

# The ALMA Redshift 4 Survey (AR4S)

## I. The massive end of the $z = 4$ main sequence of galaxies

C. Schreiber<sup>1,2</sup>, M. Pannella<sup>3,2</sup>, R. Leiton<sup>4,5</sup>, D. Elbaz<sup>2</sup>, T. Wang<sup>2,6</sup>, K. Okumura<sup>2</sup>, and I. Labbé<sup>1</sup>

<sup>1</sup> Leiden Observatory, Leiden University, 2300 RA Leiden, The Netherlands  
e-mail: cschreib@strw.leidenuniv.nl

<sup>2</sup> Laboratoire AIM-Paris-Saclay, CEA/DSM/Irfu – CNRS – Université Paris Diderot, CEA-Saclay, pt. courrier 131, 91191 Gif-sur-Yvette, France

<sup>3</sup> Faculty of Physics, Ludwig-Maximilians Universität, Scheinerstr. 1, 81679 Munich, Germany

<sup>4</sup> Instituto de Física y Astronomía, Universidad de Valparaíso, Avda. Gran Bretaña 1111, 2360102 Valparaíso, Chile

<sup>5</sup> Astronomy Department, Universidad de Concepción, Cessilla 160-C, Concepción, Chile

<sup>6</sup> School of Astronomy and Space Sciences, Nanjing University, 210093 Nanjing, PR China

Received 21 June 2016 / Accepted 16 December 2016

### ABSTRACT

We introduce the ALMA Redshift 4 Survey (AR4S), a systematic ALMA survey of all the known galaxies with stellar mass ( $M_*$ ) larger than  $5 \times 10^{10} M_\odot$  at  $3.5 < z < 4.7$  in the GOODS–south, UDS and COSMOS CANDELS fields. The sample we have analyzed in this paper is composed of 96 galaxies observed with ALMA at  $890 \mu\text{m}$  ( $180 \mu\text{m}$  rest-frame) with an on-source integration time of 1.3 min per galaxy. We detected 32% of the sample at more than  $3\sigma$  significance. Using the stacked ALMA and *Herschel* photometry, we derived an average dust temperature of  $40 \pm 2$  K for the whole sample, and extrapolate the  $L_{\text{IR}}$  and SFR for all our galaxies based on their ALMA flux. We then used a forward modeling approach to estimate their intrinsic sSFR distribution, deconvolved of measurement errors and selection effects: we find a linear relation between SFR and  $M_*$ , with a median sSFR =  $2.8 \pm 0.8$  Gyr and a dispersion around that relation of  $0.28 \pm 0.13$  dex. This latter value is consistent with that measured at lower redshifts, which is proof that the main sequence of star-forming galaxies was already in place at  $z = 4$ , at least among massive galaxies. These new constraints on the properties of the main sequence are in good agreement with the latest predictions from numerical simulations, and suggest that the bulk of star formation in galaxies is driven by the same mechanism from  $z = 4$  to the present day, that is, over at least 90% of the cosmic history. We also discuss the consequences of our results on the population of early quiescent galaxies. This paper is part of a series that will employ these new ALMA observations to explore the star formation and dust properties of the massive end of the  $z = 4$  galaxy population.

**Key words.** galaxies: evolution – galaxies: statistics – galaxies: star formation – submillimeter: galaxies

## 1. Introduction

It is now well established that galaxies have formed most of their stars around  $z = 2$  (e.g., Madau & Dickinson 2014, and references therein) and that the majority of this star-formation activity happens in galaxies that belong to the main sequence (MS) of star-forming galaxies, a tight correlation between the galaxies' stellar mass ( $M_*$ ) and star-formation rate (SFR; e.g., Noeske et al. 2007; Elbaz et al. 2007, 2011; Daddi et al. 2007; Pannella et al. 2009; Schreiber et al. 2015). At a given stellar mass, the typical SFR of galaxies belonging to this sequence has evolved dramatically through time, continuously going down by about an order of magnitude from  $z = 2$  to the present day (e.g., Daddi et al. 2007). This can be explained by a progressive depletion of gas reservoirs (e.g., Daddi et al. 2008; Tacconi et al. 2010) together with an additional decline of the star-formation efficiency over the same time period (e.g., Schreiber et al. 2016).

The situation at  $z > 2$  is less clear. Over the past decade, most of our knowledge of the early Universe has been based on observations of the stellar emission in the rest-frame UV-to-optical, which allow detecting galaxies even beyond the re-ionization era (e.g., Stark et al. 2009; Bouwens et al. 2012; Salmon et al. 2015; Oesch et al. 2016). But in the absence of direct mid- or

far-IR measurement, accurately correcting for absorption by interstellar dust is challenging. Because of the known correlation between mass and attenuation (e.g., Pannella et al. 2015), this is particularly important if one wants to study the massive end of the galaxy population (e.g., Spitler et al. 2014). While small in numbers, these massive galaxies (say, more massive than the Milky Way, with  $M_* > 5 \times 10^{10} M_\odot$ ) contribute about half of the star-formation activity in the Universe at any  $z \leq 3$  (e.g., Schreiber et al. 2015).

To date, observations of distant galaxies in the far-IR or submillimeter with single dish instruments only allowed the detection of the brightest objects, that are experiencing extreme star-formation episodes and may not be representative of the overall population (e.g., Pope et al. 2006; Capak et al. 2011; Riechers et al. 2013). Through stacking of carefully chosen samples, the average SFR and gas mass can be derived (e.g., Magdis et al. 2012; Heinis et al. 2014; Pannella et al. 2015; Schreiber et al. 2015; Béthermin et al. 2015; Tomczak et al. 2016), although these need to be corrected for the effect of source blending and clustering, which is not trivial. Stacking also allows determining the scatter of the stacked properties (Schreiber et al. 2015), however this requires higher  $S/N$  levels which cannot be reached beyond  $z = 3$  even with the deepest *Herschel* images.

Recently, the deployment of the Atacama Large Millimeter Array (ALMA) has allowed us to move forward and detect (or put stringent constraints on) the dust emission of less extreme galaxies at  $z = 4$  and beyond (e.g., Capak et al. 2015; Maiolino et al. 2015; Scoville et al. 2016). However, these first efforts were mostly focused on Lyman break galaxies (LBGs) which are necessarily blue, less dust-obscured and found preferentially at the lowest masses (e.g., Spitler et al. 2014; Wang et al. 2016). At present, we are still lacking a complete census of the massive galaxy population at these redshifts, and this is the gap we intend to fill with this paper.

We therefore introduce here the ALMA Redshift 4 Survey (AR4S<sup>1</sup>), a complete survey of massive galaxies in the *Hubble* Space Telescope (HST) CANDELS fields at  $z \sim 4$  with the ALMA telescope. By probing the rest-frame 180  $\mu\text{m}$  emission, we can put direct constraints on the infrared luminosity, and therefore on the SFR, of about a hundred galaxies at these epochs. We describe the sample in Sect. 2 and the reduction of the ALMA data in Sect. 3. We stacked the ALMA fluxes and *Herschel* images in Sect. 4.1 to measure the average dust temperature in our sample, and use it to extrapolate the SFRs for all our galaxies. We then discuss the location of our galaxies on the SFR– $M_*$  plane in Sect. 4.2, and model their sSFR distribution in Sects. 4.3 and 4.4 to provide the first robust constraints on the normalization and scatter of the MS at  $z = 4$ . Lastly, we briefly discuss the existence of quiescent galaxies in our sample in Sect. 4.5.

In the following, we assume a  $\Lambda$ CDM cosmology with  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ,  $\Omega_M = 0.3$ ,  $\Omega_\Lambda = 0.7$  and a Salpeter (1955) initial mass function (IMF), to derive both SFRs and stellar masses. All magnitudes are quoted in the AB system, such that  $m_{\text{AB}} = 23.9 - 2.5 \log_{10}(S_\nu [\mu\text{Jy}])$ .

## 2. Sample selection

### 2.1. Observed sample

We drew our sample from the CANDELS HST *H*-band catalogs in the three fields covered by deep *Herschel* imaging and accessible with ALMA, namely GOODS–south (Guo et al. 2013), UDS (Galametz et al. 2013) and COSMOS (Nayyeri et al., in prep.). Photometric redshifts and stellar masses for all galaxies in this parent sample were derived in Schreiber et al. (2015). From there, our sample is selected purely on stellar mass ( $M_* > 5 \times 10^{10} M_\odot$ ) and photometric redshift ( $3.5 < z < 4.7$ ) to ensure the highest completeness, resulting in 110 galaxies. We did not try to segregate actively star-forming galaxies from quiescent ones, since the known selection techniques (e.g., the *UVJ* selection; Williams et al. 2009) are uncertain at  $z \geq 4$  (see discussion in Sect. 4.5). Instead, we targeted all galaxies regardless of their potential star-formation activity. We added to our target list 16 lower mass galaxies with a spectroscopic redshift within  $3.5 < z < 4.7$ ; these objects are part of a second sample and will not be discussed further in the present paper.

The median *H*-band magnitude of our sample is 25.2, meaning that these galaxies are faint but still far above the  $5\sigma$  point-source limiting magnitude of  $H = 27$ . However, the *H*-band probes the rest-frame UV emission at  $z = 4$ . Even though the CANDELS HST images are the deepest available to date, this implies that our parent sample is very likely biased against the most heavily obscured galaxies at these redshifts. Indeed, a population of *H*-dropouts (but *Spitzer* IRAC bright) galaxies has

recently been identified in these fields (Wang et al. 2016). If indeed at  $z \sim 4$ , these would represent up to 20% of our selected sample and would therefore only impact our results marginally. These galaxies are currently being observed as part of another ALMA program, and will be discussed in a future work (Wang et al., in prep.). We estimate a similar completeness based on the samples of ultra-red galaxies missed by HST from Huang et al. (2011) and Caputi et al. (2012).

### 2.2. Cleaning the sample

With these short integration times, we chose to include in our selection all the potential  $z = 4$  massive galaxies, regardless of the quality of their photometry. While most galaxies in the sample do have clean measurements in all bands, we identify 14 likely spurious or contaminated sources. Ten sources, the majority, have their *Spitzer* IRAC photometry clearly contaminated by a bright neighboring source (either a star or an extended nearby galaxy). Their redshifts and stellar masses are unreliable, and they are consistently not detected on the ALMA images. Two sources are each very close ( $< 1''$ ) to another nearby galaxy which is also part of our  $z = 4$  sample. Because their redshifts are consistent with being identical with that of their neighbor, and because the ALMA emission originates from the barycenter of the two sources, we choose to consider them as a single object and re-measure their UV-near-IR fluxes and stellar mass in a larger aperture, as described in the next section. Finally, one source is out the *Spitzer* IRAC coverage, and another is partially truncated at the border of the HST *H*-band image.

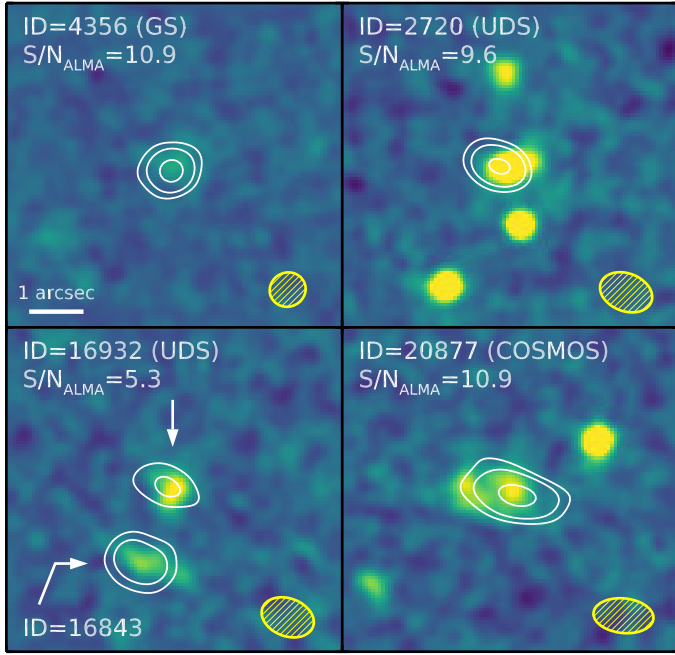
Excluding these objects, we ended up with a final sample of 96 good quality  $z = 4$  massive galaxies, which we analyzed in the following.

### 2.3. Robustness of the redshift and masses determinations

Since our targets are relatively faint, and because the *H* band is tracing the rest-frame UV, it is reasonable to wonder if the template fitting approach used by CANDELS to extract the photometry (using TFIT, Laidler et al. 2007) is adequate, and not missing part of the flux. For example, the well-known  $z = 4$  starburst GN20 has a patchy dust geometry (Hodge et al. 2012), and its rest-UV emission is not representative of the true extent of the galaxy. As a check, we therefore re-measured the photometry of all our targets, specifically looking for missing flux outside of the CANDELS *H*-band segmentation. We summed the flux on the UV-to-near-IR images in large apertures of  $0.9''$  or more, depending on the apparent morphology of the galaxy in the various bands, and indeed find on average an additional 30% flux in the bluest bands. Since the galaxies are not resolved on the *Spitzer* IRAC images, the IRAC fluxes (probing the rest-optical and constraining the stellar mass) are essentially unaffected. We re-analyzed this new photometry as in Schreiber et al. (2015) to obtain redshifts and stellar masses: reassuringly, the photometric redshifts are not affected in any significant way (10% scatter in  $\Delta z/(1+z)$ ), and stellar masses are also globally unchanged (0.06 dex increase on average, with 0.18 dex scatter). The scatter in these quantities is comparable to the one we observe when cross-matching our sample to the 3DHST catalogs (Skelton et al. 2014): 9% for the redshifts and 0.2 dex for the masses.

In the following, to avoid adding extra noise to our selection, we therefore continued to use the original CANDELS photometry except for four galaxies which were clearly missing a large fraction of their flux because of overly aggressive deblending

<sup>1</sup> Read “aras”, which is the French word for macaw birds.



**Fig. 1.** Examples of ALMA-detected galaxies in our sample. The HST F160W image is shown in the background, smoothed by a  $0.3''$  FWHM Gaussian to reveal extended features. The white contours show the extent of the ALMA emission ( $3$ ,  $5$ ,  $10$  and  $17\sigma$ ) after applying tapering. The size and orientation of the ALMA clean beam is given with a yellow hatched region (FWHM). The CANDELS ID of each target is given at the top of each cutout, together with the S/N of the integrated ALMA flux.

(see also previous section). These are ID = 23751 in GOODS–south, 5128 and 35579 in UDS, and 27853 in COSMOS.

### 3. ALMA observations and reduction

#### 3.1. Data

Our targets were observed with dedicated ALMA pointings in band 7, with a central wavelength of  $\sim 890 \mu\text{m}$  ( $888.1 \mu\text{m}$ ,  $337.8 \text{ GHz}$ ), integrating  $1.3 \text{ min}$  on each galaxy for a total observing time of  $6 \text{ h}$ . We reduced the data into calibrated visibilities using the CASA pipeline (version 4.3.1), and produce cleaned continuum images for visual inspection. The achieved angular resolution varies from one field to the other, ranging from  $0.3$  to  $0.7''$  (FWHM of the minor axis of the beam), that is,  $2$  to  $5 \text{ kpc}$  at  $z = 4$ . To measure the effective noise level on comparable grounds, we tapered the longest baselines to reach an homogeneous angular resolution of  $0.7''$  and find an RMS of  $0.15$  to  $0.22 \text{ mJy}$ . Examples of clear detections from the tapered images are given in Fig. 1.

#### 3.2. Flux measurements

We measured the  $890 \mu\text{m}$  flux of all our targets directly in the  $(u, v)$  plane using the `uvmodelfit` procedure from the CASA pipeline. The sources were first modeled with an elliptical Gaussian profile of variable total flux, centroid, width, axis ratio and position angle. When the S/N is too low, the fit becomes unstable and tends to return large position offsets ( $> 2''$ ). In these cases, we discarded the fit and used a simpler model where the position is frozen to that of the HST counterpart, and the size is fixed to the median size of the high S/N galaxies. In both cases

the adopted flux uncertainty is the formal uncertainty returned by `uvmodelfit`, and the fluxes and uncertainties are corrected for primary beam attenuation a posteriori.

The fluxes measured with this method are on average  $37\%$  higher than the peak fluxes read from the cleaned images (with tapering), indicating that most of our targets are resolved by ALMA (the median half-light radius returned by `uvmodelfit` is  $0.3''$ , or  $2 \text{ kpc}$  at  $z = 4$ ). Consequently, the uncertainties are also larger than the image RMS, with an average of  $0.31 \text{ mJy}$ . This corresponds to a  $3\sigma$  detection limit of  $\text{SFR} \approx 294 M_{\odot}/\text{yr}$  at  $z = 4$  (see Sect. 4.1 for the conversion to SFR).

We checked the accuracy of these measurements by fitting fake sources at random positions in the field of view, devoid of significant detection, and find a mean flux of  $0.02 \pm 0.02 \text{ mJy}$ , indicating that our fluxes are not biased. The RMS of these fake measurements is  $0.3 \text{ mJy}$ , and is therefore fully consistent with our average flux uncertainty. As a cross-check, we also compared our flux determination for the galaxy ID = 23751 against the published value from the ALESS program (Hodge et al. 2013). These two independent measurements agree on a total flux of  $8.0 \text{ mJy}$  with uncertainties of  $0.4$  and  $0.6 \text{ mJy}$  in our catalog and that of ALESS, respectively. More details on the reduction and flux measurements will be provided together with the complete catalog in a forthcoming paper (Leiton et al., in prep.).

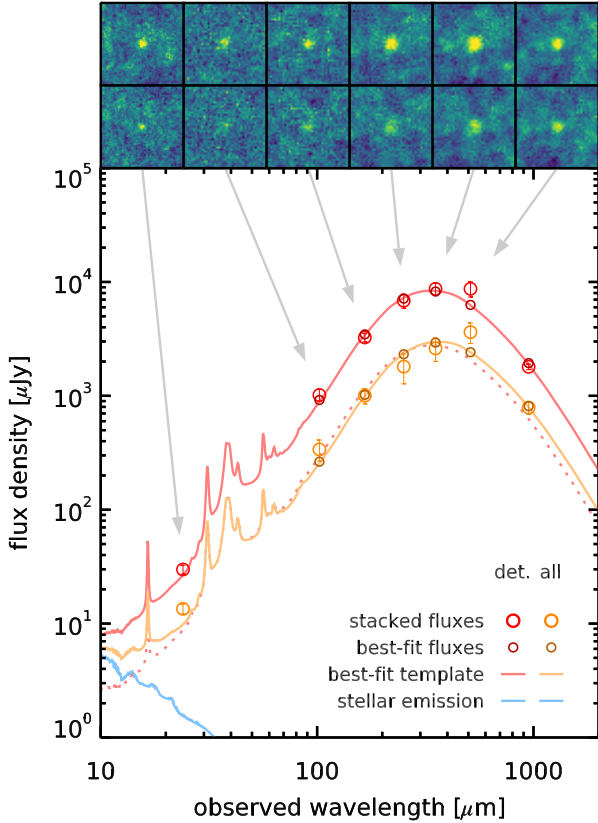
#### 3.3. Properties of the detected and non-detected galaxies

Of all targets in the present sample, we detect 46, 30 and 17 galaxies at  $> 2$ ,  $3$  and  $5\sigma$  significance, respectively. The mean flux of the sample is  $1.00 \pm 0.04 \text{ mJy}$  including non-detections.

Both detected and non-detected galaxies have a similar redshift distribution; a Kolmogorov-Smirnov (KS) test gives a probability  $> 99\%$  that they share the same photometric redshift distribution. The average redshift is  $\langle z \rangle = 3.99$ , with a standard deviation of  $0.36$ . The detections tend to have slightly fainter  $H$ -band magnitudes, although this is barely significant (the KS probability of same distribution is still  $30\%$ ). However the ALMA detections have significantly higher stellar masses than the non-detections: ALMA-detected galaxies constitute the majority of galaxies at  $\log M_* > 11.3$ , while non-detections become dominant at  $\log M_* < 10.9$  (KS test  $< 1\%$ ). This is most likely a consequence of the  $\text{SFR} - M_*$  correlation, as we will show in the following.

#### 3.4. Note on the astrometry

In their survey of the *Hubble* Ultra Deep Field (HUDF, in the center of GOODS–south), Dunlop et al. (2017) reported a systematic position offset of  $\Delta\delta = \delta_{\text{HST}} - \delta_{\text{ALMA}} = +0.24''$  between the ALMA detections and their HST counterparts, with the ALMA emission observed south of the HST. Rujopakarn et al. (2016) confirmed this offset using the JVLA and 2MASS, measuring  $\Delta\delta = +0.26 \pm 0.13''$ , and suggesting an issue in the absolute astrometry of the HST images. None of our targets falls in the HUDF, so we cannot directly double check their result. However, selecting our  $3\sigma$  detections in the whole GOODS field, we consistently find a median systematic shift of  $\Delta\delta = +0.15 \pm 0.05''$  in GOODS–south. In addition, we find a shift of  $\Delta\alpha = -0.14 \pm 0.05''$  in UDS, and no significant shift in COSMOS. These shifts are relatively small compared to the ALMA beam, and vary substantially from one source to another ( $0.2''$  scatter for the  $3\sigma$  detections). We therefore do not attempt



**Fig. 2.** Stacked mid- to far-IR SED of our full sample (orange) and only the  $3\sigma$  ALMA detections (red). The large open circles with error bars are the stacked fluxes, with SPIRE fluxes corrected for flux boosting from clustering (see text). Darker and smaller circles show the best-fit model fluxes, and the corresponding template is shown with a solid line in the background (NB: MIPS  $24\ \mu\text{m}$  was not used in the fit). The blue solid line is the stellar emission, estimated by fitting the stacked UV-to-near-IR photometry with FAST (Kriek et al. 2009). The dotted line is the best-fit SED of the ALMA detections, rescaled to the  $L_{\text{IR}}$  of the whole sample for easier comparison. The stacked cutouts are shown at the top of the plot, both for the ALMA detections (top) and the full sample (bottom). The arrows help identify which flux measurement they correspond to.

to correct for it; since we fit for the centroid for most of our targets, this would have a negligible impact anyway.

## 4. Results

### 4.1. Dust temperature and star-formation rates

To determine the SFR of the galaxies in our sample, we need to extrapolate the 8-to-1000  $\mu\text{m}$  infrared luminosity ( $L_{\text{IR}}$ ) of all galaxies from the 890  $\mu\text{m}$  flux alone. Indeed, most of the galaxies of our sample are undetected in other mid-IR or far-IR bands, and their IR spectral energy distribution (SED) cannot be constrained individually. However, we can stack them together on the *Spitzer* and *Herschel* images to recover their average SED, which can then be used to infer  $L_{\text{IR}}$  assuming it is representative of the whole sample. As in Schreiber et al. (2015), we median stack the images and measure the flux by fitting a point spread function (PSF) model at the center of the stacked image, with a freely varying background level. Because of the poor angular resolution of the images, SPIRE fluxes are boosted by the contribution of clustered nearby objects, and we statistically correct

for this boost following Schreiber et al. (2015). Uncertainties are determined by bootstrapping the sample.

We show in Fig. 2 the median-stacked SED from *Spitzer*, *Herschel* and ALMA, separately for the full sample (96 galaxies) and for the  $3\sigma$  ALMA detections only (30 galaxies). The stacked signal is stronger when we only consider the ALMA detections, and the corresponding stacked fluxes form a coherent SED in all bands. However, this sample is probably not representative of the whole population: the dust temperature ( $T_{\text{dust}}$ ) is known to increase for brighter galaxies above the MS (Magnelli et al. 2014), which are preferentially selected in such flux-limited samples (e.g., Rodighiero et al. 2011). Therefore, we expect the subsample of ALMA-detected galaxies to have on average higher  $T_{\text{dust}}$  than the rest of the population. That being said, the stacked SED of the full sample is more uncertain but does not appear to differ significantly from that of the ALMA detections.

To quantify this, we use the dust SEDs introduced in Schreiber et al. (2016) and fit the stacked photometry. This library has three free parameters:  $L_{\text{IR}}$ ,  $T_{\text{dust}}$  and  $\text{IR}8 \equiv L_{\text{IR}}/L_8$ . Since we have no data to constrain  $L_8$  (the rest-frame 8  $\mu\text{m}$  luminosity), we keep this last parameter fixed at  $\text{IR}8 = 8$  (Reddy et al. 2012; Schreiber et al., in prep.) and only fit for  $L_{\text{IR}}$  and  $T_{\text{dust}}$ . We find that this model provides a good description of our data, with a reduced  $\chi^2$  of 1.3 for both samples. The best-fit dust temperatures are  $T_{\text{dust}} = 40 \pm 2$  and  $43 \pm 1$  K for the full sample and for the ALMA detections, respectively. Therefore we do find that our ALMA detections have on average higher  $T_{\text{dust}}$  values, although the difference is only  $1.3\sigma$  and mild: it would result in a difference of only 0.1 dex when extrapolating  $L_{\text{IR}}$  from the 890  $\mu\text{m}$  flux. Consequently, we choose here to assume a single dust temperature of  $T_{\text{dust}} = 40$  K for all galaxies, and expect our individual SFRs to be systematically biased by up to 0.1 dex. We also note that high dust temperatures around 40 K have already been reported in the recent literature for  $z = 4$  galaxies (e.g., da Cunha et al. 2015; Béthermin et al. 2015).

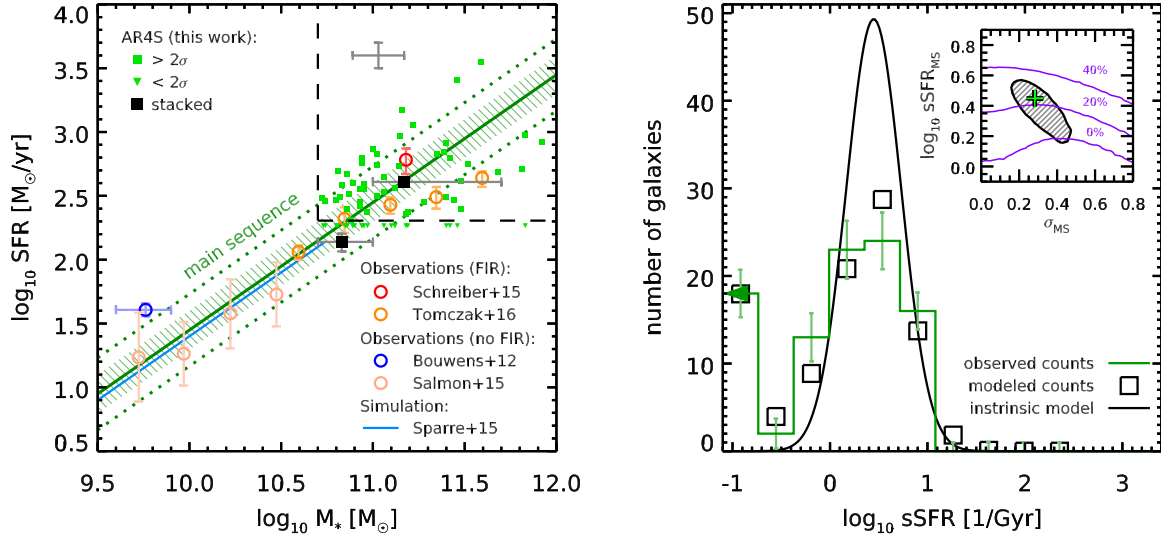
Although the stacked *Herschel* fluxes were corrected for flux boosting by clustering, we find that the 500  $\mu\text{m}$  fluxes are systematically above our best-fit SED. This suggests that our correction, which was derived as an average correction from lower-redshift samples, is insufficient at least in this band. To check that our conclusions are not affected by this potential bias, we re-run the fit excluding the SPIRE bands and find higher  $T_{\text{dust}}$  values by only 0.6 K, which is well within our error bars. We therefore conclude that our 40 K SED is not significantly biased by clustering.

Using this SED, we extrapolated the  $L_{\text{IR}}$  for all our targets and convert it into SFR with the Kennicutt (1998) conversion factor. We also include the contribution of the non-obscured UV light, although it is always negligible.

### 4.2. The $z = 4$ main sequence

We show the location of our galaxies on the  $z = 4$  SFR- $M_*$  plane in Fig. 3 (left). At first order, the detections appear to be distributed around the expected location of the MS, as inferred from previous *Herschel* stacking (e.g., Heinis et al. 2014; Schreiber et al. 2015; Tomczak et al. 2016), with a tendency for higher mass galaxies to have overall higher SFRs. However, because of our relatively low detection rate, deriving conclusions on the locus and scatter of the MS from this figure alone can prove difficult.

Instead, we use a forward modeling approach where we model the observed distribution of  $\text{sSFR} = \text{SFR}/M_*$  (shown on the right panel of the same figure) for both detected and



**Fig. 3.** *Left:* relation between the star-formation rate (SFR) and stellar mass ( $M_*$ ) at  $z = 4$ . Our ALMA sample is shown with small green squares ( $>2\sigma$ ), and downward-facing triangles ( $<2\sigma$ ). The gray cross at the top of the plot indicates the systematic uncertainty affecting our measurements (see text). The median SFR of our sample in two mass bins is shown with black squares (correcting the median for the contaminant population, see text). The green lines are our estimate of the  $z = 4$  MS locus (solid line) and scatter (dotted lines) as determined from modeling the sSFR distribution (see text and figure on the *right*), and the hatched area in the background indicates the uncertainty on the MS locus. Observations from the literature are shown with open circles, from Bouwens et al. (2012) (blue) and Salmon et al. (2015) (pink) who both derived their SFRs without mid-IR or far-IR observations, while Tomczak et al. (2016) (orange,  $z = 3.5$ ) and Schreiber et al. (2015) (red) both used *Herschel* stacking. The latter are corrected by  $-0.1$  dex to account for the difference between the mean and median. Finally, the MS relation found in the Illustris simulation is shown as a blue line (Sparre et al. 2015). *Right:* distribution of  $sSFR = SFR/M_*$  for our sample. The observed counts (including non-detections) are shown with a green histogram and error bars. The best-fit modeled intrinsic distribution is shown in black in the background, and the modeled counts (simulating measurement uncertainties and selection effects) are shown with open squares. The  $1\sigma$  confidence region for the model parameters is shown in *inset*, where the green cross gives the position of minimum  $\chi^2$ , and the purple lines indicate the contaminant fraction required by the fit (see text).

non-detected galaxies, without stacking them. The model assumes an intrinsic functional form and includes all quantifiable sources of selection and observational effects (see, e.g., Mullaney et al. 2015). We describe the method we employ in the next section.

#### 4.3. Modeling the sSFR distribution

In the following, we will make the distinction between “intrinsic” and “observed” mock quantities. The former are the real galaxy properties, free of measurement error and systematics, while the latter are an attempt at simulating the real observations we report in this work, after including all sources of noise. Owing to the modest size of our sample, we settle for a simple three-parameter model where the sSFR of star-forming galaxies is distributed according to a log-normal law of logarithmic width  $\sigma_{MS}$ , and median  $sSFR_{MS}$  (e.g., Rodighiero et al. 2011; Sargent et al. 2012; Schreiber et al. 2015; Ilbert et al. 2015). In addition, we allow a fraction  $f_c$  of galaxies in the model to be flagged as “contaminants”, in which case their sSFR is set to zero. This last step takes into account two sources of contamination in our sample. First and foremost is the existence of quiescent galaxies, which have no star formation by definition, and which could contribute as much as 30% of our sample (Straatman et al. 2014; Spitler et al. 2014). A secondary source of contamination are redshift interlopers and brown dwarfs which we would have failed to identify with the available photometry, and which are expected to be substantially fainter at submillimeter wavelengths.

We created a grid for the first two parameters (as displayed in the inset of Fig. 3, right), and model the observed sSFR distribution for each cell of that grid. To do so, we generated a mock catalog of 100 000 galaxies to which we attribute an observed redshift and stellar mass by drawing randomly and independently from their respective distributions in our real  $z = 4$  sample. We then perturbed these redshifts and masses with a Gaussian distribution of width  $\Delta z/(1+z) = 6.5\%$  and  $\Delta \log_{10} M_*/M_\odot = 0.14$  dex, respectively, to obtain the associated intrinsic quantities ( $1/\sqrt{2}$  times the values we obtained in Sect. 2.3). It is important to take into account these sources of uncertainty to constrain the intrinsic scatter of the main sequence; if we had not modeled them, our measured intrinsic scatter (see next section) would have been 0.05 dex higher. Conversely, if we had used more pessimistic values of  $\Delta z/(1+z) = 10\%$  and  $\Delta \log_{10} M_*/M_\odot = 0.2$  dex, our measured scatter would have been smaller by 0.02 dex. It turns out this variation is well within the statistical uncertainties, so our conclusions do not strongly depend on these values.

Given these intrinsic redshifts and masses, we then use the assumed MS model to derive the intrinsic sSFR of each mock galaxy, including a mild redshift evolution of  $\langle sSFR \rangle \sim (1+z)^{1.5}$  (Schreiber et al. 2015). The next step is to infer the  $890 \mu\text{m}$  flux. To do so, we first converted the sSFR into  $L_{IR}$  using the intrinsic stellar mass and the Kennicutt (1998) conversion factor. We then attributed an intrinsic  $T_{dust}$  to each mock galaxy, with an average of 40 K and a Gaussian scatter of 3 K (as observed at lower redshifts, e.g., Ciesla et al. 2014), and used the corresponding SED to infer the intrinsic ALMA flux. The observed flux is obtained simply by adding a random Gaussian noise, whose amplitude is

drawn from the flux uncertainty distribution. Given these mock observed fluxes, we applied the same procedure as in the previous section to derive the observed SFR (including, in particular, the assumption of a fixed  $T_{\text{dust}} = 40$  K), and finally obtained the observed sSFR by dividing this value by the observed  $M_*$ . We applied the same procedure to a second mock catalog where the intrinsic ALMA fluxes are all set to zero to model the contaminant population, and build histograms of the observed sSFR for both mock catalogs. These histograms are normalized to the total number of simulated galaxies (100 000). A linear combination of these two histograms is then fit to our observed data, adjusting their respective normalization to match the total number of observed galaxies, and the relative weight of each histogram allows us to derive  $f_c$ . In this fit, the uncertainties on the sSFR histograms were derived by repeatedly perturbing all the sSFRs by their respective uncertainties, and computing the standard deviation of the counts in each bin. As a last step we noted down the  $\chi^2$  of the fit and proceeded to the next cell in the grid.

#### 4.4. Result

In Fig. 3 (right) we show the intrinsic sSFR distribution that produces the smallest  $\chi^2$  when compared to our observations. This distribution has  $\sigma_{\text{MS}} = 0.28^{+0.14}_{-0.12}$  dex and  $\log_{10}(\text{sSFR}_{\text{MS}} [1/\text{Gyr}]) = 0.45^{+0.11}_{-0.18}$ , and the fit required a contaminant fraction of  $f_c = 25\%$ .

This median sSFR is lower but still consistent with our previous determination from *Herschel* stacking of  $\log_{10}(\text{sSFR}_{\text{MS}} [1/\text{Gyr}]) = 0.6 \pm 0.1$  (Schreiber et al. 2015), and the width of the distribution appears unchanged from the value of  $\sigma_{\text{MS}} = 0.31 \pm 0.02$  dex we obtained at  $z \sim 1$  (see also Ilbert et al. 2015; Guo et al. 2015). Overall, this confirms the theoretical expectation that the MS paradigm still holds in the early Universe (e.g., Sparre et al. 2015). This agreement is both qualitative and quantitative, since our determination of the locus and scatter of the MS is also in agreement with the prediction of the latest numerical simulations. Sparre et al. (2015) indeed found a scatter of 0.2–0.3 dex at  $z = 4$  in the Illustris simulation with a scatter closer to 0.3 dex at high mass, although at these redshifts they lack the volume to constrain the  $M_* > 3 \times 10^{10} M_\odot$  population. Similarly, Mitchell et al. (2014) found a 0.27 dex scatter in the GALFORM semi-analytic model at  $z = 3$ , this time with an increase toward higher redshifts (contrary to Sparre et al. who found, if anything, an opposite evolution). Our data are still insufficiently deep to rule out current models based on the scatter only, however the  $z = 3$  MS normalization of Mitchell et al. (as also discussed in their work) is particularly low. Assuming a typical redshift dependence of  $\text{sSFR} \sim (1+z)^{-2.5}$  (e.g., Dekel et al. 2013), it is inconsistent with our observations at the  $3.5\sigma$  level.

As a check, we then compute the median SFR of our sample in two bins of stellar mass above and below  $M_* = 10^{11} M_\odot$  (still including non-detections), and divide the obtained values by  $(1 - f_c)$  to statistically remove the contribution of the contaminants. The obtained SFRs,  $138 \pm 22$  and  $406 \pm 25 M_\odot/\text{yr}$ , are reported in Fig. 3 (left), and clearly illustrate the positive correlation between SFR and  $M_*$ . The low-mass point is found below the SFR– $M_*$  relation derived from the sSFR modeling, although at a significance of only  $0.7\sigma$ . This may suggest that the fraction of contaminant is higher among low-mass galaxies, which can be expected given that these are about one magnitude fainter and are therefore more likely to be wrongly characterized.

#### 4.5. Contaminants: quiescent galaxies or redshift interlopers?

The contaminant fraction  $f_c = 25\%$  is low, and remains consistent with the observed number of quiescent galaxies at these redshifts (Muzzin et al. 2013; Straatman et al. 2014). We cannot rule out, however, that a substantial fraction of these “contaminants” could be low redshift interlopers. At first order, we can obtain an estimate of how many such outliers contaminate our sample by comparing our photometric redshifts ( $z_{\text{phot}}$ ) at  $3.5 < z_{\text{phot}} < 5$  against spectroscopic determinations ( $z_{\text{spec}}$ ) from the literature. Since only one of the AR4S galaxy has a  $z_{\text{spec}} = 3.582$ , we widen our selection to include 110 galaxies of lower stellar masses. Among these, 4% have  $z_{\text{spec}} < 2$ . This spectroscopic sample has substantially bluer colors than the typical AR4S galaxy (observed  $(r - H) \sim 1$  and 3, respectively), so we expect this fraction to be larger in AR4S, possibly up to 10%. Subtracting these outliers from our contaminant population would imply that quiescent galaxies contribute only 15% of our sample, which would be more consistent with the extrapolation of the quiescent fraction from lower redshifts. On the other hand, as shown in Fig. 3 (right, inset),  $f_c$  can range from 0 to 40% within the  $1\sigma$  confidence interval for our two model parameters. We therefore refrain from drawing strong conclusions out of this number.

Independent constraints on the quiescent population could be provided by their rest-optical colors. However, as mentioned in Sect. 2, the standard *UVJ* selection of quiescent galaxies is uncertain at  $z = 4$  because the rest-frame *J* ( $1.2 \mu\text{m}$ ) band shifts into the *Spitzer* IRAC  $5.8 \mu\text{m}$  band. While the  $5.8 \mu\text{m}$  observations are deep enough in the GOODS–south field, they are shallower in UDS and COSMOS, and because they require cryogenic cooling (which is no longer available on-board *Spitzer*) their depth has not improved during the past years. Probably as a consequence of this poorly constrained  $V - J$  color, 5 out of 30 of our ALMA detections are classified as *UVJ*-quiescent (one of them being in GOODS–south). Encouragingly, the fraction of *UVJ* quiescent galaxies is larger among the ALMA non-detections (31 out of 66), but their stacked ALMA flux of  $0.4 \pm 0.1$  mJy suggests that not all are truly quiescent.

The red colors of these galaxies could be mistaken for a quiescent stellar population if these galaxies contain dust-obscured AGNs (e.g., Donley et al. 2012). Among the 5 *UVJ*-quiescent ALMA detections, only one of them is strongly detected in MIPS  $24 \mu\text{m}$ , and none is detected in the available X-ray or radio images. Among the 31 non-detections, we find one  $24 \mu\text{m}$  detection, one X-ray detection and one X-ray+radio detection. Therefore the majority of these galaxies do not show sign of AGN activity.

We conclude that the *UVJ* classification in CANDELS UDS and COSMOS is indeed unreliable at these redshifts, and until the launch of the *James Webb* Space Telescope the only way to robustly identify these  $z \geq 4$  quiescent objects is to constrain their individual star-formation activity, either with deep near-IR spectroscopy or ALMA high-frequency imaging.

## 5. Conclusions

We introduce AR4S, a systematic ALMA survey of all the known massive galaxies ( $M_* > 5 \times 10^{10} M_\odot$ ) at  $z = 4$  in the deepest fields observed with *Hubble* and *Herschel*. With only 6 h of telescope time, we detect 30 out of 95 galaxies: our strategy of targeting mass-selected samples is an order of magnitude more efficient at detecting high redshift galaxies compared to

contiguous surveys (e.g., Dunlop et al. 2017; Walter et al. 2016). Detailed properties of the detected and non-detected galaxies, as well as the full catalog, will be discussed in a forthcoming paper (Leiton et al., in prep.).

Using these new data, we first build the average dust SED of our sample and find an average dust temperature of 40 K, in agreement with recent studies of  $z \sim 4$  galaxies.

From this SED, we extrapolate the total  $L_{\text{IR}}$  (hence SFR) from the ALMA fluxes of our individual objects, and model the observed sSFR distribution with a three-parameter model inspired by observations at lower redshifts. This analysis suggests that galaxies at  $z = 4$  follow a linear relation between SFR and  $M_*$ , with a dispersion around that relation of 0.3 dex. This value is the same as that measured at lower redshifts, which is proof that the MS is already in place at  $z = 4$ , at least among massive galaxies. These new constraints on the properties of the MS are also in good agreement with the latest prediction from numerical simulations (e.g., Illustris).

Finally, our results are compatible with the existence of a population of massive quiescent galaxies early in the history of the Universe, although spectroscopic confirmation and deeper ALMA imaging would be required to draw reliable conclusions. Future works with this data will include a study of the dust attenuation properties of the galaxies in our sample (Pannella et al., in prep.) and the geometry of their star-forming regions (Elbaz et al., in prep.).

*Acknowledgements.* Most of the analysis for this paper was done using `phy++`, a free and open source C++ library for fast and robust numerical astrophysics (<http://cschreib.github.io/phypp/>). This paper makes use of the following ALMA data: ADS/JAO.ALMA#2013.1.01292.S. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada) and NSC and ASIAA (Taiwan), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ. This work is based on observations taken by the CANDELS Multi-Cycle Treasury Program with the NASA/ESA HST, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. R.L. acknowledges the financial support from FONDECYT through grant 3130558. T.W. acknowledges the financial support from the European Union Seventh Framework Program (FP7/2007–2013) under grant agreement No. 312725 (ASTRODEEP).

## References

- Béthermin, M., Daddi, E., Magdis, G., et al. 2015, *A&A*, 573, A113  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2012, *ApJ*, 754, 83  
 Capak, P. L., Riechers, D., Scoville, N. Z., et al. 2011, *Nature*, 470, 233  
 Capak, P. L., Carilli, C., Jones, G., et al. 2015, *Nature*, 522, 455  
 Caputi, K. I., Dunlop, J. S., McLure, R. J., et al. 2012, *ApJ*, 750, L20  
 Ciesla, L., Boquien, M., Boselli, A., et al. 2014, *A&A*, 565, A128  
 da Cunha, E., Walter, F., Smail, I. R., et al. 2015, *ApJ*, 806, 110  
 Daddi, E., Dickinson, M., Morrison, G., et al. 2007, *ApJ*, 670, 156  
 Daddi, E., Dannerbauer, H., Elbaz, D., et al. 2008, *ApJ*, 673, L21  
 Dekel, A., Zolotov, A., Tweed, D., et al. 2013, *MNRAS*, 435, 999  
 Donley, J. L., Koekemoer, A. M., Brusa, M., et al. 2012, *ApJ*, 748, 142  
 Dunlop, J. S., McLure, R. J., Biggs, A. D., et al. 2017, *MNRAS*, 466, 861  
 Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, *A&A*, 468, 33  
 Elbaz, D., Dickinson, M., Hwang, H. S., et al. 2011, *A&A*, 533, A119  
 Galametz, A., Grazian, A., Fontana, A., et al. 2013, *ApJS*, 206, 10  
 Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, *ApJS*, 207, 24  
 Guo, Y., Ferguson, H. C., Bell, E. F., et al. 2015, *ApJ*, 800, 39  
 Heinis, S., Buat, V., Béthermin, M., et al. 2014, *MNRAS*, 437, 1268  
 Hodge, J. A., Carilli, C. L., Walter, F., et al. 2012, *ApJ*, 760, 11  
 Hodge, J. A., Karim, A., Smail, I., et al. 2013, *ApJ*, 768, 91  
 Huang, J., Zheng, X. Z., Rigopoulou, D., et al. 2011, *ApJ*, 742, L13  
 Ilbaf, O., Arnouts, S., Le Flo'c'h, E., et al. 2015, *A&A*, 579, A2  
 Kennicutt, Robert C., J. 1998, *ARA&A*, 36, 189  
 Kriek, M., van Dokkum, P. G., Labbé, I., et al. 2009, *ApJ*, 700, 221  
 Laidler, V. G., Papovich, C., Grogin, N. A., et al. 2007, *PASP*, 119, 1325  
 Madau, P., & Dickinson, M. 2014, *ARA&A*, 52, 415  
 Magdis, G. E., Daddi, E., Béthermin, M., et al. 2012, *ApJ*, 760, 6  
 Magnelli, B., Lutz, D., Saintonge, A., et al. 2014, *A&A*, 561, A86  
 Maiolino, R., Carniani, S., Fontana, A., et al. 2015, *MNRAS*, 452, 54  
 Mitchell, P. D., Lacey, C. G., Cole, S., & Baugh, C. M. 2014, *MNRAS*, 444, 2637  
 Mullaney, J. R., Alexander, D. M., Aird, J., et al. 2015, *MNRAS*, 453, L83  
 Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, *ApJ*, 777, 18  
 Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, *ApJ*, 660, L43  
 Oesch, P. A., Brammer, G., van Dokkum, P. G., et al. 2016, *ApJ*, 819, 129  
 Pannella, M., Carilli, C. L., Daddi, E., et al. 2009, *ApJ*, 698, L116  
 Pannella, M., Elbaz, D., Daddi, E., et al. 2015, *ApJ*, 807, 141  
 Pope, A., Scott, D., Dickinson, M., et al. 2006, *MNRAS*, 370, 1185  
 Reddy, N., Dickinson, M., Elbaz, D., et al. 2012, *ApJ*, 744, 154  
 Riechers, D. A., Bradford, C. M., Clements, D. L., et al. 2013, *Nature*, 496, 329  
 Rodighiero, G., Daddi, E., Baronchelli, I., et al. 2011, *ApJ*, 739, L40  
 Rujopakarn, W., Dunlop, J. S., Rieke, G. H., et al. 2016, *ApJ*, 833, 12  
 Salmon, B., Papovich, C., Finkelstein, S. L., et al. 2015, *ApJ*, 799, 183  
 Salpeter, E. E. 1955, *ApJ*, 121, 161  
 Sargent, M. T., Béthermin, M., Daddi, E., & Elbaz, D. 2012, *ApJ*, 747, L31  
 Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, *A&A*, 575, A74  
 Schreiber, C., Elbaz, D., Pannella, M., et al. 2016, *A&A*, 589, A35  
 Scoville, N., Sheth, K., Aussel, H., et al. 2016, *ApJ*, 820, 83  
 Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, *ApJS*, 214, 24  
 Sparre, M., Hayward, C. C., Springel, V., et al. 2015, *MNRAS*, 447, 3548  
 Spitler, L. R., Straatman, C. M. S., Labbé, I., et al. 2014, *ApJ*, 787, L36  
 Stark, D. P., Ellis, R. S., Bunker, A., et al. 2009, *ApJ*, 697, 1493  
 Straatman, C. M. S., Labbé, I., Spitler, L. R., et al. 2014, *ApJ*, 783, L14  
 Tacconi, L. J., Genzel, R., Neri, R., et al. 2010, *Nature*, 463, 781  
 Tomczak, A. R., Quadri, R. F., Tran, K. H., et al. 2016, *ApJ*, 817, 118  
 Walter, F., Decarli, R., Aravena, M., et al. 2016, *ApJ*, 833, 67  
 Wang, T., Elbaz, D., Schreiber, C., et al. 2016, *ApJ*, 816, 84  
 Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labbé, I. 2009, *ApJ*, 691, 1879