odot})	0.008, 0.02
E_BV_lines ^f	0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.5
E_BV_factor ^g	1
uv_bump_amplitude	0, 1, 2, 3 ^h
powerlaw_slope ⁱ	−0.5, −0.25, 0
Ex_law_emission_lines	1 ^j , 3 ^k

^a tau_main in CIGALE refers to the e-folding time of the main stellar population model. ^b age_main in CIGALE refers to the age of the main stellar population in the galaxy. ^c tau_burst in CIGALE refers to the e-folding time of the late starburst population model. ^d f_burst in CIGALE refers to the mass fraction of the late burst population. ^e IMF of 1 in CIGALE refers to the Chabrier IMF. ^f E_BV_lines in CIGALE refers to the colour excess of the nebular lines light for both the young and old population. ^g E_BV_factor in CIGALE refers to the reduction factor to apply to E_BV_lines to compute the stellar continuum attenuation. ^h uv_bump_amplitude of 3 in CIGALE refers to the Milky Way. ⁱ powerlaw_slope in CIGALE refers to the slope delta of the power law modifying the attenuation curve. ^j Ex_law_emission_lines of 1 in CIGALE refers to the Milky Way. ^k Ex_law_emission_lines of 3 in CIGALE refers to the Small Magellanic Cloud.

protocluster, the field, or outside the redshift range of interest. The probability of a galaxy falling into one of these categories depends on several factors, including whether the galaxy has a spectroscopic redshift or not, the quality of the spectroscopic redshift and the width of the zPDF.

For each realization, we determine which environmental category each galaxy falls in and calculate the total SFR for all galaxies identified as protocluster galaxies (SFR_{pc}) and the total SFR for all galaxies identified as coeval field galaxies ($\text{SFR}_{\text{field}}$) according to the definitions given in Section 3.4. We then calculate the SFRD of the protocluster (SFRD_{pc}) by dividing SFR_{pc} by the volume of the protocluster obtained from density mapping but corrected for elongation (given in Table 6). Similarly, we calculate the SFRD of the field ($\text{SFRD}_{\text{field}}$) by dividing $\text{SFR}_{\text{field}}$ by the volume of the field. The volume of the coeval field is obtained by subtracting two quantities from the total volume of our region of interest: (1) the volume of the region enclosed in $4.48 < z < 4.64$ (which includes the protocluster); (2) the uncorrected volume of S1 and S2. The resulting coeval field volume is 686583 cMpc^3 as compared to $\sim 12621 \text{ cMpc}^3$ for Taralay. However, if we instead subtract the elongation corrected volume for S1 and S2 from the field, the volume of field increases only by 1 per cent, which has negligible effect on our results.

5.3 Contribution of lower luminosity galaxies

The spectroscopic and photometric data that this study use has limitations in terms of depth. To include the fainter galaxies that our data cannot probe and take into account the contribution of these fainter galaxies to the SFRD, we extrapolate our results for the protocluster as well as field to lower luminosity (see Appendix A4 for discussion about the effect of dust properties of faint and bright galaxies on the SED fitting process in order to estimate accurate SFR and the corrections we need to perform in order to extrapolate our

result safely). In order to do this, we begin by selecting a sample of objects with $4.2 < z_{\text{phot}} < 4.93$ whose IRAC channel 1 mag fall within 25.3 and 25.5. We look at the distribution of Far-UV magnitude of this sample and take the 80 per cent completeness limit to probe the depth of our UV and optical data in a given window of IRAC channel 1 (see appendix B of Lemaux et al. 2018 for the basic idea). The range of $25.3 < M_{\text{IRAC1}} < 25.5$ is chosen because it is brighter than the IRAC channel 1 completeness limit [25.7 in (3.6) band] stated in Weaver et al. (2022) making it likely that all objects at this brightness are detected. For the sample in this window, we sort the M_{FUV} magnitudes and remove the faintest 20 per cent objects of the sample. This is done to obtain 80 per cent completeness limit. The resulting M_{FUV} distribution is then corrected by the average difference between the range of IRAC channel 1 window we choose and the completeness limit (25.7) in order to get the M_{FUV} completeness of our sample. For example, the average difference between the window $25.3 < M_{\text{IRAC1}} < 25.5$ and the completeness limit of IRAC channel 1 (25.7) is 0.3.

The above calculation is based on an assumption that the change in M_{IRAC1} corresponds exactly to the change in M_{FUV} for the galaxy population considered here. This is not necessarily the case. In order to check this, we repeated this exercise with a different IRAC channel 1 window, $24.8 < M_{\text{IRAC1}} < 25.0$, and found that the median M_{FUV} is offset by 0.36 mag between the windows $25.3 < M_{\text{IRAC1}} < 25.5$ and $24.8 < M_{\text{IRAC1}} < 25.0$ as compared to the change in M_{IRAC1} of 0.5 mag between these two windows. To account for this difference, we correct the measured M_{FUV} distribution in our original window not by the average difference between the median IRAC1 magnitude but by the expected corresponding change in M_{FUV} coming from the above exercise. Ultimately, this results in a very small correction to the M_{FUV} distribution (~ 0.2 mag). This exercise results in the M_{FUV} completeness limit of our sample to be approximately -19.3 , a value that is not strongly dependent on the various windows chosen in this exercise.

With the depth of our data established at $M_{\text{FUV}} = -19.3$, we extrapolate our results for SFRD to $M_{\text{FUV}} = -17$ in order to include the contribution of the fainter UV galaxies not detected in the data used in this study. This also lets us compare our results with studies that report the SFRD values corrected to include the contribution of the fainter galaxies. We use the Schechter function (Schechter 1976) to extrapolate SFRD_{pc} and $\text{SFRD}_{\text{field}}$ down to $M_{\text{FUV}} = -17$. The faint limit of $M_{\text{FUV}} = -17$ is chosen because the behaviour of the galaxy luminosity function is not well known beyond $M_{\text{FUV}} = -17$ and may deviate from a simple Schechter function (e.g. O’Shea et al. 2015, de La Vieuville et al. 2019, Yung et al. 2019). We use the following equation to determine the correction factor:

$$CF = \frac{\int_{-\infty}^{-17} \left[dM (0.4 \times \log 10) \phi^* [10^{0.4(M^*-M)}] \right]^{\alpha+1} \times \exp[-10^{0.4(M^*-M)}] \right]}{\int_{-\infty}^{-19.3} \left[dM (0.4 \times \log 10) \phi^* [10^{0.4(M^*-M)}] \right]^{\alpha+1} \times \exp[-10^{0.4(M^*-M)}] \right]} \quad (5)$$

The faint-end slope α utilized in our study is determined by combining values obtained from multiple studies (e.g. Ouchi et al. 2004; Giavalisco 2005; Sawicki & Thompson 2006; Yoshida et al. 2006; Bouwens et al. 2007; 2015). To obtain a representative value and its associated uncertainty, we construct a joint PDF from the reported values and associated uncertainties in these studies. The

mean of this joint PDF serves as the final estimate of α with 16th and 84th percentiles serving as corresponding errors in our study. The value for α is $\alpha = -1.77^{+0.22}_{-0.20}$. The values for M_{UV}^* , ϕ^*

$$M_{\text{UV}}^* = (-20.95 \pm 0.10) + (0.01 \pm 0.06)(z - 6),$$

$$\phi^* = (0.47^{+0.11}_{-0.10}) 10^{(-0.27 \pm 0.05)(z-6)} 10^{-3} \text{Mpc}^{-3}$$

are taken from Bouwens et al. (2015) where we substituted $z = 4.57$ for this study. The correction factor is $\log(\text{CF}) = 0.96 \pm 0.17$ dex, which implies that faint galaxies are contributing significantly to the overall SFRD. Changing the completeness limit by 10 per cent changes the log correction factor by ~ 0.1 dex.

6 RESULTS

In this section, we report the SFRD of the Taralay protocluster at $z \sim 4.57$, SFRD of its three $\sigma_\delta \geq 5$ peaks, SFRD of the coeval field as well as the contribution of the protoclusters at $z \sim 4.57$ to the cosmic SFRD using Taralay as a proxy of all the protoclusters at this redshift. We also report on the $\text{SFR}-\sigma_\delta$ relation for all galaxies in the protocluster and coeval field.

6.1 SFRD of the field surrounding Taralay

We find that the SFRD of the galaxies in the coeval field surrounding the Taralay protocluster is $\log(\text{SFRD}/M_\odot \text{yr}^{-1} \text{Mpc}^{-3}) = -0.82^{+0.19}_{-0.29}$. The uncertainty on the $\text{SFRD}_{\text{field}}$ is a result of the combined uncertainty on the $\text{SFR}_{\text{field}}$ from performing Monte Carlo on redshifts (Section 3.2), the SED fitting (Section 5.1), the change in the SFR of the galaxies in the field due to varying the boundary of the protocluster (since σ_δ cut dictates which galaxies qualify as field galaxies or protocluster members, see Section 4), the uncertainty on the Schechter parameters (Section 5.3) and the uncertainty on the volume of the field from changing the boundary of the protocluster (see Section 5.2 for calculating the field volume). We compare our $\text{SFRD}_{\text{field}}$ with various studies in Fig. 7.

A particularly comparable study to our own is that of Khusanova et al. (2021), where the SFRD value is measured using a spectroscopic sample (from VUDS and DEIMOS) with corrections based on an adopted Far-UV luminosity function and galaxy SM function. This study uses rest-frame far-infrared continuum observations with ALMA in order to derive dust-obscured SFR. Using a somewhat similar framework to ours, the authors of this work also performed a similar faint-galaxy correction to their SFRD results. We find that our result for $\text{SFRD}_{\text{field}}$ is statistically indistinguishable from the SFRD value estimated by Khusanova et al. (2021) giving us confidence in our $\text{SFRD}_{\text{field}}$ value. This agreement indicates that the assumptions on the dust attenuation curves that went into our SED fitting in order to derive SFRD values are well accounted for.

Although the $\text{SFRD}_{\text{field}}$ value we report here is higher than what is predicted by Madau & Dickinson (2014) at $z \sim 4.57$, the values at these redshifts from Madau & Dickinson (2014) may be underestimated due the paucity of data at those redshifts a decade ago. Indeed, many of the more contemporary studies reported in Fig. 7 recover values in excess of the Madau & Dickinson (2014) best fit at these redshifts.

More specifically, values in excess of the Madau & Dickinson (2014) fit at these redshifts is supported by the findings of Kistler et al. (2009), in which the SFRD values are measured based on gamma-ray bursts, the SFRD values from measurements based on *Herschel* data from Rowan-Robinson et al. (2016), the SFRD value measured using Far-UV luminosity function and the galaxy SM function from

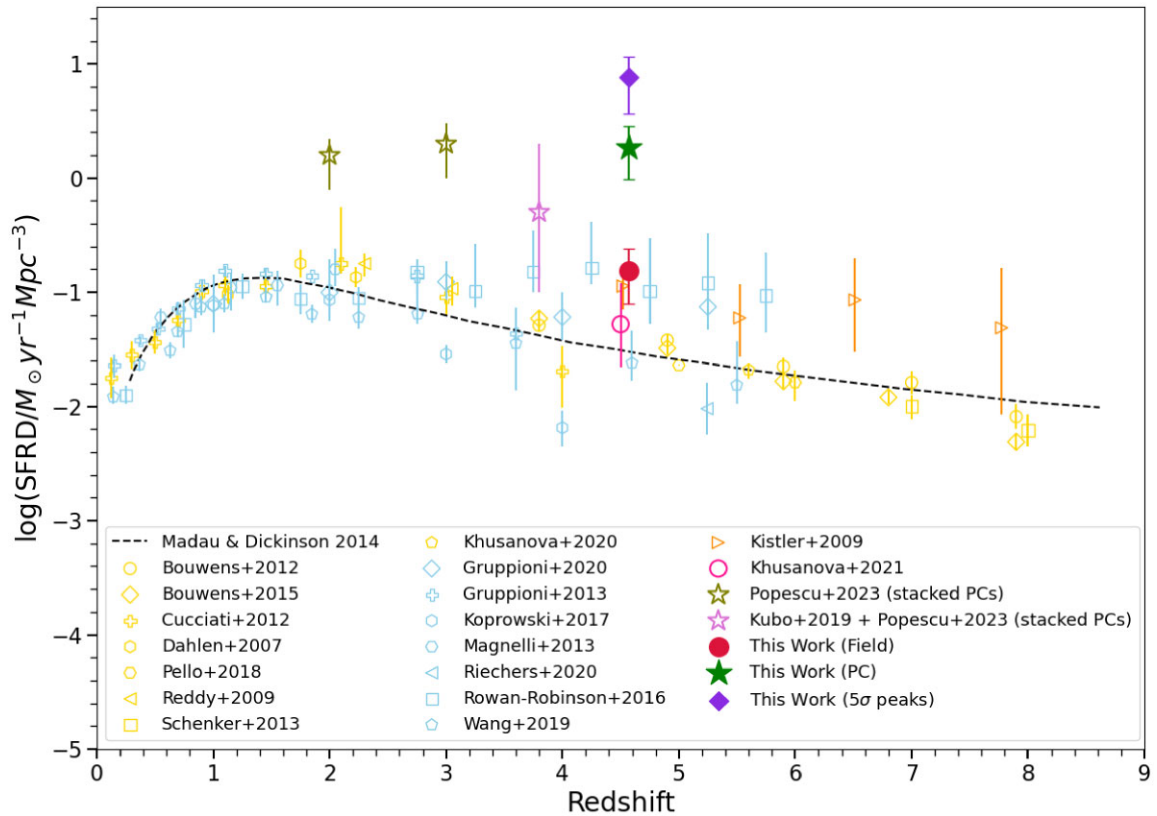


Figure 7. Comparison of protocluster SFRD with the coeval field SFRD. The green star, purple diamond, and circle cross show the SFRD values calculated in this work for the Taralay protocluster, its $\sigma_8 \geq 5$ peaks and the surrounding coeval field respectively. The rest of the points show evolution of SFRD with redshift from various studies. The yellow points are dust-corrected SFRD obtained with FUV data from Dahlen et al. (2007), Reddy & Steidel (2009), Cucciati et al. (2012), Bouwens et al. (2012, 2015), Schenker et al. (2013), Pelló et al. (2018), and Khusanova et al. (2020). The skyblue symbols are the SFRD measured from IR due to the re-radiation of dust emission from forming stars. These data points are taken from Magnelli et al. (2013), Gruppioni et al. (2013, 2020), Rowan-Robinson et al. (2016), Koprowski et al. (2017), Wang et al. (2019b), and Riechers et al. (2020). The orange data points are SFRD from Kistler et al. (2009) calculated using the number of γ -ray bursts. The pink circle is the extrapolated field SFRD value that is an average of the SFRD values leveraged on an adopted Far-UV luminosity function and those leveraged on an adopted galaxy SM function in Khusanova et al. (2021). The olive open stars are the SFRD values calculated by Popescu et al. (2023) for protoclusters at $2 < z < 4$ using stacked *WISE* and *Herschel/SPIRE* images. The orchid open star is the SFRD value of protoclusters at $z \sim 3.8$, where the SFR value is estimated by Kubo et al. (2019) and the SFRD value is calculated by Popescu et al. (2023).

Khusanova et al. (2021), the SFRD measurement based on far-IR from Gruppioni et al. (2020), and the SFRD measurement at $z > 3$ derived from radio luminosities and translated to far-IR luminosities using q_{TIR} (Novak et al. 2017). Beyond Khusanova et al. (2021), other recent works from the ALPINE survey using far-IR measurements also recover higher values for the SFRD at $4 < z < 6$ than the Madau & Dickinson (2014) relation (Gruppioni et al. 2020; Loiacono et al. 2021). Such results are similarly in tension with the SFRD measured from other studies that use dust-corrected FUV data (e.g. Bouwens et al. 2012; Cucciati et al. 2012; Schenker et al. 2013; Bouwens et al. 2015; Pelló et al. 2018; Khusanova et al. 2020). The reason that likely accounts for the difference between SFRD values derived from IR data versus the SFRD values derived from dust-corrected FUV data is the uncertainty that comes from the IRX- β relation that is used for dust-correction (e.g. Salim & Narayanan 2020 and references therein). In the future, we will investigate further the $\sim 2.5\sigma$ tension with the best fit in Madau & Dickinson (2014) by probing the field surrounding other structures at other redshifts in the C3VO survey.

6.2 SFRD of the Taralay protocluster

The SFRD of the Taralay members is $\log(\text{SFRD}/M_\odot \text{ yr}^{-1} \text{ Mpc}^{-3}) = 0.26^{+0.18}_{-0.28}$ (shown in Fig. 7) in excess of $\text{SFRD}_{\text{field}}$ by 1.08 ± 0.32 dex

($\sim 12\times$). This value is $\sim 6\sigma$ in excess than the best fit in Madau & Dickinson (2014) indicating that the protocluster galaxies are well outpacing the field. The excess with respect to the field may mean that members of Taralay are rapidly building up their collective SM through star-forming processes, well in excess of such growth in the field. Later, in this section, we will show that this is indeed the case.

We compare our results for SFRD_{pc} with the SFRD values from Popescu et al. (2023) and Kubo et al. (2019), shown by olive and pink open stars in Fig. 7, respectively. Popescu et al. (2023) stacked *Wide-field Infrared Survey Explorer* (*WISE*; Wright et al. 2010) and *Herschel*/Spectral and Photometric Imaging REceiver (*SPIRE*; Griffin et al. 2010) images for 211 protocluster candidates at $2 < z < 4$ that they selected as Planck cold sources from Planck Collaboration (2015). They define sources with redder color as cold sources that peak between 353 and 857 GHz. The redder colour corresponds to a cold dust temperature or a high redshift. The SFR of the protocluster candidates was derived through SED fitting method using CIGALE. This SFR was converted into SFRD for each candidate protocluster by using a volume approximated by a sphere of radius 10.5 comoving Mpc (cMpc) at $z = 2$. Popescu et al. (2023) also converted the SFR derived in Kubo et al. (2019), a study that stacked *Planck*, *AKARI* (Murakami et al. 2007), *Infrared Astronomical Satellite* (Neugebauer et al. 1984), *WISE*, and *Herschel*

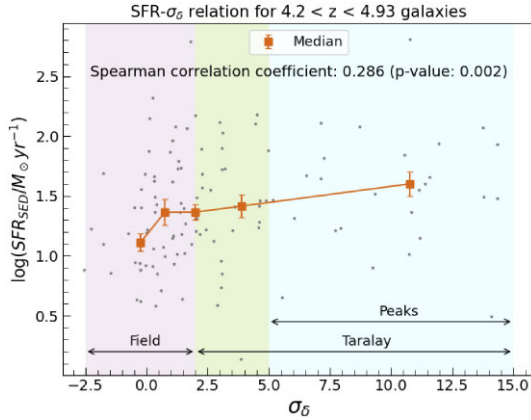


Figure 8. SFR– σ_δ (overdensity) relation for all spectroscopically confirmed galaxies that fall in the redshift range of interest $4.2 < z < 4.93$. The SFR for these galaxies are estimated with CIGALE, an SED fitting code (see 5.1). The orange squares are the median SFR plotted at the median overdensity for each bin such that each bin contains approximately equal number of galaxies. The spread on the median is calculated using σ_{NMAD} . The Spearman correlation coefficient is positive indicating a weak correlation between local galaxy density and the SFR.

images for 179 candidate protoclusters at $z \sim 4$, selected from the HSC Subaru Strategic Program, with the combined IR emission in the observed 12–850 μm wavelength range. The SFR of LBG-selected protocluster candidates from this study was converted into SFRD by using a spherical volume with a radius of 10.2 cMpc. Albeit at different redshifts, we find a good agreement between our results and the results of these studies (see Section 7 for more discussion).

It is tempting to attribute the elevated SFRD of the Taralay protocluster compared to the field solely to the fact that the protocluster hosts a great number of star-forming galaxies in a relatively small volume. Here we focus on the SFR– σ_δ relation in order to investigate whether the high SFRD_{pc} comes simply from having a large number of galaxies in the protocluster or if it is also a product of the protocluster galaxy members genuinely having an elevated SFR relative to their counterparts in the field. The SFR– σ_δ relation shown in Fig. 8 for galaxies in this sample reveals a positive correlation between the SFR and overdensity. A Spearman test results in a correlation coefficient of 0.286 and a p-value of 0.002. An identical exercise is performed with respect to SM later in this section. Lemaux et al. (2022) found a weak but significant trend for SFR-overdensity for the full VUDS+ sample of 6730 star-forming galaxies over the redshift range $2 \leq z \leq 5$. The strength of the correlation seen in our sample is >2 times higher than that measured in Lemaux et al. (2022; $\rho = 0.29$ versus $\rho = 0.13$) indicating that members of the Taralay protocluster are even more likely to have an increase in the SFR as in denser environments than the overall star-forming galaxy population at $2 \leq z \leq 5$.

To disentangle the fraction of SFRD that results from the protocluster having a higher number of galaxies versus the fraction that results from the protocluster galaxies having higher SFR on average, we investigated a scenario where we assumed that the average SFR of the protocluster galaxies is the same as the average SFR of the coeval field galaxies. In Fig. 8, the average $\log(\text{SFR}/M_\odot \text{ yr}^{-1})$ increases 0.22 for galaxies in the protocluster relative to those in the field— $\log(\text{SFRD}/M_\odot \text{ yr}^{-1} \text{ Mpc}^{-3}) = 1.25$ versus 1.47. We then reduce the SFRD_{pc} by the ratio between the average SFR of protocluster members and that of field galaxies, which results in a $\log(\text{SFRD}_{\text{pc}}/M_\odot \text{ yr}^{-1} \text{ Mpc}^{-3}) = 0.04$ as compared to the derived value of 0.26. This

means that even if protocluster members had the same average SFR as field galaxies, the SFRD_{pc} would still be higher than the SFRD_{field} by 7.3 times instead of 12. In other words, 43 per cent of the difference in the SFRD_{pc} and the SFRD_{field} is as a result of the elevated SFR of the protocluster members relative to that of field galaxies, with the remaining 57 per cent resulting from the higher galaxy number density.

In principle, it is possible that this increase in the SFR is due to an increase in the SM (e.g. Davé 2008; Sobral et al. 2014; Salmon et al. 2015; Tomczak et al. 2016) as the environment gets denser. However, performing a comparable calculation with respect to SM results in no significant evidence of correlation between the SM and overdensity. The Spearman correlation coefficient is weaker (0.155) with a p value of 0.09. We recast these results in the next section.

6.3 SFRD of the $\sigma_\delta \geq 5$ peaks of Taralay

We calculate the SFRD of the peaks of Taralay to be $\log(\text{SFRD}_{\text{peak}}/M_\odot \text{ yr}^{-1} \text{ Mpc}^{-3}) = 0.87^{+0.18}_{-0.32}$. Similar to the full Taralay protocluster, the $\sigma_\delta \geq 5$ peaks of this protocluster also show SFRD well in excess of the coeval field value.

Fig. 9 shows the cumulative distribution function (CDF) of the SFR, SM and the number of galaxies as a function of σ_δ . It is obvious on inspection that the CDF of SFR in the protocluster is skewed toward higher values in the central regions than that of the average number of galaxies, while the CDF of SM tracks the average number of galaxies throughout the entire protocluster. For example, about 68 per cent of the total SFR in the protocluster (SFR_{pc}) takes place in the $\sigma_\delta \geq 5$ peaks while the inner most regions of the peak ($\sigma_\delta \geq 10$) contains 50 per cent of the SFR_{pc}. This points to the galaxies in the peaks having accelerated evolution and is highly suggestive of inside-out growth. These galaxies might become the more quiescent galaxies that are seen ~ 850 Myr later at $z \sim 3$ (e.g. Franco et al. 2020; Forrest et al. 2020a,b; Shen et al. 2021; McConachie et al. 2022; Ito et al. 2023).

The high SFR in the peaks cannot be attributed only to the large number of galaxies in these regions as less than 30 per cent protocluster member galaxies are in the inner most regions of the peaks with the entire $\sigma_\delta \geq 5$ peaks hosting less than 50 per cent of the total protocluster member galaxies. The SM in the peaks is < 50 per cent of the total SM encased in the protocluster, which largely rules out higher SFR of the galaxies in the peak being a result of those galaxies having higher SMs than counterparts in more rarefied regions. The segregation observed in SFR between the densest regions of the protocluster, the protocluster outskirts, and the field, and the lack of SM segregation strongly indicate that the protocluster members, especially those in densest regions, are just beginning to ramp up their star formation activity. If such activity was sustained for even a relatively short time early in the formation history of the protocluster, SM segregation would almost certainly also be observed.

6.4 Fractional contribution of the protoclusters at $z \sim 4.57$ to the cosmic SFRD

We also estimate the fractional contribution from protoclusters to the cosmic SFRD at $z \sim 4.57$ using Taralay as a representative of a sample of protoclusters at these redshifts (although see Appendix A6 for discussion about using such a massive protocluster as a proxy). We estimate this fraction using the following formula:

$$\text{SFRD}_{\text{pc}} \text{ fraction} = \frac{\text{SFRD}_{\text{pc}} \times \text{vol}_{\text{pc}}}{\text{SFRD}_{\text{pc}} \times \text{vol}_{\text{pc}} + \text{SFRD}_{\text{field}} \times (1 - \text{vol}_{\text{pc}})} \quad (6)$$

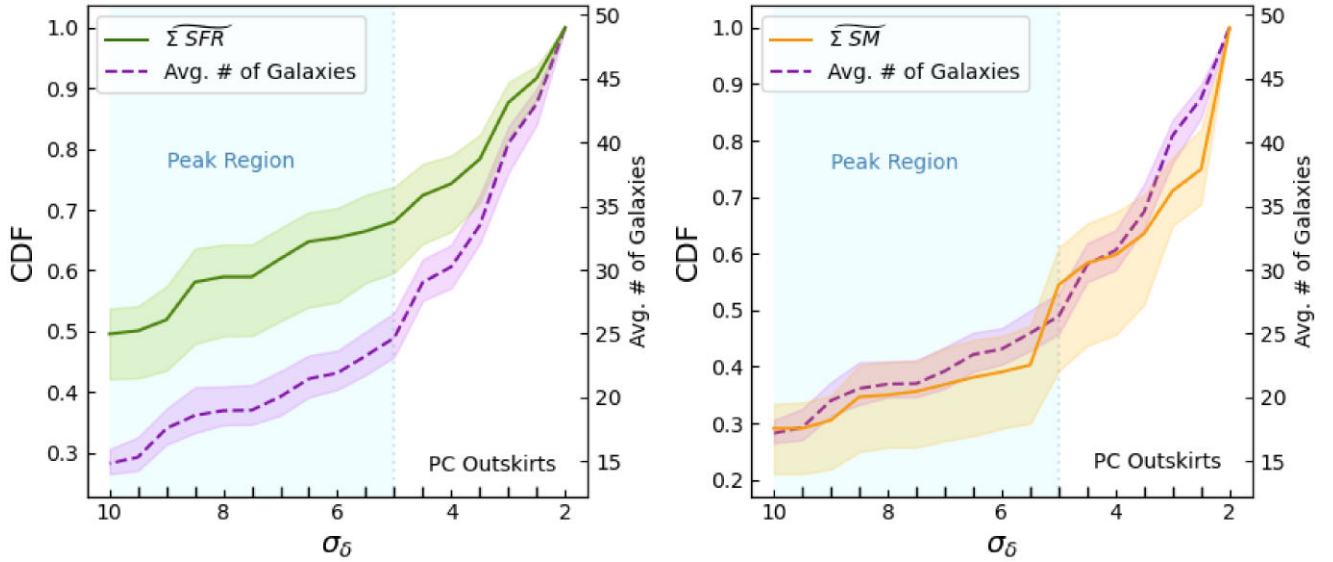


Figure 9. The cumulative distribution function (CDF) of the average total SFR, the average total SM and the average number of galaxies as a function of σ_δ . The average for all quantities is taken from 100 iterations of Monte Carlo. The distribution of the average number of galaxies shown on the right hand y-axis for both panels is slightly imprecise due to the statistical nature of the Monte Carlo process. The shaded blue region indicates the peak region of the Taralay protocluster and the $2 \leq \sigma_\delta \leq 5$ area shows the outskirts of the protocluster for both panels.

where vol_{pc} refers to the volume occupied by protoclusters. At $z \sim 4.57$, the volume filling factor for protoclusters calculated by Chiang et al. (2017) is 0.04.⁶ This makes the fractional contribution to SFRD from protoclusters at $z \sim 4.57$ to be $33.5 \text{ per cent}^{+8.0}_{-4.3 \text{ per cent}}$.⁷ The uncertainty on the SFRD_{pc} fraction comes from the uncertainty on the SFR from performing Monte Carlo on redshifts (Section 3.2), and change in the SFR and volume of the protocluster due to varying the boundary of the protocluster (Section 3.3).

We compare this result with the predictions from Chiang et al. (2017). The estimated SFRD_{pc} fraction in this study is 2.6σ and 1.7σ in excess of the predicted $\sim 22 \text{ per cent}$ and $\sim 26 \text{ per cent}$ from (Guo et al. 2013) and (Henriques et al. 2015) simulations, respectively. Though our SFRD_{pc} fraction shows moderate tension with the predictions from simulations, we would like to note that there are several caveats when comparing observational data and simulations. We discuss how the definition for a protocluster that we adopt in this study varies from the definition adopted in Chiang et al. (2017) in Appendix A5. We also remind the reader that semi-analytic models (SAMs) assign the SFR to galaxies using a prescription that differs from SED fitting results. However, regardless of these considerations, the simulations, other observational studies, and our measurements of Taralay all indicate that protoclusters at $z \geq 2$ contribute significantly to the SM growth of the universe.

We also compare our results with the SFRD_{pc} fractions based on the SFRD values of Popescu et al. (2023) in Fig. 10. To obtain the

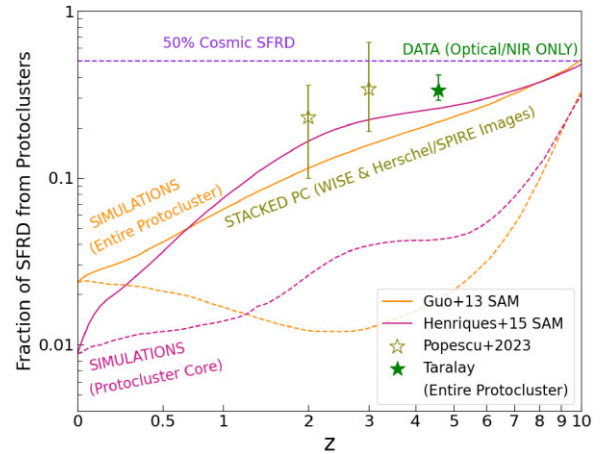


Figure 10. Comparison of the fraction of SFRD from protoclusters at $z \sim 4.57$ estimated using Taralay protocluster as a proxy of all protoclusters at this redshift with the fraction of protocluster SFRD predicted through Chiang et al. (2017) simulation. The olive stars show the fraction of SFRD from protoclusters at $z = 2$ and $z = 3$ obtained from Popescu et al. (2023). The solid lines is the fraction predicted for all protoclusters in the simulations with two different SAMs from Henriques et al. (2015) and Guo et al. (2013). The dotted lines show the fraction predicted for only the protocluster cores with SAMs mentioned above.

SFRD_{pc} fractions corresponding to the SFRD values listed in this study at $z = 2$ and $z = 3$, we first measure $\text{SFRD}_{\text{field}}$ at $z = 2$ and $z = 3$ based on the SFRD values in the $1.8 < z < 2.2$ redshift range from Dahlen et al. (2007), Gruppioni et al. (2013, 2020), Magnelli et al. (2013), Koprowski et al. (2017), and Wang et al. (2019b) and SFRD values in the $2.8 < z < 3.2$ redshift range from Gruppioni et al. (2013, 2020) and Koprowski et al. (2017), respectively. The median, 16th and 84th percentile of these samples give us the $\text{SFRD}_{\text{field}}$ and errors at $z = 2$ and $z = 3$ respectively. We find a good agreement between the SFRD_{pc} fraction of this study and the SFRD_{pc} fractions we calculate based on the SFRD values of Popescu et al. (2023).

⁶The volume filling factor in our data is 0.018 when only Taralay is considered and increases to 0.022 if S1 and S2 are included. The volume filling factor is obtained by dividing the volume of the overdense structure(s) by the volume of the coeval field reported in Section 4.

⁷Adopting the volume filling factor of our data, 0.022, rather than that of simulations, 0.04, recovered a fractional contribution to SFRD from protoclusters at $z \sim 4.57$ of $\sim 21 \text{ per cent}$. This value is also consistent with the equivalent value from simulations. However, the filling factor estimated from our data is subject to cosmic variance to a much higher level than that of the simulation.

We further discuss these results with respect to additional literature in Section 7.

We also calculate the fractional contribution from the $\sigma_8 \geq 5$ peaks of protoclusters to the cosmic SFRD using equation

$$\text{SFRD}_{\text{peak}} \text{ fraction} = \frac{\text{SFRD}_{\text{peak}} \times \text{vol}_{\text{peak}}}{\text{SFRD}_{\text{pc}} \times \text{vol}_{\text{pc}} + \text{SFRD}_{\text{field}} \times (1 - \text{vol}_{\text{pc}})} \quad (7)$$

where vol_{peak} is calculated by as the ratio of volume of peaks estimated with density mapping ($\sim 2060 \text{ cMpc}^3$) divided by the volume of the entire protocluster ($\sim 12\,620 \text{ cMpc}^3$) and then multiplying this ratio by the volume filling factor for protoclusters, 0.04. The estimated $\text{SFRD}_{\text{peak}}$ fraction is $22.2 \text{ per cent}^{+5.4}_{-7.3}$, a significant portion of the SFRD_{pc} fraction, which is indicative of the inside-out growth of protoclusters as discussed in Chiang et al. (2017). Note that we do not use the volume filling factor for the protocluster cores as outlined in Chiang et al. (2017) as their definition of core differs from our definition of peaks, discussed in Section A5.

7 DISCUSSION

The SFRD_{pc} we report in this study is based on optical/Near-IR data. To obtain the total SFR for each galaxy, we perform SED fitting that includes a dust correction. We find that our value is in agreement with the value reported in Popescu et al. (2023), a study that focuses on the stacked Far-IR data for 211 protoclusters at $2 < z < 3$ to estimate the SFR of galaxies using SED fitting (see Section 6.2 for a brief description of analysis employed by Popescu et al. (2023) to obtain the SFRD values). Because this study concentrates on the Far-IR emission of protocluster galaxies, they are only sensitive to obscured star formation activity. At such redshifts the emission in the FIR is, perhaps, a good proxy of the total SFR at $2 < z < 3$, as obscured star formation activity is predicted and measured to be an order of magnitude higher than the unobscured SFRD at these redshifts on average (see, e.g. Algera et al. 2023 and references therein). By contrast, at $z \sim 4.5$, the unobscured SFRD is thought to be in excess or comparable to the obscured SFRD (e.g. Khusanova et al. 2021; Algera et al. 2023), which makes it a reasonable approximation of the total SFRD when corrections for extinction are applied. Note, however, that there are some clear exceptions in protocluster environments, which we discuss in the next paragraph. Although the relative contribution of the unobscured and obscured SFRD in the field and in protoclusters has yet to be studied in great detail, a comparison of our results with those of Popescu et al. (2023) shows a good agreement between the obscured SFRD of protocluster galaxies at redshifts $z \sim 2.5$ with the extinction-corrected unobscured SFRD derived for Taralay at $z \sim 4.5$.

There exist some rare systems that contain overdensities of sub-mm galaxies that exhibit extreme star formation activity, such as the SPT2349-56 protocluster, which was discovered in the South Pole Telescope (SPT)’s extragalactic mm-wave point-source catalogue (Vieira et al. 2010, Mocanu et al. 2013, Everett et al. 2020) and followed up with the ALMA telescope (Miller et al. 2018). This protocluster has over 30 submillimetre-bright galaxies along with LAEs and LBGs and SFRD of over $10^5 \text{ M}_{\odot} \text{ yr}^{-1} \text{ Mpc}^{-3}$ at $z \sim 4$ (Hill et al. 2022). Such rare systems with high star formation activities also exist at lower redshifts, e.g., a system from Wang et al. (2016) at $z \sim 2.5$ that has nine starburst galaxies in the center whose SFR amounts to $3400 \text{ M}_{\odot} \text{ yr}^{-1}$ within an 80 kpc region, and four enormous Ly α nebulae from Nowotka et al. (2022) at $2 < z < 3$ with an SFRD of $1200 \pm 300 \text{ M}_{\odot} \text{ yr}^{-1} \text{ Mpc}^{-3}$; however, the SPT2349-

56 protocluster is one of the only few systems yet discovered at a comparable redshift to Taralay (see Alberts & Noble 2022 and references therein) that shows an extremely high star formation activity. The SFRD of such systems, while high in the region of the universe that they exist, averaged over the entire sky is probably much less than typical optical/Near-IR selected systems as such types of systems are extremely rare (e.g. Negrello et al. 2017, Lammers et al. 2022). In this way, Taralay might be a better representative of the underlying population of massive protoclusters at this epoch. More work is needed in the future to compare SFRD of Taralay with an ensemble of protoclusters at these redshifts.

Although the optical/Near-IR selection may lead to a sample of protoclusters that are more representative of an underlying galaxy populations, one of the disadvantages of this approach is that the optical/Near-IR diminished/dark galaxies get left out of the sample. If such galaxies, sometimes called HST-dark galaxies, i.e. the galaxies that are undetected in the current *HST* surveys due to being effectively invisible in the rest-frame ultrablue to the typical depths of *HST* survey observations (e.g. Franco et al. 2018, Wang et al. 2019a), exist in the region that is targeted in this study, Barrufet et al. (2023) shows that their presence will contribute approximately an order of magnitude less than the rest-frame UV/optically selected galaxies to the total SFRD at these redshifts. Due to the volume of the protocluster being a lot smaller than the field, the contribution of the HST-dark galaxies may affect the SFRD of the protocluster more than the field widening the gap between SFRD_{pc} and $\text{SFRD}_{\text{field}}$. The effect of these galaxies on the gap between SFRD_{pc} and $\text{SFRD}_{\text{field}}$ will depend on the location preferred by the galaxies.

Studies such as Blain et al. (2004), Almeida, Baugh & Lacey (2011), Smolčić et al. (2017), Zhou et al. (2020) claim association of sub-millimeter galaxies with overdensities, though it is not clear if these sources are fractionally over represented in protoclusters relative to the field. There is some indication from the ALPINE-ALMA survey that there might be a higher number of extremely dusty star-forming galaxies at elevated redshifts in denser environments compared to the field, including around $z \sim 4.57$ (e.g. Romano et al. 2020, Loiacono et al. 2021, Fujimoto et al. 2023). If such galaxies are more prevalent in rich environments relative to the field, the SFRD_{pc} we calculate here will be a lower limit as will its fractional contribution to the overall SFRD at this redshift.

The observational SFRD_{pc} fraction we found in this study is none the less quite large, even in excess of the predictions from simulations and indicates that the protoclusters are a significant contributors to the cosmic SFRD at high redshifts. Ito et al. (2020) also found protoclusters as a driver of SM growth in the early universe. They calculated rest-frame ultra-blue luminosity function of g-dropout galaxies in 177 protocluster candidates at $z \sim 4$ selected in the HSC Subaru Strategic Program (Aihara et al. 2018); though their SFRD fraction, 6 per cent–20 per cent, is not as high as the finding of this study.

At $z \sim 4.57$, the number density of quiescent galaxies is small (e.g. Steinhardt et al. 2014, Tasca et al. 2015, Davidzon et al. 2017, Gould et al. 2023). All such galaxies may not be detected in our combined spectroscopic and photometric sample. However, their impact on our results are minimal, as their contribution to the SFR is expected to be very low. For example, a massive quiescent galaxy in an overdense environment at $z = 4.53$ in the COSMOS field that was observed with MOSFIRE on Keck has SFR an order of magnitude less than the SFR of main-sequence galaxies at $z = 4.5$ (Kakimoto et al. 2023). This galaxy is ~ 6 proper Mpc away from the center of the PC1 of Taralay and may or may not be associated with Taralay.

8 SUMMARY AND CONCLUSION

This study was performed in order to conduct one of the first observational tests of whether the protocluster regions in the high redshift universe contribute significantly to the overall mass assembly of the universe as predicted by simulations. We chose the Taralay protocluster at $z \sim 4.57$ in the COSMOS field as the target of a large Keck/DEIMOS campaign as part of the C3VO (Lemaux et al. 2022) survey. With the total integration time of ~ 28 h, we obtained 44 new secure spectral redshifts in the redshift range of the protocluster as compared to the nine secure spectral redshifts from VUDS at the time of the discovery. We also reported on 46 new secure spectral redshifts obtained in the redshift range of the coeval field. We combined this spectroscopic data with spectral redshifts from other spectral surveys and a variety of photometric catalogs in the COSMOS field. Using this wealth of data, we measured the SFRD of the Taralay protocluster and the surrounding field. Following are the main conclusions of this paper:

(i) Using the density mapping technique, we mapped out the Taralay protocluster at $z \sim 4.57$, established its internal structure and characterized its properties. Taralay protocluster displays two sub structures PC1 and PC2 for a density isopleth of $\sigma_\delta \geq 2$ that we use as a boundary to define the outline of this protocluster. While PC1 hosts two $\sigma_\delta \geq 5$ overdense peaks, PC2, the smaller substructure only hosts one $\sigma_\delta \geq 5$ overdense peak.

(ii) The mass of Taralay protocluster is estimated to be $1.74^{+1.36}_{-0.77} \times 10^{15} M_\odot$, which makes it exceptionally massive at these redshifts. This protocluster occupies a comoving volume $12620^{+1042}_{-956} \text{ cMpc}^3$.

(iii) We measured the SFRD of the field surrounding the Taralay protocluster to be $\log(\text{SFRD}/M_\odot \text{ yr}^{-1} \text{ Mpc}^{-3}) = -0.82^{+0.19}_{-0.29}$ and found it to be consistent with the most comparable study at these redshifts (Khusanova et al. 2021) but in moderate tension with that of Madau & Dickinson (2014).

(iv) We compare the masses estimated from overdensities calculated with the density mapping technique to the dynamical masses estimated from the LoS velocity dispersion for the protocluster, two $\sigma_\delta \geq 4.5$ regions and one $\sigma_\delta \geq 2.8$ region. We find that the masses estimated from the LoS velocity dispersion show a deficit in the range of $1.5\text{--}4\sigma$ with an average deficit of 2.5σ .

(v) The SFRD of the $z \sim 4.57$ Taralay galaxy members is ~ 12 times higher or $\log(\text{SFRD}/M_\odot \text{ yr}^{-1} \text{ Mpc}^{-3}) = 1.08 \pm 0.32$ in excess of the SFRD of the coeval field galaxies signifying that the environment does play a crucial role in driving the SFRD. Protoclusters like Taralay are clearly drivers of SM growth in the early universe.

(vi) We provide one of the first observational tests of simulation predictions that protoclusters contribute significantly to the fraction of cosmic SFR density in the early universe. Our findings indicate that protoclusters drive the SM growth in the early universe contributing $33.5 \text{ per cent}^{+8.0}_{-4.3} \text{ per cent}$ to the cosmic SFRD at $z \sim 4.57$, in 2.67σ excess over the $\sim 22 \text{ per cent}$ value predicted from simulations.

(vii) We find that the contribution to the cosmic SFRD from the $\sigma_\delta \geq 5$ peaks of the Taralay protocluster is $22.2 \text{ per cent}^{+5.4}_{-7.3} \text{ per cent}$, a significant portion of the total SFRD of the protocluster and indicative of the inside-out growth pattern as predicted by simulations.

(viii) We find that the $\sigma_\delta \geq 5$ peaks of the Taralay protocluster encase 68 per cent of the SFR while hosting less than 50 per cent of the galaxies. The SFRD of $\sigma_\delta \geq 5$ peaks is $\log(\text{SFRD}/M_\odot \text{ yr}^{-1} \text{ Mpc}^{-3}) = 0.87^{+0.18}_{-0.32}$.

(ix) We find a moderately strong, significant positive correlation between SFR and overdensity for galaxies in and around the Taralay protocluster.

In the future, we will be expanding this work to ensembles of protoclusters. With an ensemble of protoclusters, it is possible to divide the protoclusters by mass, dynamical state, redshift, etc. in order to better understand the underlying mechanisms which drive and quench the rapid SM growth. We also plan on including sub-millimetre observations to attempt to characterize the role of highly dusty star-forming galaxies in protocluster environments. It appears that such sources may prefer overdense environments (e.g. Romano et al. 2020, Loiacono et al. 2021, Fujimoto et al. 2023). The inclusion of these highly dusty star-forming galaxies has the potential to significantly increase the estimated SFRD of this protocluster, and thus its contribution to the overall SFRD of the universe at these redshifts. The results we have presented in Taralay are tantalizing and, if Taralay is indeed an exemplar of massive protoclusters at these redshifts, our results indicate that protoclusters play a key role in driving SM growth in the early universe.

ACKNOWLEDGEMENTS

We thank the anonymous referee for their valuable feedback. This work was supported by the National Science Foundation under grant no. 1908422. YF acknowledges support from JSPS KAKENHI Grant Number JP23K13149. We thank the late Olivier Le Fèvre for his pioneering work with galaxy redshift surveys, including the VIMOS Ultra-Deep Survey, without which this work would have likely not been possible. PS thanks Matthew Staab for providing crucial programming support that made the analysis of this paper possible. We thank the teams that did the hard work of compiling the various photometric catalogs in the COSMOS field for making these catalogs and their associated quantities public. A part of this work is based on the Galaxy Evolution Explorer (GALEX) satellite, Canada-France-Hawaii Telescope (CFHT), Hubble Space Telescope (HST), Subaru telescope, Visible and Infrared Survey Telescope for Astronomy (VISTA), and Spitzer Space Telescope. Some of the data presented herein were obtained at the W. M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W. M. Keck Foundation. The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Maunakea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain.

9. DATA AVAILABILITY

The data used for this study will be shared upon a reasonable request.

REFERENCES

- Adams S. M., Martini P., Croxall K. V., Overzier R. A., Silverman J. D., 2015, *MNRAS*, 448, 1335
- Aihara H. et al., 2018, *PASJ*, 70, S4
- Alberts S., Noble A., 2022, *Universe*, 8, 554
- Algera H. S. B. et al., 2023, *MNRAS*, 518, 6142
- Almeida C., Baugh C. M., Lacey C. G., 2011, *MNRAS*, 417, 2057
- Arnouts S. et al., 2002, *MNRAS*, 329, 355
- Ata M., Kitaura F.-S., Lee K.-G., Lemaux B. C., Kashino D., Cucciati O., Hernández-Sánchez M., Le Fèvre O., 2021, *MNRAS*, 500, 3194
- Bañados E., Venemans B., Walter F., Kurk J., Overzier R., Ouchi M., 2013, *ApJ*, 773, 178

- Balogh M. L., Baldry I. K., Nichol R., Miller C., Bower R., Glazebrook K., 2004, *ApJ*, 615, L101
- Barrufet L. et al., 2023, *MNRAS*, 522, 449
- Beers T. C., Flynn K., Gebhardt K., 1990, *AJ*, 100, 32
- B  thermin M. et al., 2020, *A&A*, 643, A2
- Blain A. W., Chapman S. C., Smail I., Ivison R., 2004, *ApJ*, 611, 725
- Boquien M., Burgarella D., Roehlly Y., Buat V., Ciesla L., Corre D., Inoue A. K., Salas H., 2019, *A&A*, 622, A103
- Bouwens R. J., Illingworth G. D., Franx M., Ford H., 2007, *ApJ*, 670, 928
- Bouwens R. J. et al., 2012, *ApJ*, 754, 83
- Bouwens R. J. et al., 2015, *ApJ*, 803, 34
- Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
- Butcher H., Oemler A. J., 1984, *ApJ*, 285, 426
- Calvi R., Castignani G., Dannerbauer H., 2023, *A&A*, 678, A15
- Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, *ApJ*, 533, 682
- Capak P. et al., 2007, *ApJS*, 172, 99
- Cardelli J. A., Clayton G. C., Mathis J. S., 1989, *ApJ*, 345, 245
- Carlberg R. G. et al., 1997, *ApJ*, 485, L13
- Casey C. M. et al., 2012, *ApJ*, 761, 140
- Casey C. M. et al., 2015, *ApJ*, 808, L33
- Casey C. M. et al., 2023, *ApJ*, 954, 31
- Charlot S., Fall S. M., 2000, *ApJ*, 539, 718
- Chartab N. et al., 2020, *ApJ*, 890, 7
- Chiang Y.-K., Overzier R., Gebhardt K., 2013, *ApJ*, 779, 127
- Chiang Y.-K. et al., 2015, *ApJ*, 808, 37
- Chiang Y.-K., Overzier R. A., Gebhardt K., Henriques B., 2017, *ApJ*, 844, L23
- Contini E., De Lucia G., Hatch N., Borgani S., Kang X., 2016, *MNRAS*, 456, 1924
- Cooke E. A., Hatch N. A., Muldrew S. I., Rigby E. E., Kurk J. D., 2014, *MNRAS*, 440, 3262
- Cooper M. C., Newman J. A., Davis M., Finkbeiner D. P., Gerke B. F., 2012, *Astrophysics Source Code Library*, record ascl:1203.003
- Cucciati O. et al., 2012, *A&A*, 539, A31
- Cucciati O. et al., 2014, *A&A*, 570, A16
- Cucciati O. et al., 2018, *A&A*, 619, A49
- Dahlen T., Mobasher B., Dickinson M., Ferguson H. C., Giavalisco M., Kretchmer C., Ravindranath S., 2007, *ApJ*, 654, 172
- Dannerbauer H. et al., 2014, *A&A*, 570, A55
- Dav   R., 2008, *MNRAS*, 385, 147
- Davidzon I. et al., 2016, *A&A*, 586, A23
- Davidzon I. et al., 2017, *A&A*, 605, A70
- Davis M. et al., 2003, in Guhathakurta P. ed., *Proc. SPIE Conf. Ser. Vol. 4834, Discoveries and Research Prospects from 6- to 10-Meter-Class Telescopes II*. SPIE, Bellingham, p. 161
- Diener C. et al., 2013, *ApJ*, 765, 109
- Diener C. et al., 2015, *ApJ*, 802, 31
- Durkalec A. et al., 2018, *A&A*, 612, A42
- Einasto J., H  tsi G., Liivam  gi L.-J., Park C., Kim J., Szapudi I., Einasto M., 2023, *MNRAS*, 523, 4693
- Everett W. B. et al., 2020, *ApJ*, 900, 55
- Faber S. M. et al., 2003, in Iye M., Moorwood A. F. M. eds, *Proc. SPIE Conf. Ser. Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes*. SPIE, Bellingham, p. 1657,
- Faisst A. L. et al., 2020, *ApJS*, 247, 61
- Forrest B. et al., 2017, *ApJ*, 838, L12
- Forrest B. et al., 2020a, *ApJ*, 890, L1
- Forrest B. et al., 2020b, *ApJ*, 903, 47
- Forrest B. et al., 2023, *MNRAS*, 526, L56
- Franco M. et al., 2018, *A&A*, 620, A152
- Franco M. et al., 2020, *A&A*, 643, A30
- Fujimoto S. et al., 2023, preprint (arXiv:2303.01658)
- Fuller S. et al., 2020, *ApJ*, 896, 156
- Giavalisco M., 2005, *New Astron. Rev.*, 49, 440
- Ginolfi M. et al., 2020, *A&A*, 643, A7
- G  mez P. L. et al., 2003, *ApJ*, 584, 210
- Gould K. M. L. et al., 2023, *AJ*, 165, 248
- Greenslade J. et al., 2018, *MNRAS*, 476, 3336
- Griffin M. J. et al., 2010, *A&A*, 518, L3
- Gruppioni C. et al., 2013, *MNRAS*, 432, 23
- Gruppioni C. et al., 2020, *A&A*, 643, A8
- Guo Q., White S., Angulo R. E., Henriques B., Lemson G., Boylan-Kolchin M., Thomas P., Short C., 2013, *MNRAS*, 428, 1351
- Hamadouche M. L. et al., 2023, *MNRAS*, 521, 5400
- Hasinger G. et al., 2018, *ApJ*, 858, 77
- Hatch N. A., Kurk J. D., Pentericci L., Venemans B. P., Kuiper E., Miley G. K., R  ttgering H. J. A., 2011, *MNRAS*, 415, 2993
- Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, *MNRAS*, 451, 2663
- Higuchi R. et al., 2019, *ApJ*, 879, 28
- Hill R. et al., 2022, *MNRAS*, 512, 4352
- Hu W. et al., 2021, *Nat. Astron.*, 5, 485
- Huang Y. et al., 2022, *ApJ*, 941, 134
- Hung D. et al., 2020, *MNRAS*, 491, 5524
- Ichikawa T. et al., 2006, in McLean I. S., Iye M., eds, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 6269, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. SPIE, Bellingham, p. 626916
- Ilbert O. et al., 2006, *A&A*, 457, 841
- Ito K. et al., 2020, *ApJ*, 899, 5
- Ito K. et al., 2023, *ApJ*, 945, L9
- Kakimoto T. et al., 2023, A massive quiescent galaxy in a group environment at $z = 4.53$, preprint (arXiv:2308.15011)
- Kashino D. et al., 2022, *ApJ*, 925, 82
- Khusanova Y. et al., 2020, *A&A*, 634, A97
- Khusanova Y. et al., 2021, *A&A*, 649, A152
- Kimble R. A., MacKenty J. W., O'Connell R. W., Townsend J. A., 2008, in Oschmann Jacobus M. J., de Graauw M. W. M., MacEwen H. A., eds, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 7010, Space Telescopes and Instrumentation 2008: Optical, Infrared, and Millimeter*. SPIE, Bellingham, p. 70101E
- Kistler M. D., Y  ksel H., Beacom J. F., Hopkins A. M., Wyithe J. S. B., 2009, *ApJ*, 705, L104
- Konishi M. et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV*. SPIE, Bellingham, p. 84467P
- Koprowski M. P., Dunlop J. S., Micha  owski M. J., Coppin K. E. K., Geach J. E., McLure R. J., Scott D., van der Werf P. P., 2017, *MNRAS*, 471, 4155
- Koyama Y. et al., 2021, *MNRAS*, 503, L1
- Kubo M. et al., 2019, *ApJ*, 887, 214
- Laigle C. et al., 2016, *ApJS*, 224, 24
- Lammers C., Hill R., Lim S., Scott D., Ca  ameras R., Dole H., 2022, *MNRAS*, 514, 5004
- Le F  vre O. et al., 2015, *A&A*, 576, A79
- Le F  vre O. et al., 2019, *A&A*, 625, A51
- Le F  vre O. et al., 2020, *A&A*, 643, A1
- Lemaux B. C. et al., 2014, *A&A*, 572, A41
- Lemaux B. C., Tomczak A. R., Lubin L. M., Wu P. F., Gal R. R., Rumbaugh N., Kocevski D. D., Squires G. K., 2017, *MNRAS*, 472, 419
- Lemaux B. C. et al., 2018, *A&A*, 615, A77
- Lemaux B. C. et al., 2019, *MNRAS*, 490, 1231
- Lemaux B. C. et al., 2022, *A&A*, 662, A33
- Li Z., Steidel C. C., Gronke M., Chen Y., Matsuda Y., 2022, *MNRAS*, 513, 3414
- Lilly S. J. et al., 2007, *ApJS*, 172, 70
- Lilly S. J. et al., 2009, *ApJS*, 184, 218
- Loiacono F. et al., 2021, *A&A*, 646, A76
- Long A. S. et al., 2020, *ApJ*, 898, 133
- Lubin L. M., Gal R. R., Lemaux B. C., Kocevski D. D., Squires G. K., 2009, *AJ*, 137, 4867
- Madau P., Dickinson M., 2014, *ARA&A*, 52, 415
- Madden S. C. et al., 2020, *A&A*, 643, A141
- Magnelli B. et al., 2013, *A&A*, 553, A132

- Malmquist K. G., 1922, Meddelanden fran Lunds Astronomiska Observatorium Serie I, 100, 1
- Malmquist K. G., 1925, Meddelanden fran Lunds Astronomiska Observatorium Serie I, 106, 1
- Maraston C., 2005, *MNRAS*, 362, 799
- Martin D. C. et al., 2005, *ApJ*, 619, L1
- Matsuda Y. et al., 2011, *MNRAS*, 416, 2041
- McConachie I. et al., 2022, *ApJ*, 926, 37
- McLean I. S., Steidel C. C., Matthews K., Epps H., Adkins S. M., 2008, in McLean I. S., Casali M. M., eds, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy II*. SPIE, Bellingham, p. 70142Z
- Mei S. et al., 2023, *A&A*, 670, A58
- Miley G., De Breuck C., 2008, *A&AR*, 15, 67
- Miller T. B. et al., 2018, *Nature*, 556, 469
- Miyazaki S. et al., 2018, *PASJ*, 70, S1
- Mocanu L. M. et al., 2013, *ApJ*, 779, 61
- Muldrew S. I., Hatch N. A., Cooke E. A., 2015, *MNRAS*, 452, 2528
- Muldrew S. I., Hatch N. A., Cooke E. A., 2018, *MNRAS*, 473, 2335
- Murakami H. et al., 2007, *PASJ*, 59, S369
- Muzzin A. et al., 2012, *ApJ*, 746, 188
- Nantais J. B. et al., 2017, *MNRAS*, 465, L104
- Negrello M. et al., 2017, *MNRAS*, 470, 2253
- Neugebauer G. et al., 1984, *ApJ*, 278, L1
- Newman J. A. et al., 2013, *ApJS*, 208, 5
- Newman A. B. et al., 2022, *Nature*, 606, 475
- Novak M. et al., 2017, *A&A*, 602, A5
- Nowotka M., Chen C.-C., Battaia F. A., Fumagalli M., Cai Z., Lusso E., Prochaska J. X., Yang Y., 2022, *A&A*, 658, A77
- O'Donnell J. E., 1994, *ApJ*, 422, 158
- O'Shea B. W., Wise J. H., Xu H., Norman M. L., 2015, *ApJ*, 807, L12
- Old L. J. et al., 2020, *MNRAS*, 493, 5987
- Orsi Á. A., Fanidakis N., Lacey C. G., Baugh C. M., 2016, *MNRAS*, 456, 3827
- Ouchi M. et al., 2004, *ApJ*, 611, 660
- Ouchi M. et al., 2005, *ApJ*, 620, L1
- Overzier R. A., 2016, *A&AR*, 24, 14
- Overzier R., Kashikawa N., 2019, *BAAS*, 51, 180
- Pavesi R. et al., 2018, *ApJ*, 861, 43
- Pei Y. C., 1992, *ApJ*, 395, 130
- Pelliccia D. et al., 2019, *MNRAS*, 482, 3514
- Pelló R. et al., 2018, *A&A*, 620, A51
- Planck Collaboration et al., 2015, *A&A*, 582, A30
- Popescu R. et al., 2023, *ApJ*, 958, 12
- Ramakrishnan V. et al., 2023, *ApJ*, 951, 119
- Reddy N. A., Steidel C. C., 2009, *ApJ*, 692, 778
- Riechers D. A. et al., 2020, *ApJ*, 895, 81
- Romano M. et al., 2020, *MNRAS*, 496, 875
- Rowan-Robinson M. et al., 2016, *MNRAS*, 461, 1100
- Ryan R. E. J. et al., 2014, *ApJ*, 786, L4
- Salim S., Narayanan D., 2020, *ARA&A*, 58, 529
- Salmon B. et al., 2015, *ApJ*, 799, 183
- Salpeter E. E., 1955, *ApJ*, 121, 161
- Sanders D. B. et al., 2007, *ApJS*, 172, 86
- Sawicki M., Thompson D., 2006, *ApJ*, 642, 653
- Schaerer D. et al., 2020, *A&A*, 643, A3
- Schechter P., 1976, *ApJ*, 203, 297
- Schenker M. A. et al., 2013, *ApJ*, 768, 196
- Scoville N. et al., 2007, *ApJS*, 172, 38
- Shah E. A. et al., 2023, Identification and Characterization of Six Spectroscopically Confirmed Massive Protostructures at $2.5 < z < 4.5$, preprint (arXiv:2312.04634)
- Shapley A. E., Steidel C. C., Pettini M., Adelberger K. L., 2003, *ApJ*, 588, 65
- Shen L. et al., 2017, *MNRAS*, 472, 998
- Shen L. et al., 2018, *MNRAS*, 477, 4707
- Shen L. et al., 2019, *MNRAS*, 484, 2433
- Shen L. et al., 2021, *ApJ*, 912, 60
- Shen L. et al., 2022, *ApJ*, 935, 177
- Shi K. et al., 2019, *ApJ*, 879, 9
- Shi K., Toshikawa J., Cai Z., Lee K.-S., Fang T., 2020, *ApJ*, 899, 79
- Shi K., Toshikawa J., Lee K.-S., Wang T., Cai Z., Fang T., 2021, *ApJ*, 911, 46
- Smolčić V. et al., 2017, *A&A*, 597, A4
- Sobral D., Best P. N., Smail I., Mobasher B., Stott J., Nisbet D., 2014, *MNRAS*, 437, 3516
- Steidel C. C., Adelberger K. L., Shapley A. E., Erb D. K., Reddy N. A., Pettini M., 2005, *ApJ*, 626, 44
- Steinhardt C. L. et al., 2014, *ApJ*, 791, L25
- Sutherland W. et al., 2015, *A&A*, 575, A25
- Tasca L. A. M. et al., 2015, *A&A*, 581, A54
- Thomas R. et al., 2017, *A&A*, 602, A35
- Tomczak A. R. et al., 2016, *ApJ*, 817, 118
- Tomczak A. R. et al., 2017, *MNRAS*, 472, 3512
- Tomczak A. R. et al., 2019, *MNRAS*, 484, 4695
- Toshikawa J. et al., 2012, *ApJ*, 750, 137
- Toshikawa J. et al., 2016, *ApJ*, 826, 114
- Toshikawa J., Malkan M. A., Kashikawa N., Overzier R., Uchiyama H., Ota K., Ishikawa S., Ito K., 2020, *ApJ*, 888, 89
- Uchiyama H. et al., 2022, *PASJ*, 74, L27
- Venemans B. P. et al., 2007, *A&A*, 461, 823
- Vieira J. D. et al., 2010, *ApJ*, 719, 763
- Vizgan D. et al., 2022, *ApJ*, 929, 92
- Wagner C. R., Courteau S., Brodwin M., Stanford S. A., Snyder G. F., Stern D., 2017, *ApJ*, 834, 53
- Wang T. et al., 2016, *ApJ*, 828, 56
- Wang T. et al., 2019a, *Nature*, 572, 211
- Wang L., Pearson W. J., Cowley W., Trayford J. W., Béthermin M., Gruppioni C., Hurley P., Michałowski M. J., 2019b, *A&A*, 624, A98
- Weaver J. R. et al., 2022, *ApJS*, 258, 11
- Werner M. W. et al., 2004, *ApJS*, 154, 1
- Wright E. L. et al., 2010, *AJ*, 140, 1868
- Wylezalek D. et al., 2013, *ApJ*, 769, 79
- Yonekura N., Kajisawa M., Hamaguchi E., Mawatari K., Yamada T., 2022, *ApJ*, 930, 102
- Yoshida M. et al., 2006, *ApJ*, 653, 988
- Yung L. Y. A., Somerville R. S., Finkelstein S. L., Popping G., Davé R., 2019, *MNRAS*, 483, 2983
- Zanella A. et al., 2018, *MNRAS*, 481, 1976
- Zhou L. et al., 2020, *A&A*, 642, A155
- de La Vieuville G. et al., 2019, *A&A*, 628, A3
- van Dokkum P. G., 2008, *ApJ*, 674, 29
- von der Linden A., Wild V., Kauffmann G., White S. D. M., Weinmann S., 2010, *MNRAS*, 404, 1231

APPENDIX A: CAVEATS

Here we discuss some of the key caveats that should be considered when interpreting the results of our study.

A1 Estimating the overdensity

In this study, we assume that galaxies can serve as tracers of the underlying matter density field and measure the overdensity in terms of galaxies (e.g. Cucciati et al. 2014). Different types of galaxies, such as quiescent or star forming, trace the matter density field differently. Sometimes matter density is not traced very well by UV-selected galaxies, as in the case of Newman et al. (2022), which finds unexpectedly low galaxy overdensity where large-scale Ly α absorption is strongest indicating high matter density. Moreover, observations might not trace the true underlying galaxy population as low-luminosity galaxies are harder to detect, especially at high redshifts. By considering the mass of galaxies and their SFRs, we determine an appropriate bias factor from simulations that scales the

galaxy overdensity to the matter overdensity. For this study, we use a bias factor of $b = 3.6$, which is based on previous research (e.g. Chiang et al. 2013, Durkalec et al. 2018) and the upper and lower limit for the bias factor are 4.5 and 3.12, respectively, obtained from Einasto et al. (2023) and Ata et al. (2021).

A2 SED fitting

The SFR derived from SED fitting method correlates well with the SFR derived from independent measures such as [CII] lines at intermediate and high redshift (e.g. Schaerer et al. 2020). We checked if this correlation holds for galaxies that we obtained from Y. Fudamoto in *private communication* based on serendipitously detected [CII] lines with ALMA. These three galaxies have $\log(L_{\text{CII}}/L_{\odot})$ of 8.99, 9.01, and 9.10, respectively. When we compared the [CII] derived SFR with the SFR values obtained for these three galaxies through SED fitting using the parameter set given in Section 5.1, we found that the SED-fit SFR values are consistently lower than the [CII] derived SFR. If we force the fits such that the best model chooses the high end of the E(B-V) value range, the difference between the [CII]-derived SFR and the SED-fit value decreases. However, such fits are clearly disfavored by comparing their reduced χ^2 . Note that recent theoretical and observational studies suggest that [CII] is more tightly connected to the molecular gas mass in a complicated way through Kennicutt–Schmidt relation (e.g. Zanella et al. 2018; Madden et al. 2020; Vizgan et al. 2022) than to SFR.

A3 Absence of FIR data

We tested the reliability of SFR derived from SED fitting with optical/NIR data (used in this study). We selected 12 galaxies from ALPINE survey that fall in the redshift range of $4.2 < z < 4.93$ with significant detection in the continuum around restframe 155 μm . We performed the SED fitting on these galaxies with the same parameters as listed in Table 10 along with (Casey et al. 2012) dust template. We found that, relative to the fits that include the FIR data, the optical/NIR-only fits show higher SFRs by ~ 0.2 dex. For these galaxies, the extinction correction is generally overestimated by the optical/NIR-only fits. Because these 12 galaxies may or may not be representative of the true galaxy population, as these galaxies are likely the more massive/more dusty extreme of the ALPINE population, we decided to not make any changes to our analysis. Programs probing sub-mm galaxies in protoclusters are needed in order to get a fuller perspective on the star formation in protoclusters. To check how much the result of our study would change if, indeed, what we find were a generally applicable result, we reduced the SFR of each galaxy in our analysis, SFRs that are derived by optical/NIR-only fitting, by 0.2 dex and performed the analysis again. The result was that the SFRD of Taralay and the field were reduced, but well within the uncertainties. The fractional contribution of Taralay-like systems to overall SFRD at these redshifts remained the same.

A4 Dust properties of bright and faint galaxies

Here we test our assumption that the bright and faint galaxies experience the same level of dust extinction which let us extrapolate the SFRD results to include the lower luminosity galaxies in Section 5.3. If this assumption is wrong, we are either underestimating or overestimating the SFR of the fainter galaxies. This is because the different dust extinction, through the process of SED fitting, will cause the SFR to be different for the fainter galaxies compared to the SFR of the brighter galaxies. To test this assumption, we analyse

the SED models for nearly 3000 galaxies at $2.5 < z < 3.5$ in the COSMOS field. These galaxies make up our sample because at this redshift range we can probe deeper than $M_{\text{FUV}} = -19.3$, the depth of our data, at $4.2 < z < 4.93$. For this exercise, a galaxy is considered bright if its FUV absolute magnitude is less than -19 , the delineation point between bright and faint.

We perform a KS test on the $E(B - V)$ distribution of both bright and faint galaxies at $2.5 < z < 3.5$, and find that they prefer the same range of values for the colour excess of the nebular line light, for both young and old populations. The range for E.BV_lines is presented in Table 10. This finding shows that at $2.5 < z < 3.5$, the SFR traced by FUV photons is similarly affected by the dust properties of both faint and bright galaxies. We assume that this behaviour is consistent for galaxies at $4.2 < z < 4.93$ and make no further changes to our FUV magnitude correction factor.

A5 Comparison to simulations

Our study diverges from the definition and characterization of protoclusters as outlined in Chiang et al. (2017) in several key aspects. In their work, a protocluster encompasses all the dark matter and baryonic matter that will eventually merge into a cluster by $z = 0$, with a mass exceeding $10^{14} M_{\odot}$ within R_{200} . The protocluster’s volume encompasses all the matter that will contribute to the formation of the cluster by $z = 0$. In contrast, our study defines a protocluster as a structure contained within the 2σ density isopleth.

Furthermore, while Chiang et al. (2017) designates the most massive halo within the protocluster as its core at any given epoch, having a size of ~ 0.4 cMpc at $z \sim 4.57$, our study identifies the protocluster core as the region bounded inside a 5σ density isopleth with a size ~ 2 cMpc. To put our peaks on the same footing as the core defined in Chiang et al. (2017), our peak size will have to reduce beyond $\sigma_8 > 10$ where we have spectroscopic data only for ~ 3 galaxies. Hence, we adopt our definition for peaks. It is crucial to acknowledge that these differing definitions of protoclusters may contribute to the measured difference between the observed and simulated data.

A6 Using a massive protocluster as a proxy

Using the density maps, the estimated mass of Taralay protocluster at $z \sim 4.57$ is $1.74 \times 10^{15} M_{\odot}$ making this structure exceptionally massive at these redshifts.

Using such a massive protocluster, which has highly star-forming galaxies, as a representative sample of all the protoclusters at $z \sim 4.57$ to calculate the fraction of SFRD from protoclusters (see Fig. 10) may lead to an overestimated value for the derived contribution of protoclusters to the SFRD. Nevertheless, there is a possibility that environmental quenching is affecting the galaxies within this protocluster (Lemaux et al. 2018) which could result in a decrease in the SFRD. To draw a definitive conclusion, a large collection of protoclusters is required for comprehensively testing their contribution to the SFRD at high redshift.

APPENDIX B: EFFECT OF PARAMETER CHOICES ON THE MODEL SPECTRUM

The SFR values that our results are based on come from the SED fitting process. It is important therefore to discuss how the different parameters that are used in the SED fitting process impact the spectrum of a galaxy, affecting the estimated SFR. We discuss this impact here using two different approaches. First, we look at the

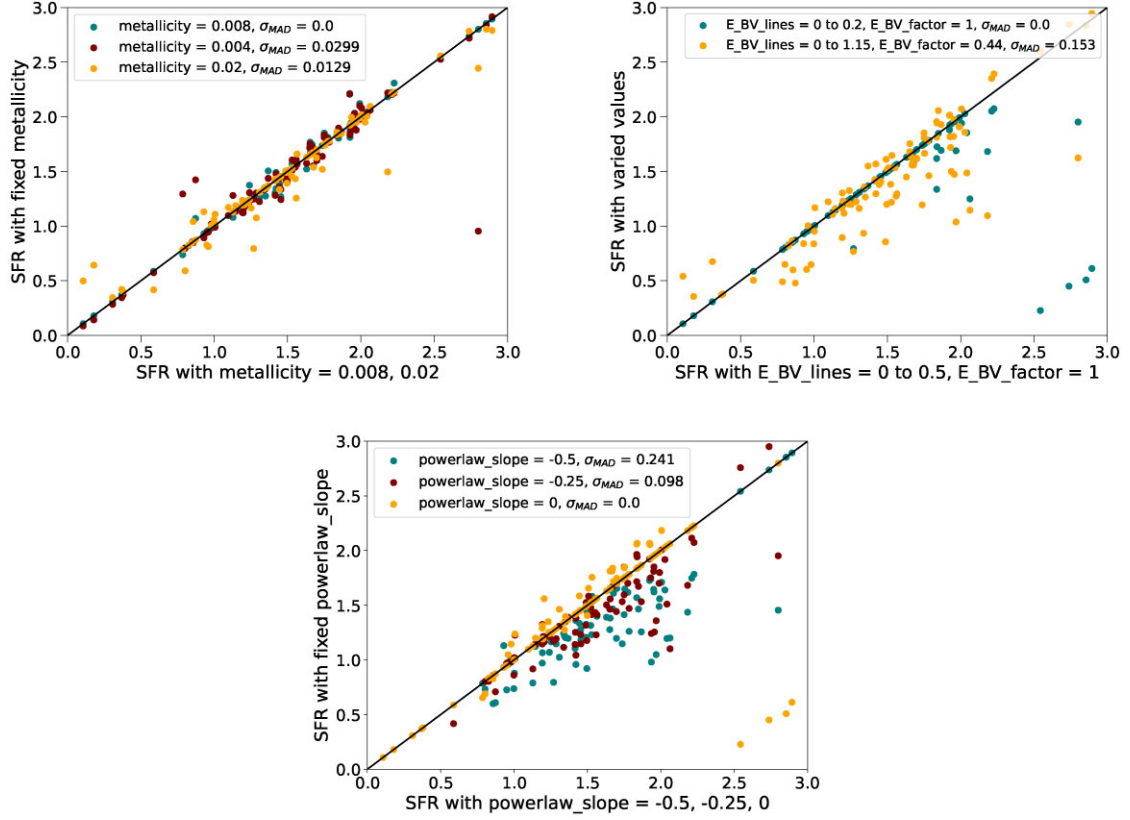


Figure B1. Comparing SFR values estimated using the ranges of parameters given in Table 10 with the SFR values estimated by forcing metallicity to take on a fix value in the top left-hand panel, changing the range of E(B-V) lines in the top right panel and forcing the powerlaw slope to a fix value in the bottom panel. For each plot, the x-axis represents SFR obtained from SED fitting the galaxies using a range of a particular parameter. These ranges for metallicity, E(B-V) lines and powerlaw_slope are shown on the x-axis of each plot. The y-axis represents SFR obtained from SED fitting galaxies using a fixed parameter. For both axes for each plot, the SFR values are in $\log(\text{SFR}/M_{\odot}\text{yr}^{-1})$. The σ_{MAD} for a distribution in the each figure shows the scatter on the SFR value. The $\sigma_{MAD} = 0$ comes from having a very few outliers with the majority of the galaxies having one to one correlation.

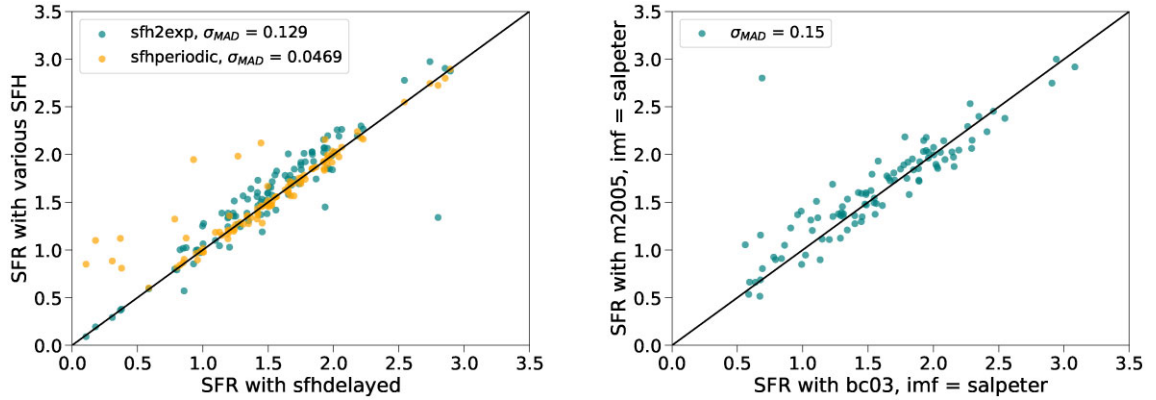


Figure B2. Comparing SFR value obtained from SED fitting using different modules of SFH and SSP. The x-axis of the first plot represents SFR obtained from doing SED fitting with sfhdelayed module while the y-axis represents SFR obtained from doing SED fitting with sfh2exp and sfhperiodic modules. For the second plot, x-axis represents SFR obtained from doing SED fitting with bc03 module with Salpeter IMF while the y-axis represents SFR obtained from doing SED fitting with m2005 module with Salpeter IMF. For both axes for each plot, the SFR values are in $\log(\text{SFR}/M_{\odot}\text{yr}^{-1})$. The σ_{MAD} for a distribution in the each figure shows the scatter on the SFR value.

scatter on the SFR of a large sample of galaxies which sheds light on how the SFR itself varies with different parameters used for SED fitting. Then we look at a model SED of a single galaxy to understand how the SED is impacted with the changes in one parameter while keeping the rest of the parameter inputs constant.

In Fig. B1, we investigate the change in SFR values by fixing metallicity and powerlaw_slope values and by varying E(B-V) lines range. The sample chosen to perform this exercise is one hundred galaxies and the same modules are chosen to estimate their SFR with SED fitting as used in the analysis of this study, i.e. sfhdelayed, bc03,

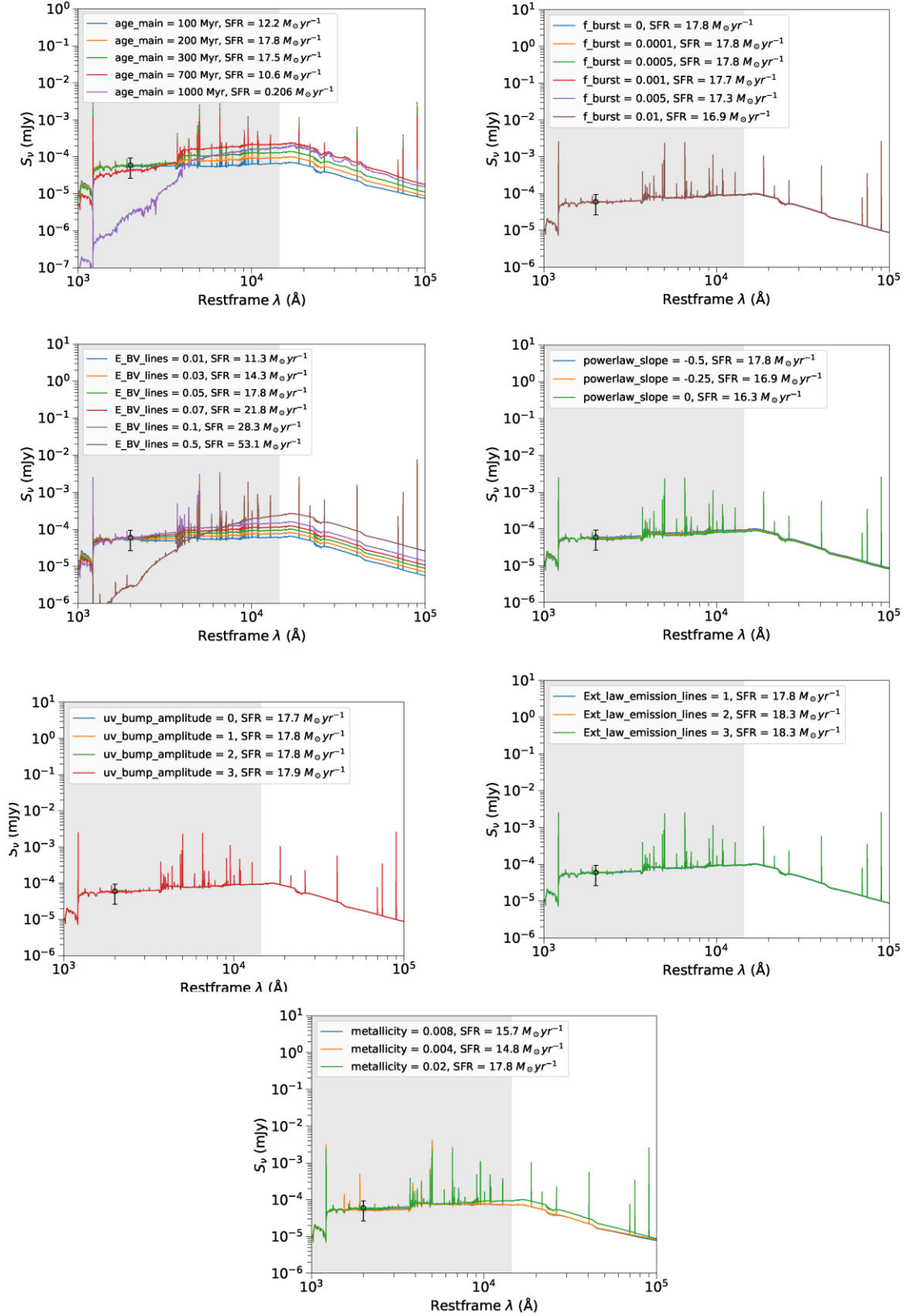


Figure B3. Rest-frame SED for a single galaxy with different parameters (see legends of each panels). The SEDs from each panel are such that they have nearly the same SFRs while changing the parameters indicated in the legends in the panels. Data points at $\lambda_{\text{rest}} = 2000 \text{ \AA}$ in each panels indicate typical errors of the observations. The differences in the SED of the galaxy are within the uncertainty except for the SED obtained with age_{main} of 1000 Myr and $E_{\text{BV_lines}}$ of 0.5.

nebular, dustatt_modified_starburst, redshifting (See Section 5.1 for the ranges of these parameters). We find that the scatter on the SFR resulting from changing metallicity is fairly small indicating that a specific value of metallicity is not a strong driver of change in SFR. The scatter on SFR from varying the powerlaw_slope and E.BV lines is larger indicating these parameters may have an impact on the accuracy of the estimated SFR. The largest systematic uncertainty of ~ 0.2 dex coming from the powerlaw_slope value of -0.5 affecting the SFR of both the protocluster and the coeval field galaxies equally. Assuming that factors like IMF, E(B-V), powerlaw_slope do not dramatically change for the two populations, this uncertainty does not change our results since we compare the SFR of the two populations.

In the left-hand panel of Fig. B2 we compare the SFR estimated with sfhdelayed, the SFH module chosen for perform this study, with two of the other SFH modules available from CIGALE to see how the change in our choice of module can impact our results. We find that the scatter on the SFR is small and a particular choice of SFH is not likely to impact the estimated SFR significantly. In the right-hand panel of this figure, we change the library of SSP to investigate the impact of using bc03 module instead of m2005 module. Because m2005 module is not compatible with the Chabrier IMF originally used in the SED fitting of this study, for this exercise, we use the Salpeter IMF with both bc03 and m2005 module. We do not see a significant difference in the estimated SFR by choosing different SSP modules.

Next, we investigate the impact of using different parameter ranges on the spectrum of a single galaxy. The SED of the example galaxy in Fig. B3 changes drastically in the rest-frame UV (shaded region) covered by our data for age_main of 1000 Myr and E.BV_lines of 0.5 showing that it is mainly these two factors that can affect the spectrum of a galaxy. However, our data cannot differentiate between the models with age_main differences of ~ 100 Myr and E.BV_lines differences of ~ 0.05 as the variations in the models are within the uncertainty of our data (typically of order $\sim 1e^{-5}$ mJy for this example).

APPENDIX C: SED FITTING WITH CIGALE VERSUS LEPHARE

We compare various physical parameters obtained from SED fitting with CIGALE using the parameter set given in Section 5.1 to the

best-fitting parameters given in the COSMOS2020 catalog (Weaver et al. 2022) and find no significant difference that can be attributed to the choice of using a particular SED fitting software. Here we show this comparison for SM and SFR. The sample used for comparison consists of 106 galaxies in the redshift range of $4.2 < z < 4.93$ that have secure spectroscopic redshifts. This sample has also undergone an IRAC channel 1 and/or IRAC channel 2 cut such that the 106 galaxies have magnitudes brighter than the IRAC channel 1 and/or IRAC channel 2 completeness limit listed in (Weaver et al. 2022).

¹Department of Physics and Astronomy, University of California, Davis, One Shields Ave., Davis, CA 95616, USA

²Gemini Observatory, NSF's NOIRLab, 670 N. A'ohoku Place, Hilo, HI 96720, USA

³INAF Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Piero Gobetti 93/3, 40129 Bologna, Italy

⁴University of Hawai'i, Institute for Astronomy, 2680 Woodlawn Drive, Honolulu, HI 96822, USA

⁵Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843-4242 USA

⁶George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, Texas A&M University, College Station, TX 77843-4242 USA

⁷Aix Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326, F-13388 Marseille, France

⁸Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

⁹INAF-IASF Milano, Via Alfonso Corti 12, I-20159 Milano, Italy

¹⁰Dipartimento di Fisica e Astronomia Galileo Galilei, Università degli Studi di Padova, Vicolo dell'Osservatorio 3, I-35122 Padova Italy

¹¹Academia Sinica Institute of Astronomy and Astrophysics (ASIAA), No. 1, Section 4, Roosevelt Road, Taipei 10617, Taiwan

¹²Center for Frontier Science, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan

¹³Department of Pure and Applied Physics, Graduate School of Advanced Science and Engineering, Faculty of Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku, Tokyo 169-8555, Japan

¹⁴INAF-IASF, via Bassini 15, I-20133, Milano, Italy

¹⁵Astronomy Department, University of Massachusetts, Amherst, MA, 01003 USA

¹⁶Instituto de Astrofisica, Departamento de Ciencias Fisicas, Facultad de Ciencias Exactas, Universidad Andres Bello, Fernandez Concha 700, Las Condes, Santiago RM, Chile

¹⁷University of Bologna – Department of Physics and Astronomy 'Augusto Righi' (DIFA, Via Gobetti 93/2, I-40129 Bologna, Italy

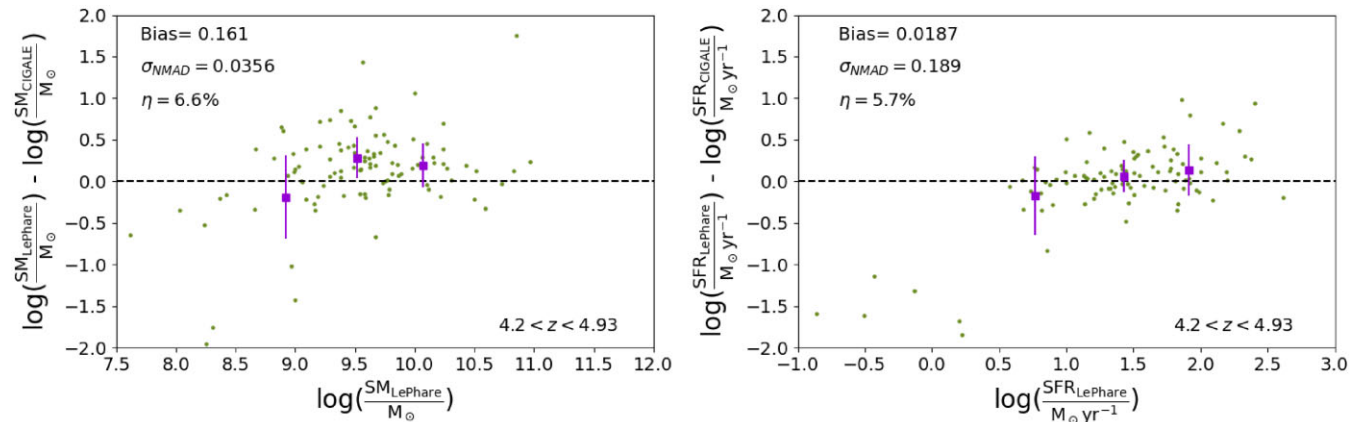


Figure C1. Difference between SM and SFR estimated with CIGALE and LePhare, two SED fitting software. The σ_{NMAD} , catastrophic outlier rate ($\eta = |\log(X_{LePhare}) - \log(X_{CIGALE})| / (1 + \log(X_{LePhare})) > 3\sigma_{NMAD}$), and bias which is defined as the median of the difference between $\log(X_{LePhare})$ and $\log(X_{CIGALE})$ are shown for both panels where X represents SM/M_\odot for the left-hand panel and $SFR/M_\odot \text{ yr}^{-1}$ for the right-hand panel.

¹⁸*Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*

¹⁹*Institute for Advanced Research, Nagoya University, Nagoya 464-8601, Japan*

²⁰*Department of Physics, Graduate School of Science, Nagoya University, Nagoya 464-8602, Japan*

²¹*California Institute of Technology, MC 249-17, 1200 East California Blvd., Pasadena, CA 91125, USA*

This paper has been typeset from a \TeX/L\AA\TeX file prepared by the author.