ALPINE: the ALMA Large Program to INvestigate [CII] at Early times

Data delivery document

ALMA Large Program PI: Olivier Le Fèvre (LAM/Marseille)
ALMA Large Program ID: 2017.1.00428.L

ALPINE team at the DR1:
Ricardo Amorin, Sandro Bardelli, Matthieu Béthermin, Médéric Boquien, Peter Capak, Rosamaria Carraro, Paolo Cassata, Andrea Cimatti, Miroslava Dessauges-Zavadsky, James Dunlop, Andreas Faisst, Yoshi Fudamoto, Seiji Fujimoto, Mauro Giavalisco, Michele Ginolfi, Carlotta Gruppioni, Nimish Hathi, Shoubaneh Hemmati, Thomas Hughes, Edo Ibar, Gareth Jones, Yanu Khusanova, Anton Koekemoer, Guilaine Lagache, Federica Loiacono, Olivier Le Fèvre, Brian Lemaux, Roberto Maiolino, Daniel Masters, Hugo Méndez-Hernández, Tohru Nagao, Desika Narayanan, Pascal Oesch, Riccardo Pavesi, Janine Pforr, Francesca Pozzi, Dominik Riechers, Michael Romano, Wiphu Rujopakarn, Daniel Schaerer, John Silverman, Margherita Talia, Lidia Tasca, Romain Thomas, Sune Toft, Laurence Tresse, Livia Vallini, Daniela Vergani, Fabian Walter, Wei-Hao Wang, Lin Yan, Giovanni Zamorani, Elena Zucca

Contact about the ALMA data: Matthieu Béthermin (matthieu.bethermin@lam.fr or matthieu.bethermin@astro.unistra.fr)

Contact about the ancillary data: Andreas Faisst (afaisst@ipac.caltech.edu)

Reference papers: Le Fèvre et al. (2020, A&A, 643A, 1L, survey design), Béthermin et al. (2020, A&A, 643A, 2B, ALMA data), and Faisst et al. (2020, ApJS, 247, 61F, ancillary data). Please cite these three papers as reference for the ALPINE survey.

Note: some parts of this document were adapted from Béthermin et al. (2020, A&A, 643A, 2B) and Béthermin et al. (2020, ESO Messenger, 180, 31B).

Introduction

The aims of the ALPINE survey

Before the inception of the Atacama Large Millimeter/submillimeter Array (ALMA), the main constraints on the evolution of galaxies in the adolescent Universe ($4 < z < 6$) came from rest-frame ultraviolet (UV) light that was redshifted to optical and near-infrared wavelengths. Thanks to its unprecedented sensitivity in the infrared, ALMA has opened a new window through which to explore the cold and dusty Universe at these early times. Pioneering studies demonstrated that ALMA can detect both the dust continuum and the far-infrared (FIR) fine-structure line [CII] (see, for example, Capak et al., 2015). Continuum and line emission can be targeted simultaneously with ALMA and are both valuable tracers of dust-obscured star formation. ALPINE (the ALMA Large Program to INvestigate [CII] at Early times; Le Fèvre et al., 2020; Faisst et al., 2020; Béthermin et al., 2020) builds the first comprehensive, statistically representative, multi-wavelength sample of normal (main-sequence) star-forming galaxies at the end of the epoch of reionisation, with observations of all galaxies in the sample from the rest-frame UV to the far-infrared. In total, 118 spectroscopically selected galaxies at red- shifts of $4.4<z<5.9$ were observed as part of the ALPINE survey.

The main goals of this survey are:
– To test the [CII] line as a star formation tracer at high redshift, since some theoretical models predict a deficit in [CII] per unit star formation in the low-metallicity galaxies that dominate at high red- shift as compared to the local SFR-[CII] relation.
– To use both the dust continuum and [CII] to estimate the contribution of obscured star formation in known spectroscopic sources at $4<z<6$, and to use this estimate to understand how this contribution modulates the total star formation rate density at these epochs.
– To estimate the precise relationship between the stellar mass of a galaxy and its star-formation activity at $4 < z < 6$ by combining UV (emission tracing exposed young stars) and infrared (emission tracing young stars hidden by dust) data to estimate the total star formation rate (SFR).
– To understand the basic interstellar medium (ISM) properties of these systems from their dust and [CII] luminosities.
– To measure the dynamical masses of these systems using [CII] to constrain their gas fractions.
– To measure the merger rate using the velocity and position information of [CII].
– To identify and quantify possible gas outflows from the [CII] line profiles.

Sample selection and ancillary data

The ALPINE sample is primarily rest-frame UV-limited at $M1500<–20.2$, a limit which corresponds to galaxies about 2.5 times fainter than those typical at this epoch (i.e., $L_{\odot,UV}$). This limit naturally leads to a sample selected to an SFR limit of $\gtrsim 10\ M_{\odot}\ yr^{-1}$. We found that this cut maximizes the sample size while simultaneously minimizing the amount of observing time with ALMA. To set the expectations of the ALMA observations, the [CII] emission fluxes were conservatively predicted using the relation between the observed UV luminosity and [CII] line emission based on a pilot sample by Capak et al. (2015). Because of ALMA's narrow frequency bands, the redshifts of all targeted galaxies were already determined precisely by one of several large spectroscopic surveys on the targeted legacy fields (COSMOS and ECDFS). Most of the spectroscopic redshifts were measured from observations taken as part of the VUDS survey (Le Fèvre et al., 2014) and the Keck/DEIMOS 10k survey (Hasinger et al., 2018). To mitigate potential biases associated with spectroscopic selections, the target sample was selected by several different methods, for example, via colors (Lyman-break dropout technique), narrow bands (Lya emission selection), photometric redshifts, and serendipitous detections. Furthermore, the spectroscopic redshifts are derived from UV absorption lines as well as the Lya emission feature. In total, 13 target galaxies are located in the ECDFS field (from the VISTA Deep Extragalactic Observations survey, VIDEO) and 105 in the COSMOS field from Cosmic Evolution Survey. As shown in Faisst et al. (2020), the ALPINE galaxies represent the average population of galaxies at these redshifts well in terms of the mass of their stellar content and SFRs. As such, ALPINE observations enable, for the first time, the study of the panchromatic properties of the average galaxy at these cosmic times with a high degree of statistical certainty.

The ALPINE team combined ALMA observations in the far-infrared with a wealth of exquisite ancillary imaging and spectroscopy products at rest-frame UV and optical wavelengths, all of which constitutes the first large multi-wavelength survey of galaxies at these redshifts. All galaxies have deep photometry from UV to near-infrared (NIR) from ground-based telescopes, the Hubble Space Telescope, and the Spitzer Space Telescope. These data allow accurate constraints of galaxies' physical properties (stellar mass, SFR, age) from high-quality spectral energy distribution (SED) fitting. In addition, the Spitzer [3.6 μm]–[4.5 μm] colors provide estimates of their Ha emission strength, hence additional constraints on their rates of star formation. The deep rest-frame UV spectroscopy available for all galaxies provides valuable insights into the metallicity and stellar wind properties via UV absorption lines and the Lya emission line. Furthermore, the UV continuum slope is a good measure of dust opacity along the line of sight and can, together with the ratio of far-infrared to UV luminosity, constrain the dust properties of these galaxies. The ancillary imaging and spectroscopic data, as well as various measurements of physical parameters, are detailed in Faisst et al. (2020).

ALMA observations

The ALPINE-ALMA large program targeted 122 individual $4.4<z_{spec}<5.9$ and SFR>10 $M_{\odot}$/yr galaxies with known spectroscopic redshifts from optical ground-based observations. In this redshift range, the [CII] line falls in the band 7 of ALMA (275-373 GHz). To avoid an atmospheric absorption feature, no source has been included between z=4.6 and 5.1. To minimize the calibration overheads, we created many groups of two sources with similar redshift, which are observed using the same spectral setting. In our sample, the typical optical line width is sigma~100 km/s (or FWHM~235 km/s). At the targeted frequency, the coarse resolution (31.250 MHz) offered by the Time Division Mode (TDM) is sufficient to resolve our lines (25-35 km/s) and results in a total size of our raw data below 3 TB for the whole sample. The [CII] lines of the targeted sources are covered by two contiguous spectral windows (1.875 GHz each), while we placed two remaining spectral windows in the other side band to optimize the bandwidth and thus the continuum sensitivity. To maximize the integrated flux sensitivity, we requested compact array configurations (C43-1 or C43-2) corresponding to a >0.7 arcsec resolution to avoid diluting the flux of our sources into several synthesized beams.

We aimed for a 1-sigma sensitivity on the integrated [CII] luminosity $L_{[CII]}$ of 0.4×10^8 $L_{\odot}$ assuming a line width of 235 km/s. At higher redshift (lower frequency), we need to reach a lower noise in Jy/beam to obtain the same luminosity (~0.2 mJy/beam in

235 km/s band at z=5.8 versus ~0.3 mJy/beam in the same band at z=4.4). In contrast, at low frequency, the noise is lower because of the higher atmospheric transmission and the lower receiver temperature. The two effects compensate each other, and the integration times are similar for our entire redshift range (15-25 min on source). Each scheduling block containing the observations of the calibrators and two sources can be observed using a single 50 min-1h15min execution. In total, we had 61 scheduling blocks (SBs) for a total of 69.3 h including overheads.

ALPINE was selected in cycle 5 and most of the observations were completed during this period. Between 2018/05/08 and 2018/07/16, 102 of our sources were observed. Observations had to be stopped from mid-July to mid-August because of exceptional snowstorms. Two additional sources were observed after the snowstorms (2018/08/20). After that, the configuration was too extended and the 18 last sources were carried over in cycle 6. They were observed between 2019/01/09 and 2019/01/11.

We realized during the data analysis that four ALPINE sources were observed two times with different names: vuds_cosmos_5100822662 and DEIMOS_COSMOS_514583, vuds_cosmos_5101288969 and DEIMOS_COSMOS_679410, vuds_cosmos_510786441 and DEIMOS_COSMOS_455022, vuds_efdcs_530029038 and CANDELS_GOODSS_15. For simplicity, we use hereafter only the VIMOS ultra deep survey (VUDS) name of these objects. We combined the two ALPINE observations of each of these sources to obtain deeper cubes and maps.

Content of the data delivery

This ALPINE data delivery contains the following products:

- continuum maps after excluding the channels contaminated by the [CII] line from the ALPINE targets (*continuum* in the file name). The serendipitous continuum detections in the field may be contaminated by a line (see Sect. 3.4 of Béthermin et al. 2020). Line-free continuum images of these sources are available on the team website (https://cesam.lam.fr/a2c2s/data_release.php).
- Data cubes including the continuum (*cube* in the file name).
- Moment-0 maps of the [CII] line (*CIImom0* in the file name).

For each of these products, we provide:

- the signal map (*.image.fits*),
- the primary beam map (*.flux.fits*),
- the synthesized beam map (*.psf.fits*).

The signal maps were NOT corrected for the primary beam gain. Our motivation was to have a homogeneous noise across the field for source extraction. In the ALPINE catalog, the source fluxes are corrected a posteriori by the primary beam correction at their position.

The full name structure of the files is
member.[uid].lp_olefevre.[source name].[product].[type].fits, where:

- [uid] is the member OU ID (e.g. uid___A001_X1284_X1872)
- [source name] is the source name (e.g. DEIMOS_COSMOS_400160)
- [product] is the type of product (continuum, CIImom0, or cube).
- [type] is the type of data (image for the signal map, flux for the primary beam, and psf for the synthesized beam).

The catalog of the entire survey is available in Béthermin et al. (2020) and at the CDS (https://vizier.cds.unistra.fr/viz-bin/VizieR?-source=J/A%2BA/643/A2).

Data reduction methods

Pipeline calibration and data quality

The data were initially calibrated at the observatory using the standard ALMA pipeline of the Common Astronomy Software Applications package (CASA) software (McMullin 2007). We checked the automatically-generated calibration reports and identified a few antennae with suspicious behaviors (e.g., phase drifts in the bandpass calibration, unstable phase or gain solutions, anomalously low gains or high system temperatures), which were not flagged by the pipeline. For example, we had to flag the DV19 antenna for all the cycle 6 observations, for which the bandpass phase solution drifted by ~180 deg/GHz in the XX polarization. For half of the observations, no problems were found, and we used directly the data calibrated by the observatory pipeline. Most of the other observations were usually good with only 1 or 2 antennae with possible problems. Four SBs have between 3 and 5 potentially problematic antennae. Considering the very low impact of a single antenna on the final sensitivity, we thus decided to be conservative and fully flag these suspicious antennae and subsequently excised them from our analysis. The list of antennae excluded before the imaging process is listed in the README file.

While the reduction process was generally smooth, we encountered a couple minor issues. The pipeline sometimes flagged the channels of a spectral window overlapping with the noisy edge channels of another spectral window. It was solved by adding the *fracspw*=0.03125 option to the *hifa_flagdata* task before re-running the pipeline script from the observatory. This option flags the edge channels corresponding to 3.125% of the width of the spectral window, while the default is to flag two channels on each side in TDM mode, that is 4/128 = 0.0315. In theory, this command is equivalent to the default routine. In practice, it is not affected by the subtle bug flagging the channels of the other spectral windows when they overlap, which solves our problem. In a few cases, the pipeline used an inconsistent numbering of the spectral windows and we had to manually correct these problematic SBs.

Data cube imaging and production of [CII] moment-0 maps

The datacube were imaged using the *tclean* CASA routine using 0.15 arcsec pixels to well sample the synthesized beam (6 pixels per beam major axis in the field with the sharpest synthesized beam). The clean algorithm is run down to a flux threshold of 3 $sigma_{noise}$, where $sigma_{noise}$ is the standard deviation measured in a previous nonprimary-beam-corrected cube after masking the sources. The determination of the final clean threshold is thus the result of an iterative process. The noise converges very quickly with negligible variations between the second and the third iteration. In practice, the exact choice of the clean threshold has a very low impact on the final flux measurements, since our pointings mostly contain one or a few sources, which are rarely bright. In addition, the natural weighting produces sidelobes and high signal-to-noise ratio (S/N) sources can produce nonnegligible artifacts in the dirty maps or unproperly cleaned maps. We checked that the amplitude of the largest sidelobes is below 10% of the peak of the main beam. The sidelobe residuals after cleaning down to 3 sigma should thus be below 0.3 sigma.

The standard ALPINE products were produced using a natural weighting of the visibilities. This choice maximizes the point-source sensitivity and produces a larger synthesized beam than other weighting schemes, which limits the flux spreading across several beams for slightly extended sources. These cubes are thus optimized to measure integrated properties of ALPINE targets.

We also produced continuum-free cubes. The continuum was subtracted in the uv-plane using the *uvcontsub* CASA routine. This routine takes as input a user-provided range of channels containing line emission, and masks them before fitting a flat continuum model (order 0) to the visibilities. To identify the channels to mask, we used the line properties extracted using the method presented in Sect. 6.1 of Béthermin et al. (2020). We first ran a line search algorithm in a 1-arcsec radius region around the optical coordinates and extracted the spectrum at this position (point source extraction). We then used the position and width of the line in the spectrum to produce a moment-0 map and measured the 2-sigma [CII] contours of the object. Finally, we extracted the spectrum in this 2-sigma contours to get the total integrated spectrum from which we derived the line properties (central frequency, width, amplitude). We use several iterations of the cube production and the line extraction to obtain the final version of these products. To avoid any line contamination, we chose to be conservative and excluded all the channels up to 3-$sigma_v$ from the central frequency of the best Gaussian fit of the line. When a [CII] spectrum exhibits a non-Gaussian excess in the wings, we masked manually an additional ~0.1-0.2GHz to produce conservative continuum-free cubes.

Finally, we generated maps of the [CII] integrated intensity by summing all the channels containing the line emission, i.e. the moment-0. The integration windows were manually defined using the first extraction of the. Contrary to the continuum subtracted cubes, the integration window is not defined in a conservative way but designed to avoid adding noise from channels without signal in the moment-0 maps.

Continuum imaging

We produced continuum maps using the similar method as for the cubes (same clean routine, pixel sizes, and weighting fort the cube imaging, except that the continuum maps were produced using multi-frequency synthesis (MFS, Conway 1990) rather than the channel-by-channel method used for the cubes. The MFS technique exploits the fact that various continuum channels probe various positions in the uv plane to better reconstruct 2-dimensional continuum maps. We excluded the same line-contaminated channels as for the uv-plane continuum subtraction used to produce the cubes. Only the lines of the ALPINE target sources were excluded. Some off-center continuum sources with lines were serendipitously detected in the field. A specific method has been used to measure their continuum flux (not available on the ALMA archive, but on the team website).

Data quality assessment

As explained in the previous section, the quality of the pipeline calibration was checked by our team for each of the 61 scheduling blocks of the survey, and couple of antennae were excluded from the imaging (listed in the REAME file of each MOU). We also inspected visually the products (maps and cubes) to check for artifacts but did not identify anything problematic. Finally, we compared the sources fluxes extracted in map- and uv-space and found an excellent agreement (see Sect. 3.6 and 6.3 of Béthermin et al. 2020).

The original ALPINE products were imaged using a topocentric reference frame (no Doppler tracking). However, considering the low spectral resolution and the broad lines, the difference with LSRK is negligible. This does not affect the current ALPINE results. We forced the LSRK keyword in the FITS header to ingest the data in the ALMA archive (else the archive would reject it) despite the cubes are actually in TOPO. We advise people interested by large statistical compilation of velocity offsets to perform a new imaging if they have a doubt about the significance of the effect.

The source DEIMOS_COSMOS_351640 is not included in this ALMA archive release, because the source name in the observing tool was wrong (DEIMOS_COSMOS_416105). The ALMA archive does not allow to correct the source name after the observations or use inconsistent names between the metadata and the file names. The only solution to ingest it in the ALMA archive would have been to use the wrong name. We preferred not to do it to avoid confusion. The data for this source can be found on the team website (https://cesam.lam.fr/a2c2s/data_release.php).

Other useful resources

The full catalog of the survey containing both ALMA properties (Béthermin et al. 2020) and the ancillary data properties (Faisst et al. 2020) is available at: https://vizier.cds.unistra.fr/viz-bin/VizieR?-source=J/A%2BA/643/A2
Note: Be careful, the SFR in this file is derived from the UV-to-near-IR SED (see Faisst et al. 2020). The infrared SFR must be derived from the infrared luminosity (LIR) listed in the table (details in Béthermin et al. 2020).

Additional material (calibrated measurement sets, tapered products, continuum images of the serendipitous sources after excluding their lines) is available on the team website:
https://cesam.lam.fr/a2c2s/data_release.php