Facultad de Ciencias Físicas
Máster en Astrofísica

ALMA [OIII] 88μm Emitters. Signpost of Early Stellar Buildup and Reionization in the Universe.

Galaxias detectadas en [OIII] 88μm con ALMA. Indicios de la reionización y formación estelar temprana.

Por:
Rafael Navarro Carrera

Tutores:
Luis Colina (Centro de Astrobiología, INTA-CSIC)
Javier Álvarez (Centro de Astrobiología, INTA-CSIC)

Trabajo de Fin de Máster

Madrid, Junio de 2022

Contents

1 Abstract

Future observations of galaxies in the epoch of reionization (EoR) with James Webb Space Telescope (JWST) will unveil new information about these objects. The present work will evaluate the observational strategy of General Observing Time, Cycle 1 (GO1) proposal #1840 [Alvarez-Marquez et al., 2021]. The proposal makes use of JWST's Near Infrared (NIR) Camera instrument (NIRcam) for photometric measurements and Near Infrared Spectrometer (NIRSpec) for spectroscopic measurements.

Observations will contribute to determining physical and spatial galactic structure whilst also tracing star populations and interstellar medium (ISM). In this work, simulations of NIRCam observations will be first made, in order to test the observational strategy of the proposal, and then reduced with JWST pipeline in preparation for reduction of real data.

Finally spectral energy distribution (SED) fitting will be applied in order to test whether more information can be extracted from the sources: star formation history (SFH) or physical properties such as stellar mass.

After a detailed analysis of the observational strategy, it has been concluded that #1840 proposal GO1 observations will provide with new information about spatial structure of epoch of reionization sources, among estimations of their star formation histories and physical properties by performing a careful spectral energy distribution fitting.

Resumen

Futuras observaciones de galaxias de la época de reionización con el Telescopio Espacial James Webb (JWST) permitirán extraer información extremadamente novedosa acerca de estos objetos. El presente trabajo evaluará la estrategia observacional que se va a seguir en las observaciones de General Observing Time Cycle 1 (GO1), propuesta #1840 [Alvarez-Marquez et al., 2021]. La propuesta hace uso de los instrumentos NIRCam (Near Infrared Camera) para realizar fotometría y NIRSpec (Near Infrared Spectrometer) para realizar espectroscopía.

Las observaciones permitirán determinar la estructura espacial de las fuentes y la distribución de las diferentes poblaciones estelares y del medio interestelar. En el presente trabajo se simualrán observaciones con NIRCam para comprobar la estrategia observacional de la propuesta, y posteriormente se reducirán empleando la pipeline de JWST, lo que será útil para entrenar el proceso de reducción de cara a futuras observaciones.

Tras todo ello, se estudiarán técnicas como el ajuste de distribuciones espectrales de energía (SEDs) para extraer información acerca de cómo estos objetos formaron sus poblaciones estelares, técnica que permitirá caracterizar algunas de sus propiedades físicas.

Tras un análisis detallado de la propuesta observacional, se ha podido concluir que las observaciones de la propuesta #1840 GO1 aportarán nueva información sobre la estructura de las fuentes observadas, y que a partir de ellas será posible estimar la historia de formación estelar y algunos parámetros físicos empleando la técnica de ajuste de SEDs.

2 About the project

The Epoch of Reionization (EoR) is the period of time when matter in the universe stopped being neutral. The causes and origin of this phenomenon are not yet fully understood. New information on galaxies from the EoR: physical properties, age of stellar populations and emission fraction of ionizing photons will allow us to further deepen our knowledge of the cosmic dawn. Thanks to the exceptional angular resolution and sensibility of JWST, a unique window into the early universe will open. New discoveries about high redshift (z) objects, until now just detected but not well characterized or even totally unknown, will start to appear during JWST's lifespan.

GO1 JWST proposal #1840 [Alvarez-Marquez et al., 2021] will shed light on the EoR by means of studying a sample of high z, [OIII] 88 μm emitting galaxies previously detected with Atacama Large Millimiter Array (ALMA) in order to understand how their photons can ionize atoms. At the same time they will characterize their morphology and the properties of the ISM. This work is preparation for future JWST observations. It will consist of simulations of NIRCam observations, data reduction tests with JWST's pipeline and analysis of SED fitting techniques to extract physical properties of EoR objects.

2.1 Scientific goals

Objetive 1: Stellar and gaseous structure of the host galaxies. By observing stellar and gaseous structure it might be possible to differentiate between clumpy or disk driven galaxies. Using deep NIR ($0.16 \sim 0.55\ \mu m$) imaging techniques with JWST's NIRCam instrument, structure of different galactic components is expected to be spatially resolved (best case 350 pc at $z \sim 7$). Young ($\leq 10\ Myr$), mature ($\geq 100\ Myr$) star populations, and ISM components can be individually traced by observing rest frame UV, NIR and optical continuum plus nebular line emission.

The synergy between JWST and ALMA allows to better study the structure of ISM. High resolution [OIII]88 μm and [CII]158 μm emission detected by ALMA together with deep NIRCam imaging and NIRSpec spectroscopy will allow to characterize in detail the ISM structure of targets.

Objetive 2: Age and mass of the stellar populations. By characterizing mass and age of the star populations of galaxies, it may become possible to determine whether star formation has occurred during numerous short bursts or in a sustained and constant manner starting at very high redshifts $z \sim 12 - 15$. Using photometric measurements such as Balmer break height and observations in multiple photometric bands; star population age, mass, star formation ratio (SFR) and dust attenuation can be estimated. Deep imaging will be made with NIRCam covering rest frame UV to optical wavelengths, with bands selected according to a z based criteria in order to always measure Balmer break with a pair of NIRCam filters without contamination from nearby emission lines.

SED fitting using photometric data from NIRCam, spectroscopic data from NIRSpec, and photometric medium infrared (MIR) data from MIRI will further characterize different physical properties, but more importantly, will estimate the star formation history (SFH) of the galaxies. Separating contributions from young ($\leq 10\ Myr$) and mature ($\geq 100\ Myr$) star populations (if present) can help to date the beginning of star formation in the universe.

The ultimate goal is to better understand the ionization mechanisms of early galaxies, and to test different hypotheses about how light from the first stars was responsible for the re-ionization of the universe.

Table 1: Sample of ALMA detected, EoR [OIII]88μm emitters that will be observed within #1840 GO1 proposal [Alvarez-Marquez et al., 2021]. Estimations of redshift (z), UV magnitude (M_{UV} [mag]) and of star formation ratio (SFR [$M_\odot$ yr^{-1}]) are available for all galaxies. Unlike stellar mass ($M_*[M_\odot]$), estimated for only some candidates. [J. Álvarez Márquez, Private Communication].

Name	z	M_{UV} [mag]	SFR [$M_\odot$ yr^{-1}]	$M_*[M_\odot]$
J1211-0118	6.03	-22.8	086.0	-
J0235-0532	6.09	-22.8	054.0	-
RXC-J2248-ID3	6.11	-20.1	003.0	6.0E+07
J0217-0208	6.20	-23.3	096.0	-
SDF-LBG-ID34	6.31	-21.8	046.0	4.1E+09
COS-2987030247	6.80	-22.0	022.7	1.7E+09
COS-3018555981	6.85	-21.9	019.2	1.4E+09
BDF-3299	7.11	-20.6	005.7	-
SXDF-NB1006-2	7.21	-21.5	208.0	2.0E+08
MACS0416-Y1	8.31	-21.2	057.0	2.4E+08
A2744-YD4	8.38	-20.8	012.3	1.2E+09
MACS1149-JD1	9.11	-19.2	004.2	1.1E+09
B14-65666	7.15	-22.4	200.0	7.7E+08
A1689-zD1	7.21	-20.0	012.0	1.7E+09

2.2 Detection of galaxies in the EoR

In the Epoch of Reionization (EoR) matter in the universe started to be ionized by photons of new formed stars. By the end of the EoR matter was fully ionized and light could travel freely without being absorbed. It covers the period of time between $5 < z < 15$ or from $400Myr$ to $1Gyr$ after the beginning of the universe.

Theory predicts that first galaxies should start to appear from gravitationally bound objects around $z \sim 12$, that the fraction of ionizing photons that escape from them should be responsible for the ionization of the matter in the universe, and that this process should be completed towards $z \sim 6$ [Ceverino et al., 2019].

Ionization mechanisms that drove physical processes during EoR are not well known to this date. Strong UV emitters are the preferred candidates for explaining reionization [Barkana and Loeb, 2001], but a careful modeling of these sources is needed, and observations of EoR objects need to be made in order to test predictions of models. Physical properties of high z, EoR galaxies will further constraint the models.

The criteria for selecting galaxies of this sample has been the detection of [OIII]88μm emission from ground-based radio telescope ALMA. It has been proven that [OIII] emission can trace star-forming (high redshift) galaxies [Suzuki et al., 2016] in a similar way to Hα emission, with the advantage that Hα is not longer visible from ground for galaxies with $z > 3$.

3 Methodology

In order to prepare JWST observations associated with #1840 GO1 proposal, it's important to: 1) Simulate real data that will be recovered from observations and prepare in advance the reduction process. 2) Test whether the observational strategy is adequate to fulfill all (or some) of the scientific goals detailed in *Section 2.1*.

3.1 FIRSTLIGHT simulation

The lack of real JWST observations from any of it's instruments implies that all of the predictions will be made using simulations albeit with the most recent calibration data from ground. When the observatory is fully commissioned, inflight calibration data will be made public, but no substantial changes are expected in case of nominal performance of the instruments.

Sufficiently good observations of objects from the sample in Table 1 haven't been made to this date with any other instrument in the optical or NIR, because there does not exist any other (ground or spatial) observatory capable of performing deep imaging with high enough spatial resolution and sensitivity to generate NIRCam imaging simulations. This is the main motivation to obtain data from FIRSTLIGHT [Ceverino et al., 2017], [Ceverino et al., 2018], [Ceverino et al., 2019], [Ceverino et al., 2021] cosmological simulation.

FIRSTLIGHT is a high resolution zoom-in cosmological simulation that covers a complete range of halo masses from 10^9 $M_\odot$ to 10^{11} $M_\odot$ within two 10 h^{-1}Mpc and 20 h^{-1}Mpc boxes and with 10 pc resolution. Dark matter particles are 10^4 $M_\odot$ and stellar particles are 10^2 $M_\odot$.

Galaxies from FIRSTLIGHT simulation have been chosen according to a variety of criteria (selection processes are detailed in *Sections 3.2* and *5.1*). They will be the input data from now on, so all results are not for real galaxies in the sample, but for FIRSTLIGHT counterparts. They will be used as reference to generate simulated NIRCam observations, and to determine the way in which star formation history and physical properties can be recovered from NIRCam photometric measurements by performing SED fitting.

Figure 1 shows FIRSTLIGHT sample limited to objects within $5.8 < z < 6.5$ and $6.5 < z < 7.2$, which matches the redshift range of the real targets (superimposed as yellow stars in the figure). M_{UV} [mag] and $M_*[M_\odot]$ vs SFR [$M_\odot$ yr^{-1}] relations are shown, being these the observationally derived data for galaxies in the sample. The process of selecting candidates and transforming data from intrinsic properties obtained from FIRSTLIGHT will be detailed along this work.

3.2 Examination of the sample

The EoR galaxy sample that will be observed within the proposal is shown in Table 1. The sample is mainly composed of $M_* \geq 10^8$ $M_\odot$, star-forming galaxies. SFR values verify $SFR \geq 10$ $M_\odot$/yr for the majority of objects (apart from three exceptions). As can be seen in Figure 1, the range of values of M_{UV} spans from -23.3 to -19.2 mag, covering 4 orders of magnitude. The heterogeneity of objects guarantees that conclusions will not be biased towards one unique type of system (e.g. luminous post-merger galaxies). However, the observed objects are in the luminous/massive end of the EoR galaxy spectrum as can be seen in the Figure 1, because of limitations of ALMA sensitivity and the limited observation time of the proposal.

In Figure 2, relations between [OIII]88 μm/[CII]158 μm and Lyα equivalent width ($EW_{Ly\alpha}$) are shown. The quantity [OIII]88 μm/[CII]158 μm is thought to be related with escape fraction of Lyman photons (ionizing radiation), which is key to understanding the ionizing mechanisms that drove processes during EoR. $EW_{Ly\alpha}$ shows that not all of candidates are Hα emitters.

Figure 1: Comparison plots between galaxies of the sample (marked as yellow stars) and FIRSTLIGHT objects (shown as colored dots). Two objects COS-3018555981 (red star) and RXCJ2248-ID3 (blue star) have been identified together with their FIRSTLIGHT closest matches FL939 (red dot) and FL895 (blue dot) according to four parameters: z, SFR [$M_\odot$ yr^{-1}], M_{UV} [mag] and M_*[$M_\odot$].

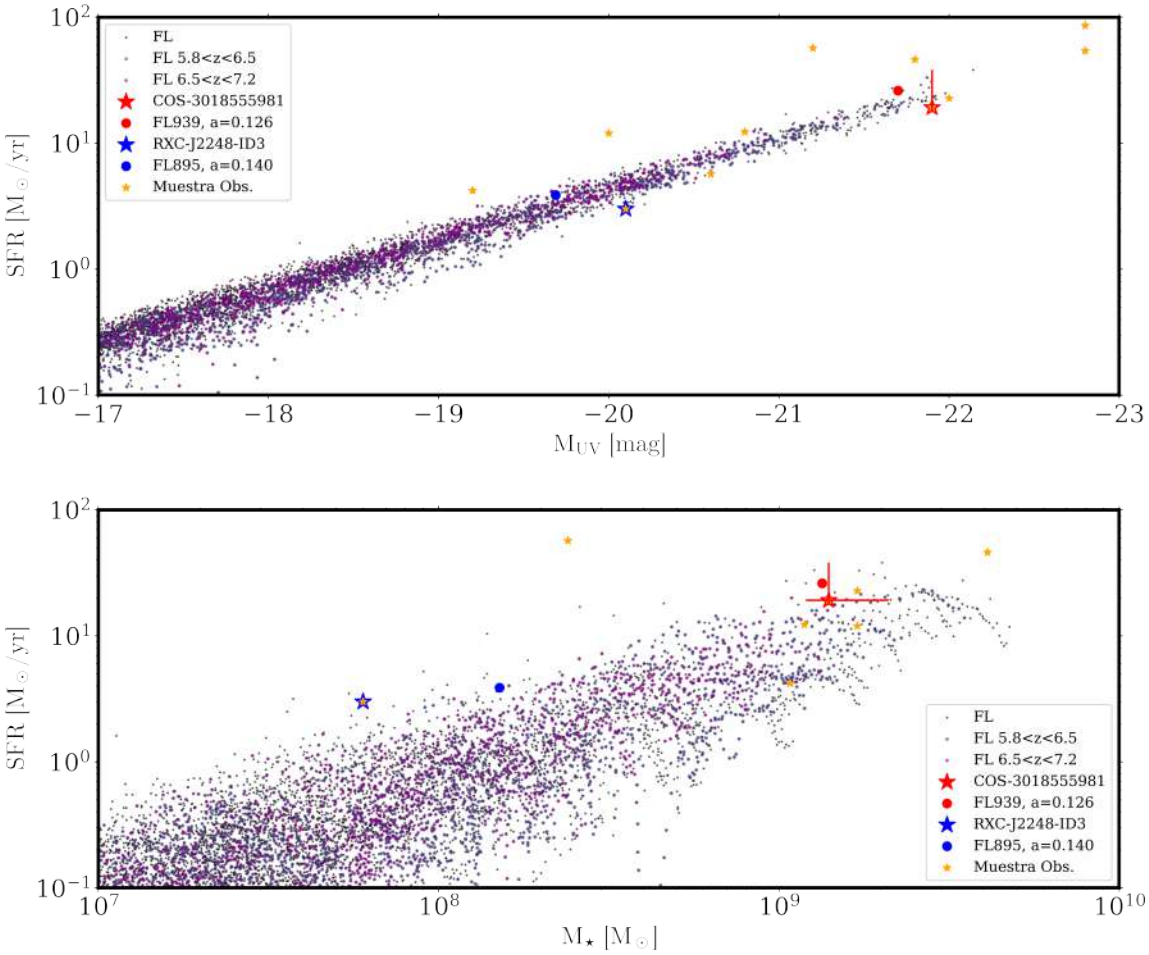


Figure 2: View of the complete sample of [OIII]88μm emitters at $z > 6$. It shows the ratio of [OIII]88μm/[CII]158μm (left) and the $EW_{Ly\alpha}$ (right) as a function of M_{UV}. The colored dots represent sources with no Guaranteed Time Observations or GTO (9, green), with partial GTO (3, orange), and fully covered by GTO programs (2, red). The gray dots represent spectroscopically detected Lyα emitters at $z > 6$. [J. Álvarez Márquez, Private Communication].

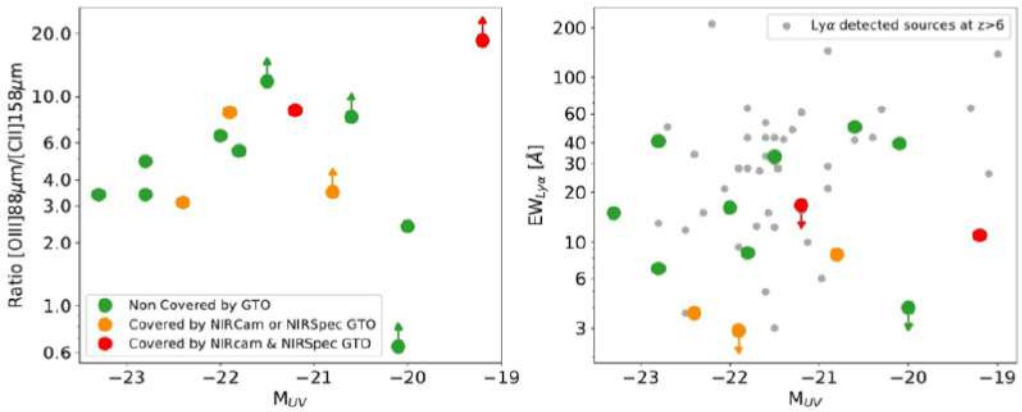
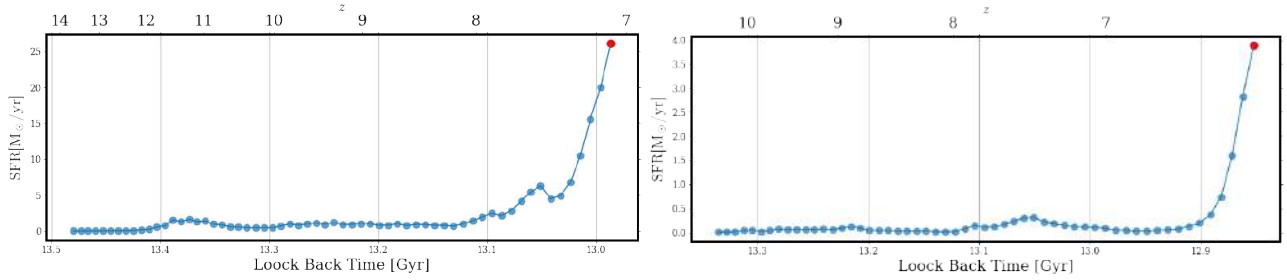


Figure 3: Star formation histories (SFH) of FIRSTLIGHT candidates. Red dot indicates the 'present' snapshot, that being the time when SED has been extracted.



(a) FLF939, candidate of COS-3018555981

(b) FL895, candidate of RXC-J2248-ID3

Two objects of the sample have been selected to perform the analysis and simulations: **COS-3018555981** and **RXC-J2248-ID3**. They are two different prototypes of EoR objects: COS-3018555981 [Smit et al., 2018] is a very luminous and massive system with remarkably high SFR of $19.2 M_{\odot} yr^{-1}$ in the present. RXC-J2248-ID3 [Balestra et al., 2013] is a less massive gravitationally lensed object with a still high but more moderate SFR of $3.0 M_{\odot} yr^{-1}$ in the present.

As shown in Figure 1, the two FIRSTLIGHT candidates with most similar z, SFR [$M_{\odot}$ yr^{-1}], M_{UV} [mag] and $M_*[M_{\odot}]$ (simultaneously) have been chosen to perform all of the simulations. They are thought to share most properties with the real systems. These objects can be seen together with their real counterparts in Table 2. COS-3018555981 and FL939 as red star and dot respectively, RXC-J2248-ID3 and FL895 as a blue star and dot in the same manner.

Figure 3 shows that both FL939 (candidate for COS-3018555981) and FL895 (candidate for RXC-J2248-ID3) show high values of star formation ratio in the present. Clear signs of emission by new formed stars can be seen in Figure 4 such as: negative Balmer break height, strong nebular emission lines and important UV emission. NIRCam observation bands corresponding to the ones found in the proposal [Alvarez-Marquez et al., 2021] are labeled and marked in grey. The observational strategy will be discussed in *Section 3.3*.

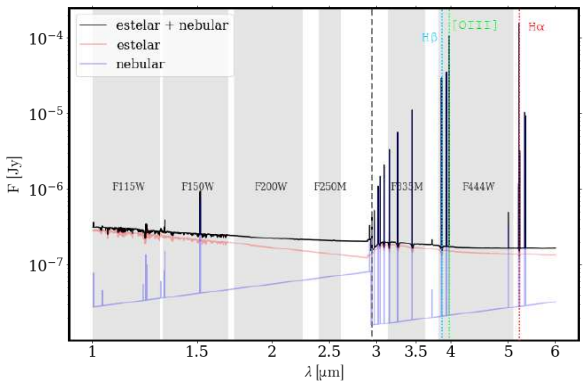
After studying FIRSTLIGHT simulation in order to better characterize candidates FL939 and FL895 it has been found that:

- FL939 is a post-merger system, that is experiencing a very intense ($26.0 M_{\odot}\ yr^{-1}$) star forming episode. It is located in the bright and massive region of the sample.
- FL895 is currently experiencing a merger. It is less massive and bright than FL939, but also experiencing a star formation episode ($3.9 M_{\odot}\ yr^{-1}$) in the present.

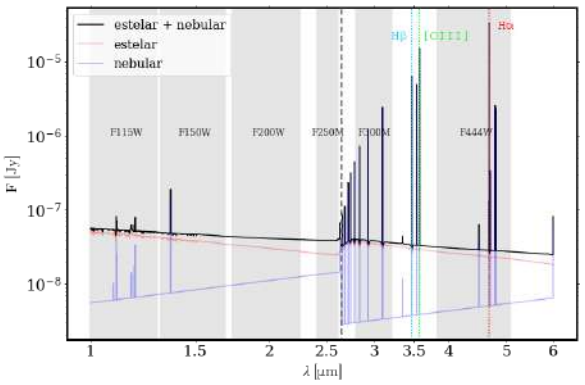
Table 2: Follows the same structure that Table 1, but including only the two objects analyzed in detail: COS-3018555981 and RXC-J2248-ID3 with their closest FIRSTLIGHT counterparts FL939 and FL895, respectively.

Name	z	M_{UV} [mag]	SFR [$M_{\odot}$ yr^{-1}]	$M_*[M_{\odot}]$
RXC-J2248-ID3	6.11	-20.1	03.0	6.0E+07
FL895	6.14	-19.7	03.9	1.5E+08
COS-3018555981	6.85	-21.9	19.2	1.4E+09
FL939	6.94	-22.0	26.0	1.3E+09

Figure 4: View of the spectral energy distributions (SED) of main candidates. Stellar (dark red curve), nebular (dark blue curve) and total SED (black curve) are shown. Three most intense nebular emission lines have been marked with vertical dotted lines: Hα λ6563Å (red), [OIII] λ5007Å (green), Hβ λ4681Å (blue). Position of Balmer break has been also indicated with a vertical discontinuous grey line.



(a) FL939, candidate of COS-3018555981



(b) FL895, candidate of RXC-J2248-ID3

3.3 Observational strategy

In this work, simulations for the main instrument used in the proposal, NIRCam, will be generated. Simulations and calculations use the bands, exposure patterns, pointing locations of #1840 GO1 proposal that will be observed by JWST.

NIRCam is a NIR instrument on JWST that has different working modes, but will only be used in imaging configuration. The observation range in this project is from 1 μm to 5 μm. Coverage of rest frame UV and optical ranges of the spectrum is a key advantage over ground based or spatial optical observations that can only sample UV rest frame for EoR galaxies. NIRCam is composed of two different channels. The short wavelength channel (0.6μm to 2.3μm) has four detectors per unit, with $0.031''/px$ resolution. The long wavelength channel (2.4μm to 5μm) has one detector per unit with $0.063''/px$ resolution. Both observe simultaneously the same sky area of $2.2' \times 2.2'$. Two units (A and B) are available, each containing one set of $(4 + 1)$ $2.2' \times 2.2'$ detectors, but only one is used habitually.

As can be seen in Figure 4, observations will be made in various bands, marked as grey overlays. Selection of bands is based on:

- Rest frame UV: 3 wide filters F115W, F150W and F200W. Important for testing continuum emission from young stellar population. Estimated SNR is 30 in this band.
- Balmer Break: 2 medium filters that depend on z (as can be seen comparing the images) so that bands are located before and after the break. Medium filters are chosen to minimize contamination by nearby nebular lines. Estimated SNR is 15 (which provides an uncertainty of 0.1 mag in Balmer jump height).
- Optical continuum + nebular emission: 1 wide filter F444W, to study mature star emission and H$\alpha\lambda$6565 (RXC-J2248-ID3 / FL895) or [OIII]λ5007(+H$\beta\lambda$4862) (COS-3018555981 / FL939) emission depending on z. This selection effect can be seen comparing both images within Figure 4. Estimated SNR is 30 in this band.

Exposures do vary in their readout pattern and number of integrations. Number of primary dithers is fixed to 4, in order to cover gaps between detectors in the short channel (although targets are located inside one detector) and to reject detector defects. Table 3 shows the configuration for the studied targets, as well as total exposure time.

Table 3: Observation bands and exposure times for objects COS-3018555981 and RXC-J2248-ID3. Note that medium filters do change between targets depending on their redshift in order to correctly characterize the Balmer's break height.

Name	Redshift (z)	Bands	Exp. Time [s]
RXC-J2248-ID3	6.11	SHORT: F115W, F150W, F200W LONG: F250M, **F300M**, F444W	3348
COS-3018555981	6.85	SHORT: F115W, F150W, F200W LONG: F250M, **F335M**, F444W	2220

4 NIRCam imaging

As described above, one of the main objectives of this project is to determine how future NIRCam observations of EoR objects will come out taking into account the observational strategy that has been described in *Section 3.3*. In order to do so, the generation of mock observations mimicking real in flight data is necessary. Reduction of these images using JWST's pipeline is the next step to convert raw images to stacked and calibrated results.

For simplicity, not all exposures neither all objects will be simulated. It has been esteemed sufficient to work with three different bands that cover most of the span of the observed range: UV with F115W, post Balmer break with F330M or F335M, and optical continuum and strong nebular lines with F444W; for the two main objects presented in this project: COS-3018555981 and RXC-J2248-ID3.

FIRSTLIGHT data will be used as input to the simulation package: MIRAGE (Multi Instrument Ramp Generator), the STScI (Space Telescope Science Institute) observation simulator for NIRcam and NIRISS (Fine Guidance Sensor and Near Infrared Imager and Slitless Spectrograph). Data will then be processed with JWST pipeline, run through a background noise removal tool and finally analyzed.

4.1 Making the simulations

The integration of MIRAGE within the STScI ecosystem allows to reduce the workload and make predictions using the proposal's observational strategy: from readout patterns to pointing locations. After exporting proposal's data from the APT (Astronomers Proposal Tool), in which all JWST proposals are created, this data can be imputed directly to MIRAGE.

Input from FIRSTLIGHT needs to be correctly formatted. MIRAGE requires a $.fits$ image with the real pixel scale and adequate relative pixel intensity values in the observed band. It also requires the value of integrated AB magnitude in the desired band, defined as $m_{AB} = -2.5\ log_{10}\left(f_\nu[erg\ s^{-1}cm^{-2}Hz^{-1}]\right) - 48.60 = -2.5\ log_{10}\left(f_\nu[Jy]\right) - 8.90$ [Oke and Gunn, 1983]. Therefore conversion from FIRSTLIGHT intrinsic parameters to observed values in each band requires to perform some manipulation of data.

Magnitudes have not been calculated purely from FIRSTLIGHT. The observational strategy is tuned for the real sample, not FIRSTLIGHT simualtion (which does not completely match the real targets). Since F160W photometry has been measured by Hubble Space Telescope, F115W magnitude has been rescaled making the approximation that SED remains constant in the UV (as can be seen in Figure 4). Other bands have been adjusted taking into account the relative differences to F115W by deriving synthetic photometry from FIRSTLIGHT data.

The data conversion process consists of the following steps:

1. Reading FIRSTLIGHT data formatted as a cube conformed of two spatial dimensions and a third one that contains wavelength information. 2D physical images are given in an array of $61px \times 61px$ with a scale of $100pc/px$.

2. Rebinning the image to the correct pixel size according to NIRCam plate scale and angular size of the object in an universe with cosmological parameters as in Planck18 [Plank Collaboration, 2020]. Special attention has been put into not changing relative pixel values in the process of rebinning.

3. Converting luminosities provided by FIRSTLIGHT to fluxes by dividing by the geometrical factor $4\pi d_{lum}^2$. Luminosity distance d_{lum} is calculated with Planck18 cosmological parameters. Then redshifting wavelengths with the relation $\lambda_{obs} = \lambda_{emit} \cdot (1+z)$. Observed flux then needs to be adjusted so that the integrated monochromatic flux between two redshifted wavelengths λ^{obs} and $\lambda^{obs} + d\lambda^{obs}$ is the same that the integrated monochromatic flux between two rest frame wavelengths λ^{emit} and $\lambda^{emit} + d\lambda^{emit}$. Monochromatic flux is virtually constant in a differential interval, so it follows that $dF = \int_{\lambda}^{\lambda+d\lambda} f_{\hat{\lambda}} d\hat{\lambda} = f_{\lambda} d\lambda$. This implies the following relation between f^{obs} and f^{emit}:

$$\frac{f^{obs}}{f^{emit}} = \frac{d\lambda^{obs}}{d\lambda^{emit}} \rightarrow f^{obs} = \frac{f^{emit}}{1+z} \ .$$

4. Finally, performing synthetic photometry using the correct band of the observation taking into account the filter response curve and converting the flux in $[erg/s/cm^2/A]$ to $[Jy]$ and finally to $[ABmag]$.

The effect of adapting FIRSTLIGHT resolution and pixel count to NIRCam long and short channel resolution can be seen in Figure 5a, where pixel numbers are indicated in order to give a more accurate idea of real dimensions of each image. A schematic view of the process of integrating the flux and determining the value of $ABmag$ (derived from the constant value of flux that produces the same energy as the integration of the actual monochromatic flux) is shown in Figure 5b.

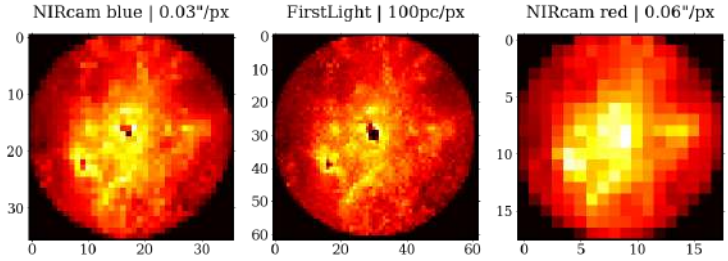
Once data has been transformed from FIRSTLIGHT's format to MIRAGE's format, and the proposal information has been exported from APT; MIRAGE is able to generate simulated exposures that mimic real in flight data. It generates one $.fits$ image (with a number of layers equal to the number of readouts) for each (primary) dither and band. The typical configuration for this proposal is 4 different 9 layer $.fits$ for each band and detector. In the case of short wavelength channel, only the detector that contains the object has been simulated to save computation time.

4.2 Calibrating the data

After creating all simulations a considerable number of uncalibrated images are obtained. This data is presented in a raw format (equivalent to real observations with JWST), that isn't useful for retrieving scientific data, because exposures are not staked and because images are not calibrated and detector defects are present. Figure 6a shows the aspect of a test raw image obtained from MIRAGE (a preliminary test before performing real simulations) that contains field stars and galaxies randomly distributed. Note the presence of vertical bands with different intensity result of the readout process of the HgCdTe NIR detector. After running the pipeline, detector defects are removed and a calibrated image is obtained, as can be seen in Figure 6b.

Figure 5: Relevant parts of the processing of FIRSTLIGHT data in order to input it to MIRAGE.

(a) Comparison between FIRSTLIGHT, NIRCam short/blue channel and NIRCam long/red channel. FL939 at $z = 6.94$ is displayed, with a scale of 5.371 kpc/″.



(b) Calculation of *ABmag* from FIRSTLIGHT datacubes for FL939 and F335M band. Red dots indicate measured values of monochromatic flux (including extinction). The green line indicates average monochromatic flux, which would generate the same energy than the integration of the monochromatic flux over the given interval (in grey). *ABmag* is calculated with the relation $m_{AB} = -2.5 \log_{10} f_\nu + 8.90$. The black line is FIRSTLIGHT reference SED without extinction.

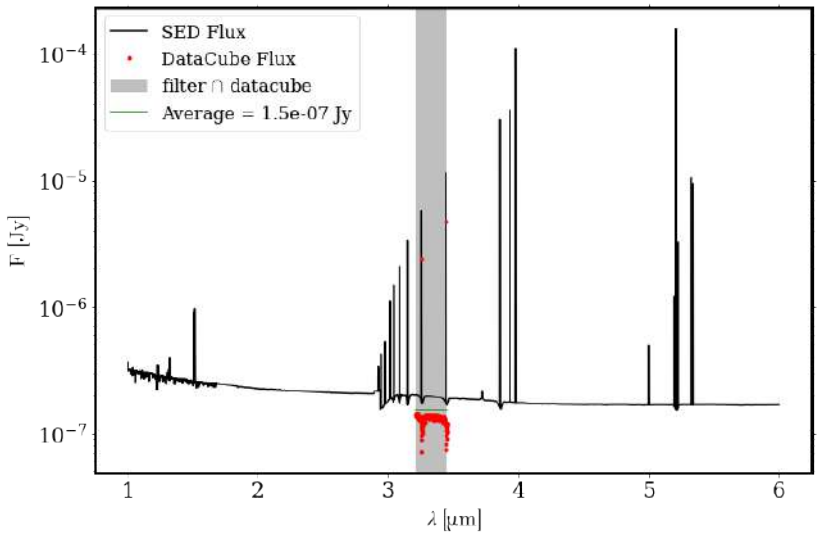
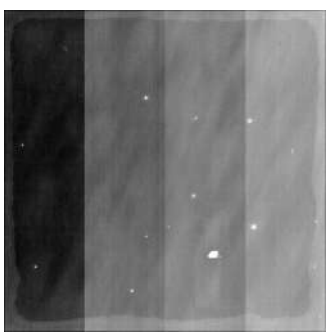


Figure 6: Comparison between raw data with detector defects from MIRAGE and calibrated data after the pipeline.

(a) Raw data out of MIRAGE.



(b) Data after calibration.

The pipeline consists of 3 main stages:

1. Detector Processing: corrects for bias (pedestal level from readout), removes dark current (thermal noise of electrons) and computes jumps and ramps of the non destructive NIR readout.
2. First Imaging Processing: removes background, assigns astrometrical physical coordinates to the image (WCS) and divides by flat field.
3. Second Imaging Processing: combines calibrated data from multiple exposures into a distortion corrected image.

Between steps 2 and 3, that is, before combining exposures together, an aditional noise reduction algorithm has been applied (implemented with IDL and IRAF by a collaborator). It removes large scale background noise patterns from the image. This allows for a better quality and easier background subtraction when analyzing the final images.

4.3 Analysis of the images

Figures 7a and 7b show steps of the NIRCam image simulation process for FL895 and FL939 FIRSTLIGHT objects respectively. In both cases, the structure of the mosaic is similar.

Each column depicts one of the studied bands:

1. F115W band samples rest frame UV, and traces young star population.
2. F335M or F330M samples post Balmer break region, a good estimator of star population age. Selection of band depends on z of the object.
3. F444W samples optical rest frame and strong nebular emission lines such as [OIII]λ5007 for FL939 or H$\alpha\lambda$6565 for FL895. This is an effect of different z values in objects of the sample.

Each row shows the image at different stages of the processing:

1. *firstlight*: image obtained directly from FIRSTLIGHT, integrating into the correct band. The 3D datacube has been collapsed into a 2D image as shown here.
2. *input mirage*: image after rebinning to the correct pixel number according to NIRCam resolution and the angular size of the object. Overlayed contours show distribution of light in the objects.
3. *cal. output*: calibrated product after performing simulations with MIRAGE and running data through the JWST pipeline. Contours from *input mirage* are also included.

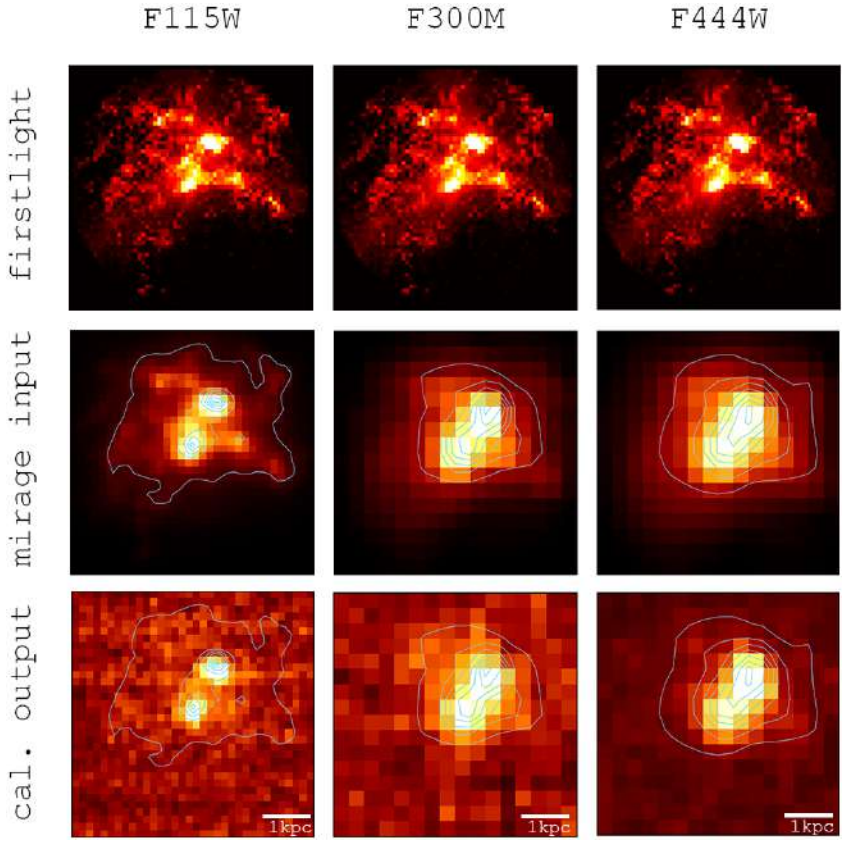
The immediate conclusion is that the observational strategy produces viable images of EoR galaxies that will be studied in the sample. Both FL939 and FL895 calibrated outputs show the structure of sources. Central regions are seen as bright patches. External and fainter regions are also present, even without a careful stretching.

FL895 can be seen as two spatially resolved sources in F115W band, whilst in the other bands effects of point spread function (PSF) and smaller resolution make it more difficult to resolve the two independent sources. With a more careful image stretching better separation could be achieved. Some bright clumps in the outskirts of the galaxy also seen in the original FIRSTLIGHT image are still present.

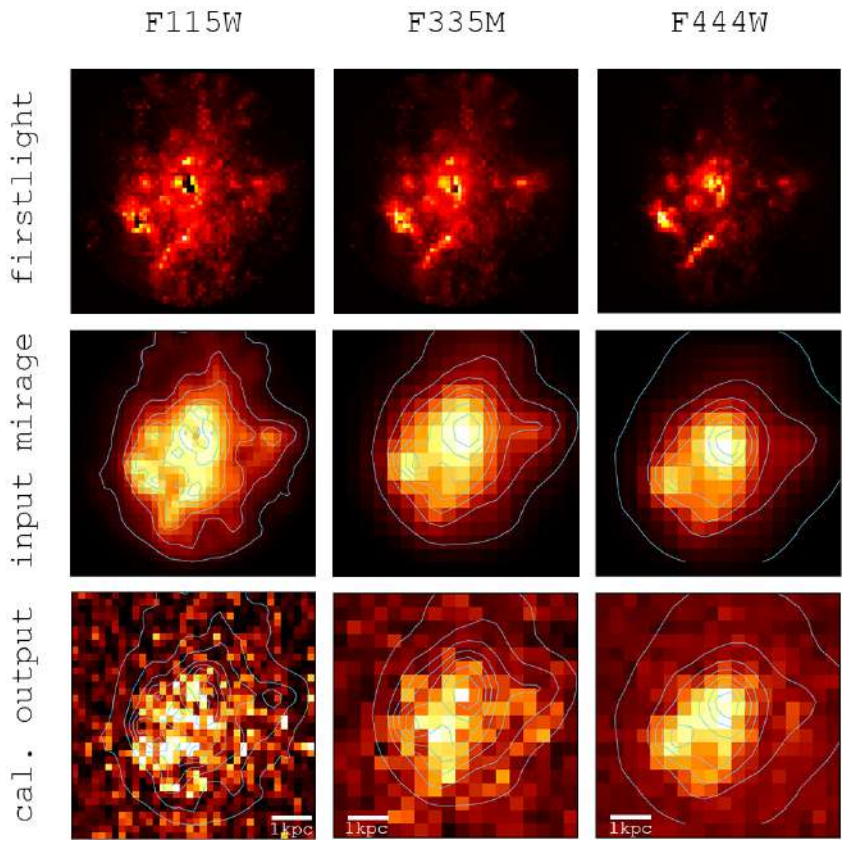
FL939 is a post-merger system that shows different physical structure, also with two bright regions but showing more contribution from external regions. The effect of extinction can be seen in the center of both bright regions mainly in band F115W in data directly extracted from FIRSTLIGHT. The final image in band F115W shows less structure because detection is poorer.

Figure 7: View of the simulated images with MIRAGE. Calibrated data from the pipeline (cal. output), after rebinning (mirage input), and directly from FIRSTLIGHT (firstlight) is shown. Interpretation can be found at *Section 4.3*.

(a) FL895



(b) FL939



Compatibility problems between MIRAGE and the pipeline have been reported. Dark reference files do not perfectly match, and this can introduce artifacts in the images. The chosen solution has been to recover linear data from MIRAGE that is already dark subtracted, but this approach involves building a pipeline with custom steps.

Strategies that might help to obtain a better image from the galaxies do exist. Figures 8a and 8b show the effect of stacking two observations for FL939. A noticeable increase in SNR is achieved, being specially interesting for characterizing the physical structure of galaxies e.g. light distribution, presence of clumps; and to study fainter external regions. The sum of two images provides the following increase of SNR: $SNR_{combined}/SNR_{original} = \sqrt{S_0 + S_1}/\sqrt{S_0}$ where S_0 is the original signal and $S_0 + S_1$ is the added signal of original and new data (for each pixel). An average increase of $SNR_{combined}/SNR_{original} = \sqrt{2}$ is to be expected when the two summed exposures have similar instrumental configuration and exposure times.

Figures 9a and 9b depict the effect of rebinning, in this case reducing the number of pixels per axis by a factor of 2. This increases the signal of each pixel, being specially useful for bands such as F115W where due to a very high resolution signal might be low in individual pixels.

Figure 8: Comparison between original image of FL939 in band F335M (b) and combined data from bands F335M and F444W using a median stack. The increase of SNR can be seen because with equal stretch, contrast between signal and background is enhanced.

(a) F335M + F444W (b) F335M

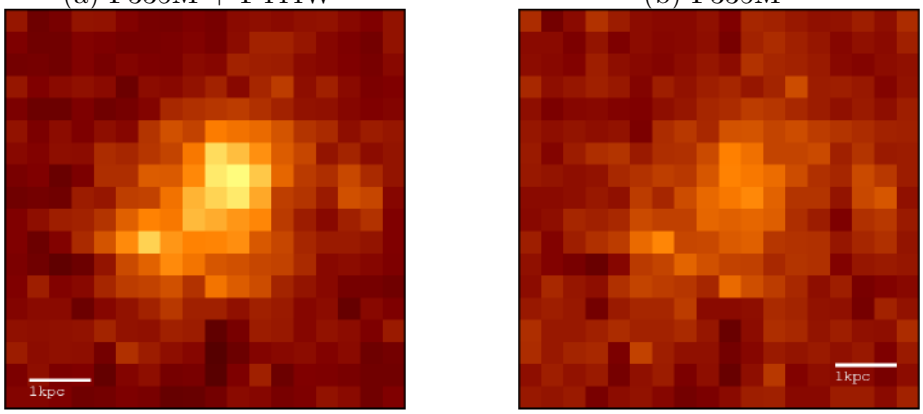
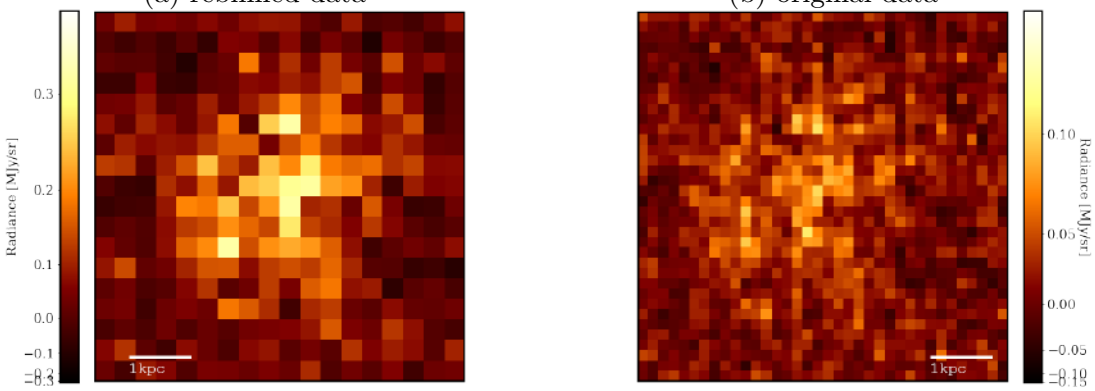


Figure 9: Comparison between original image of FL939 in band F115W (b) and re-binned data with half number of pixels (a).

(a) rebinned data (b) original data



5 SFH AND PHYSICAL PROPERTIES

After having studied the spatial distribution of stellar populations and ISM in the simulated sources via direct visualization of images obtained in *Section 4.3*, estimation of SFH and physical properties is to be done.

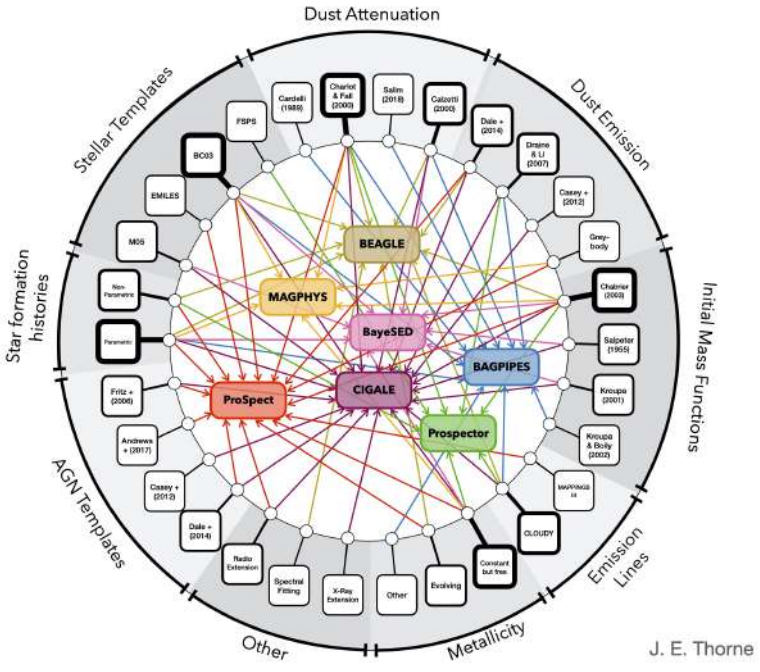
It is of special relevance to determine the SFH of EoR sources with enough certainty to separate contributions from new formed stars and older stellar populations with an age of more than 100 Myr. This is relevant because evidence of presence of old star populations (mainly dust emission) has been found in some objects of the sample. Confirming it could help to narrow down the beginning of the star forming period in the universe.

In order to do so, a careful analysis of NIRCam photometry, NIRSpec spectroscopy of strong emission lines, and MIRI medium infrared photometry will be used altogether. SED fitting has been the chosen technique, as it can make use of photometry + spectroscopy to determine which galaxy model best fits observations, estimating SFH and physical parameters this way.

CIGALE [Boquien et al., 2019] will be the used SED fitting software, as it offers some key advantages over other approaches like combining photometric and spectroscopic measurements, using different initial mass functions (IMFs), or employing bayesian inference. The possibility of including radio, X-ray or AGN extensions to the models is also interesting. Figure 10 shows the most used SED fitting packages and their different capabilities. CIGALE is one the most complete options, lacking non parametrical analysis but offering bayesian methods and extension to radio and X-ray wavelengths.

It first generates a grid of models combining the desired properties of models (as star formation bursts, dust extinctions or ISM metallicities) included in a configuration file. Then it determines the best fitting model out of all of them by minimizing the value of χ^2. The best fitting SED together with its associated SFH and physical properties is then returned. Estimation of the desired parameters using Bayesian inference techniques can also be retrieved easily.

Figure 10: View of the most used SED fitting packages, along with the models they use. Credit for the image: Jessica E. Thorne [Thorne et al., 2021].



5.1 BUILDING THE DATABASE

FIRSTLIGHT has been used as starting point for obtaining data for CIGALE. A large sample has been selected and divided into the following groups:

1. Galaxies with star formation in the present, in which young stellar population ($\leq 10Myr$) clearly dominates SFH and SED.
2. Galaxies with star formation in the past ($>$ 100Myr) with intense stellar formation then, but not in the present.
3. Galaxies with star formation both in the past and present, which is the ideal scenario for testing if the observational strategy will be able to separate old and young components (in systems that actually have them).

Figures 11a, 11b and 11c show prototypical SFHs for each group in which FIRSTLIGHT objects have been classified. Contrast between instantaneous SFR in star forming bursts and quiescent periods has been taken into consideration in order to select viable candidates, because small relative changes might not be detected using SED fitting.

Figures 12a, 12b and 12c show the SEDs of prototypical objects for each group (extracted at the snapshot marked with a red dot in the corresponding SFHs). Features such as Balmer break height, balance between UV and optical regions of the spectrum, and prevalence of emission lines can be associated with differently aged star populations.

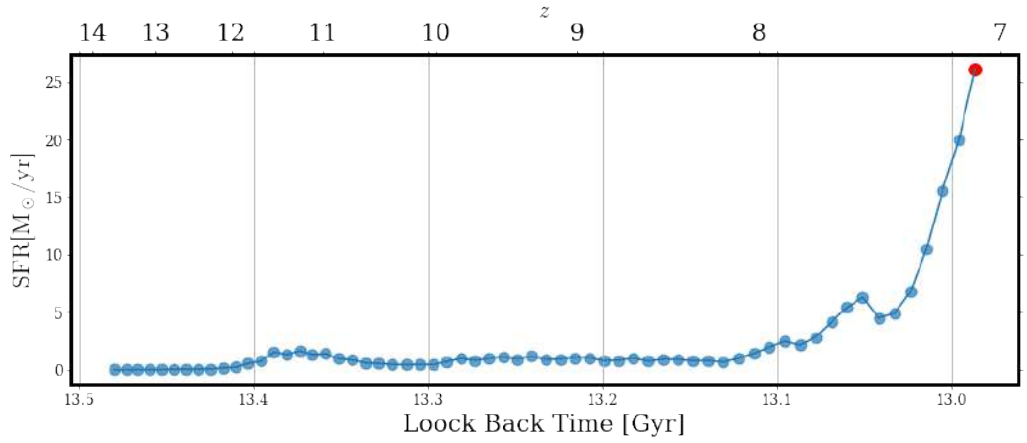
Table 4 shows the complete sample with physical properties for each object and also the group they belong to. A similar process to the one described in *Section 4.1* has been followed. Synthetic photometry has been calculated for each candidate in order to generate a reference file will magnitudes (and errors) with all objects in the desired bands. Random gaussian noise has been added to each measurement to emulate real observations, according to the expected SNR in each band.

Table 4: Most relevant physical parameters for objects in the database for performing SED fitting with CIGALE.

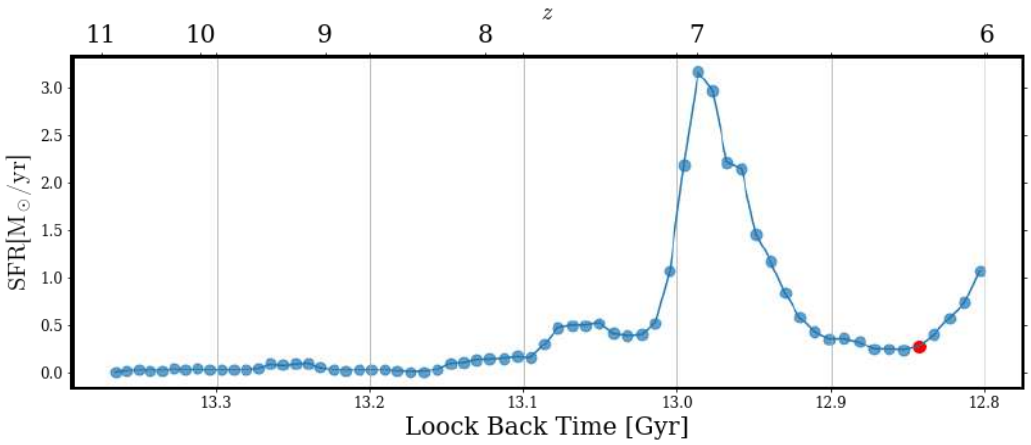
Name	z	M_{UV} [mag]	SFR [$M_\odot$ yr^{-1}]	M_*[$M_\odot$]	group
FL939	6.94	-21.7	26.0	1.34E+09	present
FL895	6.14	-19.7	03.9	1.51E+08	present
FL863	6.94	-19.2	03.2	1.11E+08	present
FL897	6.75	-18.7	01.6	9.09E+07	present
FL861	6.25	-21.0	10.8	6.03E+08	present
FL866	5.94	-19.6	03.1	2.89E+08	present
FL897	6.14	-17.3	00.2	1.31E+08	past
FL863	6.09	-17.8	00.3	2.08E+08	past
FL874	6.41	-16.9	00.1	1.70E+08	past
FL903	6.52	-17.8	00.3	2.55E+08	past
FL773	6.09	-16.4	00.1	7.00E+07	past
FL900	6.87	-19.1	02.1	2.04E+08	past+present
FL932	5.99	-20.0	03.9	7.68E+08	past+present
FL921	5.94	-20.2	04.9	5.07E+08	past+present
FL904	6.09	-18.6	01.1	2.66E+08	past+present
FL829	5.94	-18.9	01.4	2.04E+08	past+present
FL841	6.14	-18.2	00.8	1.16E+08	past+present

Figure 11: Comparison between prototypical SFHs of objects of each group of the sample. Blue dots represent FIRSTLIGHT snapshots, and the red dot represents the selected snapshot to extract data from FIRSTLIGHT.

(a) Star formation in the present: FL939



(b) Star formation in the past: FL863



(c) Star formation both in the past and present: FL932

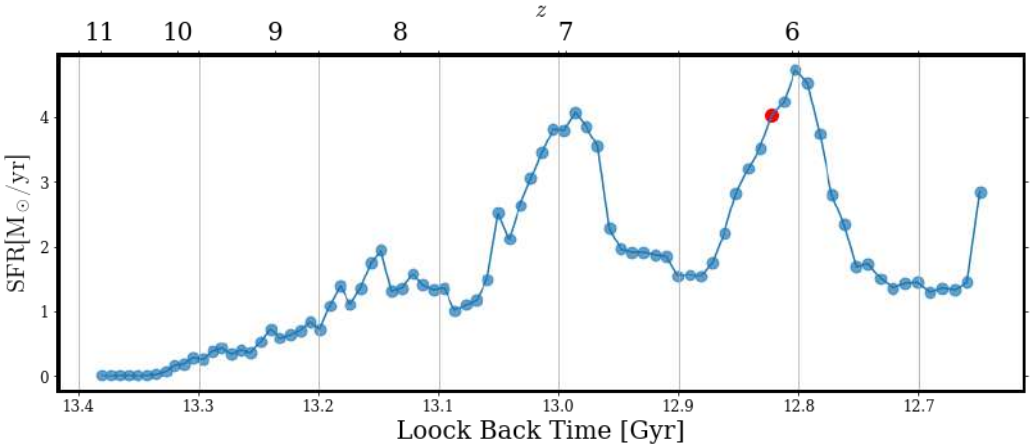
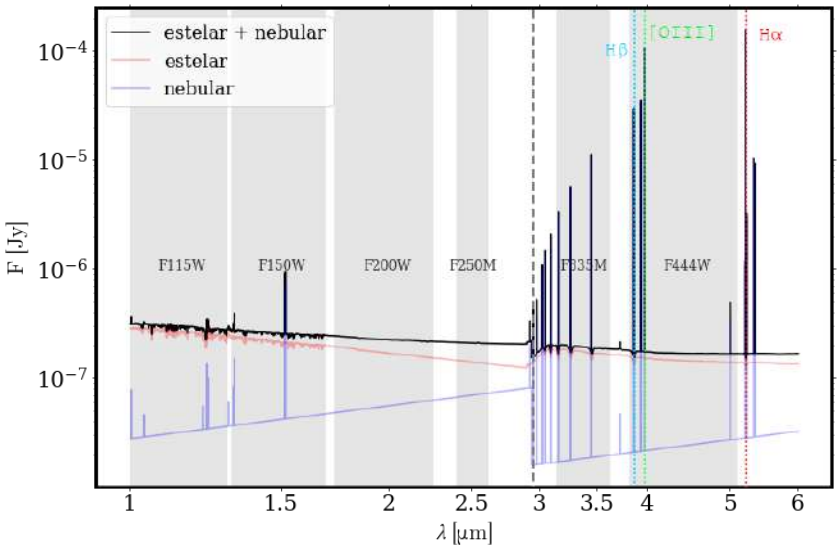
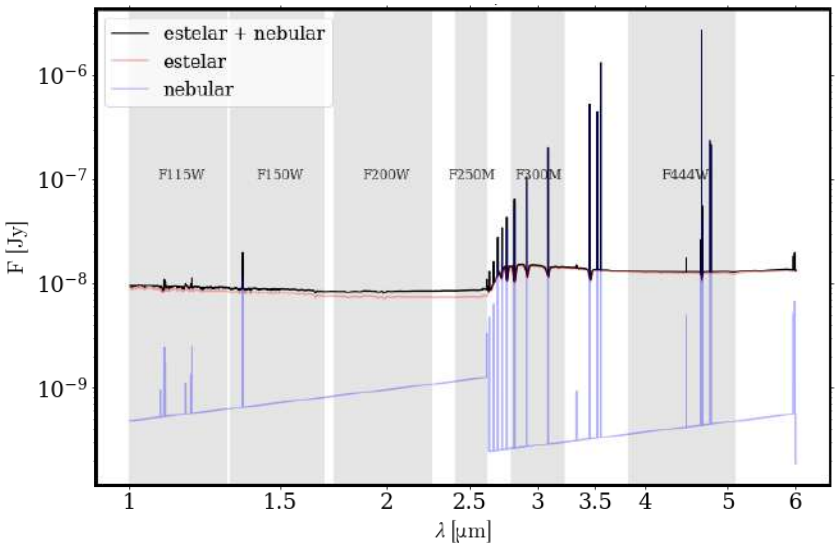


Figure 12: Comparison between prototypical SEDs of objects of each group of the sample.

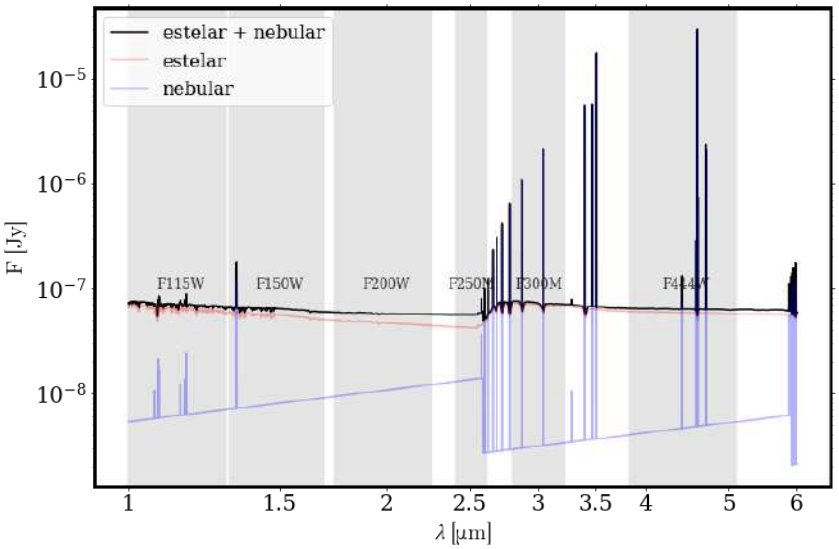
(a) Star formation in the present: FL939



(b) Star formation in the past: FL863



(c) Star formation both in the past and present: FL932



5.2 SED fitting Analysis

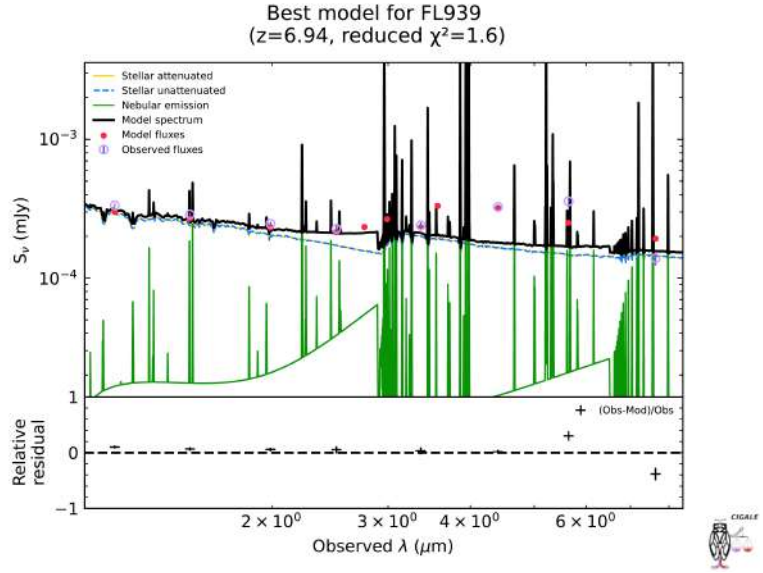
A set of initial SED fitting runs for source FL939 have been done in order to test the correct operation of CIGALE and to ensure that the created database is fully operative and yields consistent results out of SED fitting.

Figures 13a and 13b show SED fitting results for FL939. Comparison with real data (SED in Figure 12a and SFH in Figure 11a) yields positive detection for a recent star forming burst of an approximate duration of 25 Myr.

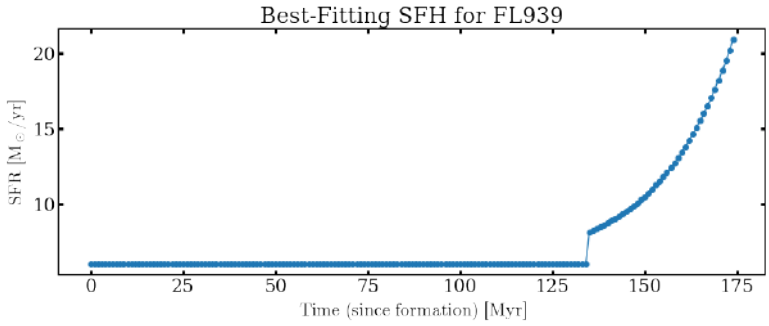
Bayesian analysis of fitted parameters gives the following results: $M_* = (1.0 \pm 0.3) \times 10^9\ M_\odot$, $SFR(10Myr) = 26 \pm 6\ M_\odot/yr$, which are close to the real FL939 parameters described in Table 4 (SED fitting predictions are inside the uncertainty range in both cases). This ensures that estimation of physical parameters such as stellar mass, SFR at different timescales, SFH, and others such as metallicity, gaseous mass, UV slope, can be retrieved for objects in the sample (at least in the ones similar to FL939).

Figure 13: Main results of SED fitting with CIGALE for object FL939.

(a) SED of the best fitting model. Almost all computed photometry values (red) are in good concordance with *observed* (FIRSTLIGHT derived) results marked in purple. The last two values corresponding to MIRI data are not well fitted, this could be an indicator for a deviation in older stellar population.



(b) SFH corresponding to the best fitting model. SFH has been modelled as a double exponential with a constant component and a rising burst in the last few Myr.



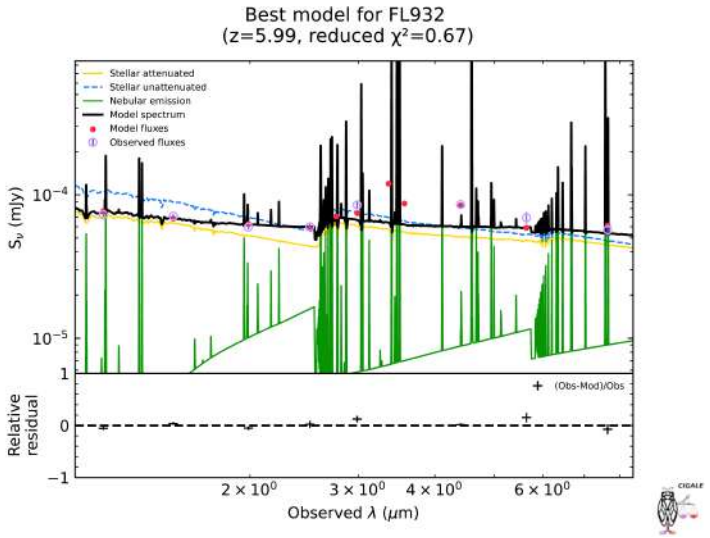
It is also relevant to study whether SED fitting will be able to untangle complex SFHs such as the ones in objects with two star forming bursts, one in the past and one in the present. Such is the case of FL932, with properties detailed in Table 4 and Figures 11c (SFH) and 12c (SED). Another set of parameters has been prepared for the (double exponential) SFH model: a past decaying burst plus a present rising burst. Results can be seen in Figures 14a and 14b.

All of the generated models should show that there is no dust extinction ($E(B-V) = 0$ mag). The followed approach has been to build a series of catalogs with different extinction values $E(B-V) \in (0.00, 0.01, 0.10, 0.20, 0.30, 0.40)$, and to study them one by one to determine in which way extinction affects SED fitting results. This analysis is to be done, for the time being only the zero extinction catalog has been tested, as mentioned in *Section 6*.

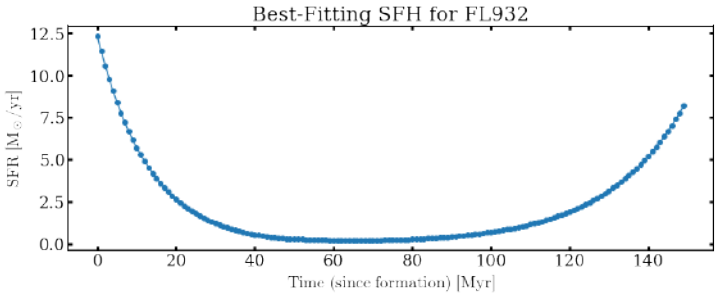
The parameters obtained for SED fitting of FL932 are $M_* = (2.6 \pm 0.1) \cdot 10^8\ M_\odot$, $SFR(10Myr) = 7 \pm 1\ M_\odot/yr$, $SFR(100Myr) = 1.7 \pm 0.3\ M_\odot/yr$. Stellar mass doesn't match as well as in the case of FL939, because reconstruction of complex SFHs requires a more advanced analysis that hasn't been considered. However, this results show that two different aged components might be distinguished in galaxies that host older and younger stellar populations, and with a more careful work, correct estimations of physical parameters can probably be obtained with observations of the proposal.

Figure 14: Main results of SED fitting with CIGALE for object FL932.

(a) SED of the best fitting model. It shows good agreement for almost more bands. Here CIGALE has applied extinction of $E(B-V) = 0.1$ mag, which does not agree with the $E(B-V) = 0.0$ mag (no extinction) of the catalog.



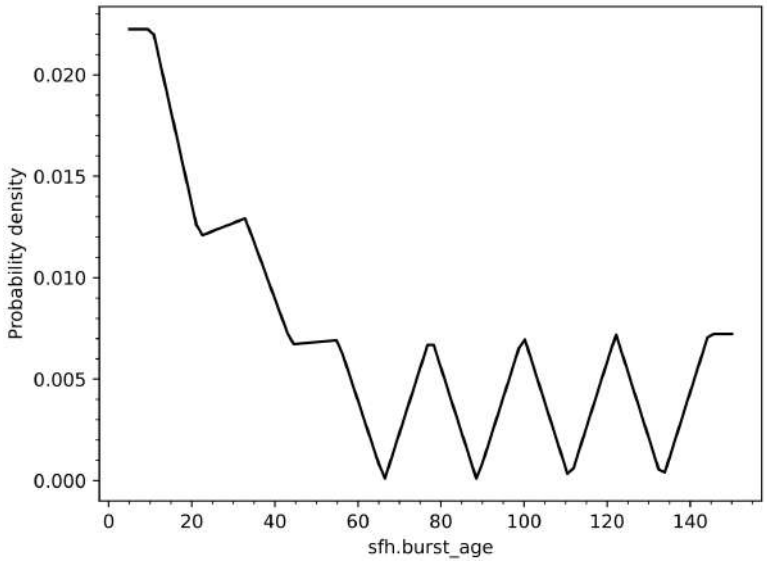
(b) SFH corresponding to the best fitting model. SFH has been modelled as a double exponential with a decaying past component and a rising present burst in the last few Myr.



The main advantage of CIGALE is the possibility of including Bayesian inference in the SED fitting process. This way, not only the best fitting values for parameters can be obtained. As can be seen in Figure 15, the posterior probability density function (PDF) has been obtained for the age of recent star formation burst of FL939. Values smaller than 25 Myr are the most probable ones, as it was predicted by the classical χ^2 minimization method shown in Figure 13b. These estimations are available for all the parameters, and allow a better interpretation of results, which can sometimes be difficult when working on faint EoR objects.

An overview of the model that has been used for performing SED fitting with CIGALE will be detailed. The chosen stellar template is BC03 [Bruzual and Charlot, 2003] with a initial mass function as defined by Chabrier [Chabrier, 2003]. Calzetti relations for dust attenuation [Calzetti et al., 2000] have been used to both generate the catalogs and run CIGALE. Neither AGN, radio or X-ray extensions have been used to perform the modeling.

Figure 15: View of the posterior probability density function (PDF) for the age of the late stellar burst in FL939.



6 Future Work

A more detailed analysis could have been done specially in the determination of SFH and physical parameters using SED fitting. Some pending tasks are:

1. Doing a better modeling of two star populations systems, varying SFH shape and trying to obtain a better quality fit.
2. Testing all the extinction catalogs and not only the one with $E(B-V)=0$.
3. Performing a blind fit of the entire sample and trying to determine which objects have two populations without any beforehand information.
4. Making NIRCam simulations with MIRAGE for all bands in order to test that the SNR is in agreement with predictions.

Another aspect to take into consideration would be to use FIRSTLIGHT MASSIVE catalog, a new cosmological simulation that host more massive galaxies than FIRSTLIGHT, similar to the ones seen in the observed sample. All the analysis routines here developed would be easily adaptable to other FIRSTLIGHT candidates.

7 Conclusions

Images obtained in *Section 4.3* show that Objective 1 (determining stellar and gaseous structure of the host galaxies) will be fulfilled with the proposed observations. The observational strategy has enough sensitivity and spatial resolution to differentiate structure of stellar and nebular components in galaxies of the sample.

As shown in *Section 5.2* SED fitting will be useful to estimate SFHs and physical properties, probably being able to detect as well the presence of old and young stellar populations in galaxies. This will broaden our knowledge about EoR, and get us closer to determining the mechanisms of ionization in the early universe.

To sum up, the proposed observations will provide unseen details of EoR objects. They will allow us to estimate unknown properties such as star formation history and also to test already derived parameters like stellar mass. This will be possible by simultaneously observing rest frame UV, optical and NIR ranges of the spectrum, which can be done using IR telescopes such as JWST. New possibilities that arise with last generation observatories will be game changing for modern high redshift extragalactic astronomy.

References

[Alvarez-Marquez et al., 2021] Alvarez-Marquez, J., Hashimoto, T., Arribas, S., Bakx, T., Ceverino, D., Colina Robledo, L., Inoue, A., Marques-Chaves, R., Matsuo, H., Mawatari, K., Pereira Santaella, M., Tamura, Y., and Yoshida, N. (2021). ALMA [OIII]88um Emitters. Signpost of Early Stellar Buildup and Reionization in the Universe. page 1840.

[Balestra et al., 2013] Balestra, I., Vanzella, E., Rosati, P., Monna, A., Grillo, C., Nonino, M., Mercurio, A., Biviano, A., Bradley, L., Coe, D., Fritz, A., Postman, M., Seitz, S., Scodeggio, M., Tozzi, P., Zheng, W., Ziegler, B., Zitrin, A., Annunziatella, M., Bartelmann, M., Benitez, N., Broadhurst, T., Bouwens, R., Czoske, O., Donahue, M., Ford, H., Girardi, M., Infante, L., Jouvel, S., Kelson, D., Koekemoer, A., Kuchner, U., Lemze, D., Lombardi, M., Maier, C., Medezinski, E., Melchior, P., Meneghetti, M., Merten, J., Molino, A., Moustakas, L., Presotto, V., Smit, R., and Umetsu, K. (2013). CLASH-VLT: spectroscopic confirmation of a z = 6.11 quintuply lensed galaxy in the Frontier Fields cluster RXC J2248.7-4431. *A&A*, 559:L9.

[Barkana and Loeb, 2001] Barkana, R. and Loeb, A. (2001). In the beginning: the first sources of light and the reionization of the universe. *PHYSREP*, 349(2):125–238.

[Boquien et al., 2019] Boquien, M., Burgarella, D., Roehlly, Y., Buat, V., Ciesla, L., Corre, D., Inoue, A. K., and Salas, H. (2019). CIGALE: a python Code Investigating GALaxy Emission. *Astronomy & Astrophysics*, 622:A103.

[Bruzual and Charlot, 2003] Bruzual, G. and Charlot, S. (2003). Stellar population synthesis at the resolution of 2003. *MNRAS*, 344(4):1000–1028.

[Calzetti et al., 2000] Calzetti, D., Armus, L., Bohlin, R. C., Kinney, A. L., Koornneef, J., and Storchi-Bergmann, T. (2000). The Dust Content and Opacity of Actively Star-forming Galaxies. *ApJ*, 533(2):682–695.

[Ceverino et al., 2017] Ceverino, D., Glover, S. C. O., and Klessen, R. S. (2017). Introducing the FirstLight project: UV luminosity function and scaling relations of primeval galaxies. *MNRAS*, 470(3):2791–2798.

[Ceverino et al., 2021] Ceverino, D., Hirschmann, M., Klessen, R. S., Glover, S. C. O., Charlot, S., and Feltre, A. (2021). FirstLight IV: diversity in sub-L_* galaxies at cosmic dawn. *MNRAS*, 504(3):4472–4480.

[Ceverino et al., 2018] Ceverino, D., Klessen, R. S., and Glover, S. C. O. (2018). FirstLight - II. Star formation rates of primeval galaxies from z=5-15. *MNRAS*, 480(4):4842–4850.

[Ceverino et al., 2019] Ceverino, D., Klessen, R. S., and Glover, S. C. O. (2019). FirstLight III: rest-frame UV-optical spectral energy distributions of simulated galaxies at cosmic dawn. *MNRAS*, 484(1):1366–1377.

[Chabrier, 2003] Chabrier, G. (2003). Galactic Stellar and Substellar Initial Mass Function. *PASP*, 115(809):763–795.

[Hilbert et al., 2019] Hilbert, B., Sahlmann, J., Volk, K., Osborne, S., Dthatte, Perrin, M., Chambers, L., Slavich, E., Taylor, J., Tollerud, E., and Lim, P. L. (2019). spacetelescope/mirage: First github release.

[Oke and Gunn, 1983] Oke, J. B. and Gunn, J. E. (1983). Secondary standard stars for absolute spectrophotometry. *ApJ*, 266:713–717.

[Plank Collaboration, 2020] Plank Collaboration (2020). Planck 2018 results. *Astronomy & Astrophysics*, 641:A6.

[Smit et al., 2018] Smit, R., Bouwens, R. J., Carniani, S., Oesch, P. A., Labbé, I., Illingworth, G. D., van der Werf, P., Bradley, L. D., Gonzalez, V., Hodge, J. A., Holwerda, B. W., Maiolino, R., and Zheng, W. (2018). Rotation in [C II]-emitting gas in two galaxies at a redshift of 6.8. *Nature*, 553(7687):178–181.

[Suzuki et al., 2016] Suzuki, T. L., Kodama, T., Sobral, D., Khostovan, A. A., Hayashi, M., Shimakawa, R., Koyama, Y., Tadaki, K. i., Tanaka, I., Minowa, Y., Yamamoto, M., Smail, I., and Best, P. N. (2016). [O III] emission line as a tracer of star-forming galaxies at high redshifts: comparison between Hα and [O III] emitters at z=2.23 in HiZELS. *MNRAS*, 462(1):181–189.

[Thorne et al., 2021] Thorne, J. E., Robotham, A. S. G., Davies, L. J. M., Bellstedt, S., Driver, S. P., Bravo, M., Bremer, M. N., Holwerda, B. W., Hopkins, A. M., Lagos, C. d. P., Phillipps, S., Siudek, M., Taylor, E. N., and Wright, A. H. (2021). Deep Extragalactic VIsible Legacy Survey (DEVILS): SED fitting in the D10-COSMOS field and the evolution of the stellar mass function and SFR-$M_\star$ relation. *MNRAS*, 505(1):540–567.