Interpreting the *Spitzer*/IRAC colours of $7 \leq z \leq 9$ galaxies: distinguishing between line emission and starlight using ALMA

G. W. Roberts-Borsani [ORCID],[1,2]★ R. S. Ellis[2] and N. Laporte[3,4,2]

[1]*Department of Physics and Astronomy, University of California, Los Angeles, 430 Portola Plaza, Los Angeles, CA 90095, USA*
[2]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[3]*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[4]*Cavendish Laboratory, University of Cambridge, 19 JJ Thomson Avenue, Cambridge CB3 0HE, UK*

ABSTRACT

Prior to the launch of JWST, *Spitzer*/IRAC photometry offers the only means of studying the rest-frame optical properties of $z > 7$ galaxies. Many such high-redshift galaxies display a red [3.6]$-$[4.5] micron colour, often referred to as the 'IRAC excess', which has conventionally been interpreted as arising from intense [O III]+Hβ emission within the [4.5] micron bandpass. An appealing aspect of this interpretation is similarly intense line emission seen in star-forming galaxies at lower redshift as well as the redshift-dependent behaviour of the IRAC colours beyond $z \sim 7$ modelled as the various nebular lines move through the two bandpasses. In this paper, we demonstrate that, given the photometric uncertainties, established stellar populations with Balmer (4000 Å rest frame) breaks, such as those inferred at $z > 9$ where line emission does not contaminate the IRAC bands, can equally well explain the redshift-dependent behaviour of the IRAC colours in $7 \lesssim z \lesssim 9$ galaxies. We discuss possible ways of distinguishing between the two hypotheses using ALMA measures of [O III] $\lambda 88$ micron and dust continuum fluxes. Prior to further studies with JWST, we show that the distinction is important in determining the assembly history of galaxies in the first 500 Myr.

Key words: Galaxy: evolution – galaxies: high-redshift – cosmology: early Universe – cosmology: dark ages, reionization, first stars.

1 INTRODUCTION

The last few years have seen impressive progress in studies of galaxies in the so-called 'reionization era' corresponding to the redshift interval $7 < z < 10$. However, the number of spectroscopically confirmed examples remains limited and much has been deduced from spectral energy distributions (SEDs) of photometric samples. In addition to demographic studies based on star formation rate densities (Oesch et al. 2014; McLeod, McLure & Dunlop 2016) and luminosity functions (Atek et al. 2015; Bouwens et al. 2015), a key area of interest is studies of the gaseous and stellar properties of early systems. The latter topic is central to understand both the ionizing capability of early galaxies as well as the age of their stellar populations (for a recent review see Stark 2016).

Although much of the progress has been made using photometric samples based on Hubble imaging, both in deep fields (Grogin et al. 2011; Koekemoer et al. 2011; Ellis et al. 2013) and through lensing clusters (Bradley et al. 2014; Lotz et al. 2017; Salmon et al. 2018; Coe et al. 2019), the *Spitzer Space Telescope* has made a key contribution since, at $z \gtrsim 5$, the two bandpasses at 3.6 and 4.5 μm sample the rest-frame optical. At redshifts of $z \gtrsim 6.6$–6.8, it is claimed that the redshift-dependent trend of the IRAC colours is consistent with intense nebular emission lines shifting through the

bandpasses (Labbé et al. 2013; Smit et al. 2015; Roberts-Borsani et al. 2016), although the precision of this exercise is dependent mostly on samples with only photometric redshifts. With this in mind, the surprising spectroscopic confirmation with Lyα (Oesch et al. 2015; Zitrin et al. 2015; Stark et al. 2017) of the four brightest $z > 7$ galaxies in the CANDELS survey selected to display red *Spitzer*/IRAC 3.6m$-$4.5 μm colours (and hence an 'IRAC excess') of >0.5 mag (Roberts-Borsani et al. 2016), reinforced the hypothesis that the IRAC excess arises from intense [O III] $\lambda\lambda$ 4959,5007 Å plus Hβ emission within the 4.5 μm band. To explain the IRAC colours, the rest-frame equivalent widths (EWs) of [O III]+Hβ must be of order 500 Å.

Although it will not be possible to confirm this suggestion with direct spectroscopy until the launch of JWST, the so-called '[O III] hypothesis' has been widely accepted for several reasons. First, at lower redshift $z \sim 6.6$–6.8 where [O III] passes through the 3.6 μm bandpass, the required blue 3.6$-$4.5 μm colour is seen for a sample of galaxies, several of which are now spectroscopically confirmed with ALMA (Smit et al. 2018; see also Sobral et al. 2015; Pentericci et al. 2016; Matthee et al. 2017). Finally, as a proof of concept, galaxies whose rest-frame [O III] emission exceed EW $\simeq 1000$ Å, while difficult to reproduce via modelling except in very young star-forming systems, have been studied at $z \simeq 2$ (Maseda et al. 2014).

A more recent development has been the location of IRAC-excess galaxies whose photometric redshifts lie at $z > 9$; at these redshifts [O III] $\lambda 5007$ Å – the strongest contaminating emission

★ E-mail: guidorb@astro.ucla.edu

line – is shifted beyond both IRAC filters. Although both H β and [O III] λ4959 Å remain in the filter until $z \sim 9.3$, their relatively low strengths make it difficult to reproduce strong IRAC-excesses. Thus, far only one system, MACS1149-JD1, hereafter JD1, has been spectroscopically confirmed at $z = 9.11$ (Hashimoto et al. 2018). Analysis of its SED attributes the IRAC excess to the Balmer break at 4000 Å consistent with a mature $\sim$200–300 Myr old stellar population providing a first tantalizing glimpse of 'cosmic dawn' at $z \simeq 15 \pm 3$. This discovery raises the question of the extent to which the IRAC excess seen in galaxies at $7 < z < 9$ might also, in part, be due to a similar Balmer break. The distinction is important since it would imply many luminous $z \simeq 7$–9 galaxies may have older stellar populations and larger stellar masses than previously thought, with interesting consequences for the presence of earlier star formation. An additional issue is whether JD1 is representative of the galaxy population at $z \simeq 7$–9 (see Katz et al. 2019).

The goal of this paper is to explore the extent to which an IRAC excess and its redshift-dependent trend might be due, in part, to starlight rather than solely intense [O III] line emission. A plan of the paper follows. In Section 2, we examine predicted IRAC colours in the context of both hypotheses, using carefully chosen template galaxies as well as contemporary stellar population models that incorporate nebular line emission. In Section 3, we turn to what data might be needed to distinguish between the two hypotheses. Prior to spectroscopy with JWST, we consider the flux of [O III] λ88 μm that is accessible with ALMA and examine the IRAC colour for those spectroscopically confirmed galaxies for which [O III] λ88 μm fluxes are available. In Section 4, we discuss our results and the implications on the early assembly of galaxies. Throughout this paper, we refer to the *HST* F160W and *Spitzer*/IRAC 3.6 and 4.5 micron bands as H_{160}, [3.6] and [4.5], respectively, for simplicity. We also assume $H_0 = 70 \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$. All magnitudes are in the AB system (Oke & Gunn 1983).

2 MODELLING THE IRAC COLOURS OF $7 < z < 9$ GALAXIES

In exploring the redshift-dependent behaviour of the IRAC 3.6 and 4.5 μm colours, under the hypotheses of contributions from [O III] line emission or a Balmer break due to a more mature stellar population, we begin by selecting two template SEDs fitted to actual data for spectroscopically confirmed $z > 7$ galaxies. For the case of intense [O III] emission, we use a spectroscopic template fitted to the ground-based and *HST*/*Spitzer* photometry of EGSY8p7 at $z = 8.68$ (Zitrin et al. 2015), one of the four IRAC-excess bright sources first identified in the CANDELS survey (Roberts-Borsani et al. 2016). The template and *HST*/*Spitzer* photometry used here are taken directly from the latter study, and we refer the reader to that paper for details on the construction of the templates and derivation of the photometric data points.

Similarly, for the case of a mature stellar population with a prominent Balmer break, we select the SED fit to MACS1149-JD1 at $z = 9.11$ from Hashimoto et al. (2018). This fit represents a composite of a mature $\sim$200–300 Myr population augmented with a younger component invoked to match the intensity of [O III] emission at 88 μm discovered with ALMA; full details can be found in Hashimoto et al. (2018). Since this spectral fit includes nebular line emission which, while not contributing significantly to the IRAC bands at $z > 9$ will do so at lower redshift, we also explore the effect of suppressing all optical emission lines from the fitted spectrum of JD1. The observed (rest-frame) EWs of the combined [O III] and H β lines are $\mathrm{EW_{obs}}([O \, III]+H \beta) \approx 6690$ Å ($\mathrm{EW_{rest}}([O \, III]+H \beta) \approx$
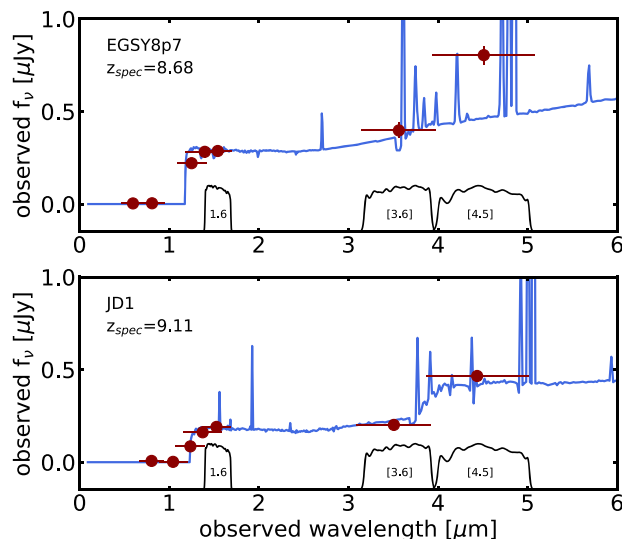

Figure 1. The two synthetic spectra (the blue lines) and the associated observed photometry (the dark red points with the error bars; Roberts-Borsani et al. 2016; Zheng et al. 2017) used in this study to demonstrate the similarity of the redshift-dependent *Spitzer*/IRAC [3.6]−[4.5] evolution. The spectra are of EGSY8p7 (Zitrin et al. 2015; Roberts-Borsani et al. 2016), a supposed extreme [O III]+H β line emitter at $z = 8.68$ (top), and JD1 (Hashimoto et al. 2018), a Balmer break galaxy at $z = 9.11$ (bottom). The JD1 spectrum shown here is uncorrected for magnification. Shown at the bottom of each panel are the *HST*/H_{160} and *Spitzer*/IRAC 3.6 and 4.5 μm response filters, for reference.

770 Å) and $\mathrm{EW_{obs}}([O \, III]+H \beta) \approx 3408$ Å ($\mathrm{EW_{rest}}([O \, III]+H \beta) \approx 375$ Å) for EGSY8p7 and JD1, respectively. The adopted spectral templates and associated photometry are shown in Fig. 1, and for all subsequent analysis, we normalize both spectra by their flux at 0.325 μm (rest frame), where the spectra are free from emission or absorption features, to ensure their 3.6 and 4.5 μm photometry can be directly compared.

We are now in a position to explore how these template spectra affect the IRAC colours over the redshift range $6.8 < z < 9$. In Fig. 2, we present the redshift evolution of the [3.6]−[4.5] and H_{160} −[3.6] colours in steps of $\Delta z = 0.02$ for each of our fiducial spectra, following a similar approach by Labbé et al. (2013). At each redshift interval, the colours are measured directly from the redshifted spectrum using the relevant filter response curves. The simulation shows that the Balmer break in JD1 can mimic the effect of intense line emission to within ± 0.1 mag in the [3.6]−[4.5] colour, particularly at redshifts $z \gtrsim 7.5$, and even produce redder colours at $z \gtrsim 8.5$. Since the JD1 template includes a contribution from nebular emission lines (e.g. [O II], H β, and [O III]), including some from H β and [O III] λ4959 Å at $z = 9.11$ in the 4.5 μm band, we explored suppressing all optical line emission in this template but find consistently red IRAC colours with virtually identical colour–colour evolution (particularly at $z \gtrsim 8$) and little difference in normalization: the masked spectrum consistently produces red IRAC [3.6]−[4.5] colours $\lesssim 0.2$ mag lower than its unmasked counterpart) on the simulated colours (see discussion next and Fig. 2).

The similarity between the redshift-dependent trends of a Balmer break and intense line emission over the chosen redshift interval $7 < z < 9$ may seem surprising given earlier conclusions of a similar exercise undertaken by Labbé et al. (2013). Those authors explored the effect by analysing average SEDs (containing nebular emission lines) of $z_{phot} \approx 7$ and $z_{phot} \approx 8$ galaxies selected as Lyman break
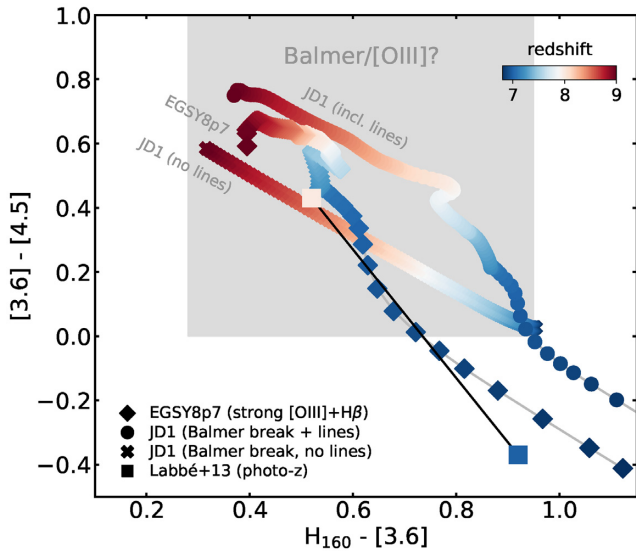
Figure 2. The [3.6]−[4.5] versus H_{160} −[3.6] colours as a function of redshift from $6.8 \leq z \leq 9$ for the spectral templates of spectroscopically confirmed galaxies EGSY8p7 (the diamonds; assumed to be an extreme [O III]+Hβ emitter) and JD1 (circles; a Balmer break galaxy) and the average SED results of Labbé et al. (2013; the squares). Each are connected by the grey lines or a black vector for clarity. Since the template of JD1 includes some optical nebular emission, we also show the results for JD1 with masked emission lines (the crosses) to highlight the contribution from the stellar continuum only. The grey-shaded region indicates the colour space over which the different spectra produce similar colours for the selected redshift interval. A clear difference is evident between spectroscopic and photometric results, for which blue *Spitzer*/IRAC colours at $z_{\mathrm{phot}} \sim 7$ are likely driven by strong nebular line emitters at $z \sim 6.6$–6.8.

'dropouts' in the HUDF09+ERS and CANDELS-GOODS South fields. They were able to reproduce the average [3.6]−[4.5] and H_{160} −[3.6] colour differences only through the inclusion of high EW nebular emission lines (EW([O III]+Hβ) $\sim$ 560–670 Å), rejecting an increased dust content or stellar age as major contributors to the colour evolution. We show their average $z \sim 7$ and $z \sim 8$ colours together with our results in Fig. 2.

However, while samples of $z \sim 8$ galaxies selected as Y-dropouts over the aforementioned fields have average photometric redshifts of $z \sim 7.9$ (Bouwens et al. 2011; Oesch et al. 2012; Bouwens et al. 2015), $z \sim 7$ galaxies selected as z_{850} dropouts have average redshifts of $z \sim 6.8$ (Bouwens et al. 2015). As shown by Smit et al. (2015) and Roberts-Borsani et al. (2016), the *Spitzer*/IRAC [3.6]−[4.5] colour changes dramatically from $z \approx 7 \rightarrow 6.8$ due to the strong [O III]+Hβ lines transiting between filters, resulting in a change from especially red to especially blue colours. Thus, the blue colours observed in the Labbé et al. (2013) $z \sim 7$ stack are conceivably due to the presence of $z \sim 6.8$ galaxies that boost the $z = 7 \rightarrow 8$ IRAC colour difference (as is clearly illustrated by our spectroscopic results where the blue colours are apparent only at $z < 7$ and the red colours at $z \geq 7$), thereby requiring more intense contributions to the colour evolution. The actual colour difference is in fact much smaller ($\sim$0.3–0.4 mag as opposed to $\sim$0.8 mag, see Fig. 2) and consistent with the trends observed for both strong nebular emission and Balmer break SEDs.

At this point, we caution the reader that this exercise is not motivated to claim that the IRAC excess seen in many sources at $7 < z < 9$ cannot be due to intense [O III]+Hβ emission. Rather, we wish to point out that the existence of a Balmer break for JD1 at

$z > 9$ may imply *some contribution of starlight to the IRAC excess seen in sources at $7 < z < 9$* and to explore whether such starlight is prominent in existing spectroscopically confirmed galaxies at $7 < z < 9$. Of course, these results will depend on the choice of spectral template. In the case of JD1 we selected the only case known to date of a galaxy with an IRAC excess that cannot be explained solely via intense line emission (Hashimoto et al. 2018). For the line emitter template, the results will differ slightly depending on which object is chosen from the samples available in the literature. Given this uncertainty plus recent discussions on the possible exceptional case of JD1 (Bingelli et al. 2019), it is helpful to examine the redshift-dependent trends produced by stellar population synthesis models as well as to expand the discussion to compare all these predictions with actual $z > 7$ data in the literature.

For the population synthesis models, we use the Pégase3 suite (Fioc & Rocca-Volmerange 2019), which includes self-consistent modelling of nebular line emission and dust evolution. Clearly, such models have an abundance of free parameters but, for the present exercise, our main goal is to demonstrate that *relative contributions* of synthetic galaxy spectra selected at various ages from a simple star formation history can also reproduce the trends we see using our observed spectral templates. For the current experiment, our simulated galaxy adopts a Chabrier (2003) IMF with a constant star formation history beginning at $z = 15$ and ending at $z = 2$. We select simulated spectra at various time intervals corresponding to galaxy ages from 1 Myr ($z \sim 15$) to 600 Myr ($z \sim 6$). Nebular emission and dust evolution are included in the modelling and, for the former, we accommodate the possibility of a multiplicative factor L_{neb} for the absolute strength of the lines to allow for extreme emission. As with the EGSY8p7 and JD1 spectra, the Pégase3 spectra are normalized to their flux at 0.325 µm (rest frame) prior to analysis.

Fig. 3 shows the IRAC [3.6]–[4.5] colour versus redshift trend for our fiducial galaxy templates as well as the Pégase3 models. For each of our SEDs (i.e. EGSY8p7, JD1 and each of the Pégase3 spectra corresponding to various time intervals of the simulated galaxy's evolution), the spectrum is redshifted across our range of interest and the colour measured through the relevant response filters, to assess the relative colour contributions from each spectrum's features. To illustrate the effects of adopting normal ($L_{\mathrm{neb}} = 1$) and extreme ($L_{\mathrm{neb}} \gg 1$) nebular emission for, respectively, this portion of the analysis and that adopted later, we also overplot the youngest galaxy spectrum with $L_{\mathrm{neb}} = 5$, which is sufficient to match even those most extreme blue and red colours at $z \sim 6.6$–6.8 and $z \sim 7$–7.5, respectively. Additionally, the figure also shows the JD1 template adjusted to exclude the contribution from optical nebular lines. To compare with actual data, Table 1 represents a compilation of 13 $z > 7$ spectroscopically confirmed galaxies drawn from the literature, each with available *HST* and *Spitzer*/IRAC photometry, and we plot their photometric data alongside the spectroscopic results in Fig. 3. For completeness, we also add spectroscopically confirmed sources at $z \simeq 6.8$ sources with especially blue ([3.6]−[4.5] < −0.5 mag) colours (Sobral et al. 2015; Laporte et al. 2017b; Smit et al. 2018; Matthee et al. 2019) that demonstrate the influence of [O III]+Hβ emission in the 3.6 µm band at lower redshift.

Focusing initially on the comparison between Pégase3 and our chosen galaxy templates, we can see very similar trends, albeit with some difference in normalization. Within the $7 < z < 9$ redshift range, Pégase models corresponding to younger ages closely track the evolution of the EGSY8p7 template, whose red colours are dominated by strong nebular emission lines, whilst evolved stellar ages are required to explain the evolution of the JD1 templates (both
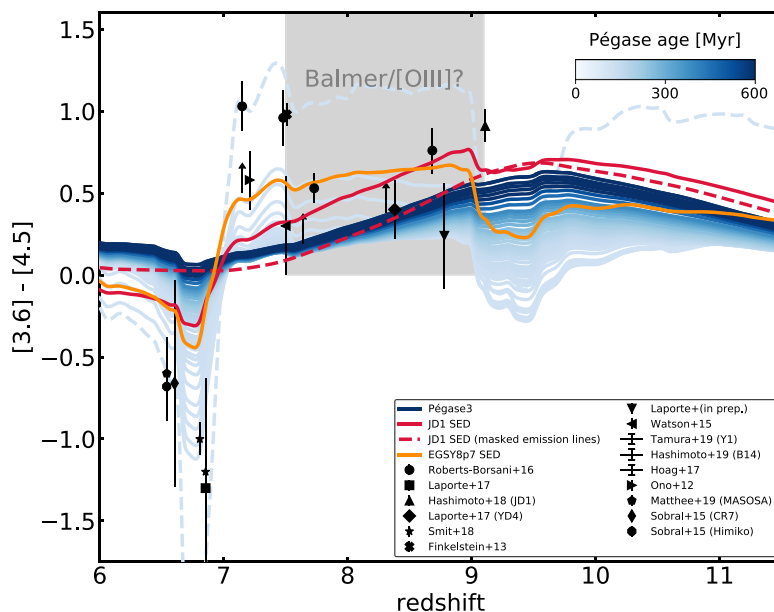
Figure 3. The *Spitzer*/IRAC [3.6]−[4.5] colour evolution as a function of redshift. The black points represent spectroscopically confirmed galaxies at $z > 7$ with red IRAC colours (see Table 1) with additional data for $z \sim 6.6$–6.8 sources with especially blue IRAC colours ([3.6]−[4.5] < −0.5). The evolution for the synthetic spectra of EGSY8p7 (orange), JD1 (red, solid), JD1 with masked emission lines (red, dashed) can be compared with synthesis models including normal ($L_{neb} = 1$) line emission from Pégase3 (the blue shades, the solid lines) for a variety of galaxy ages, assuming a constant SFR from $z = 15$ to $z = 2$. The dashed blue line illustrates the colour evolution for the youngest Pégase3 spectrum and $L_{neb} = 5$, which is sufficient to reproduce the most extreme colours of spectroscopically confirmed galaxies. The light grey shaded region highlights an approximate redshift interval over which ambiguity exists as to the primary mechanism for red *Spitzer*/IRAC colours.

Table 1. A list of spectroscopically confirmed galaxies at $z > 7$ with red IRAC colours.

ID	z_{spec}	H_{160}	3.6−4.5 μm	Reference
MACS1149_JD1	9.11	25.70 ± 0.01	0.91 ± 0.10	Hashimoto et al. (2018)
GN-z10-3	8.78	26.74 ± 0.12	0.24 ± 0.32	Laporte et al. (in prep.)
EGS8p7	8.68	25.26 ± 0.09	0.76 ± 0.14	Zitrin et al. (2015)
A2744_YD4	8.38	26.42 ± 0.04	0.4 ± 0.18	Laporte et al. (2017a)
MACS0416_Y1	8.31	26.04 ± 0.05	>0.38	Tamura et al. (2019)
EGS-zs8-1	7.73	25.03 ± 0.05	0.53 ± 0.09	Oesch et al. (2015)
MACS1423-z7p64	7.64	25.03 ± 0.10	>0.19	Hoag et al. (2017)
z8_GND_5296	7.51	25.55 ± 0.07	0.98 ± 0.07	Finkelstein et al. (2013)
A1689-zD1	7.50	24.70 ± 0.10	0.30 ± 0.30	Watson et al. (2015)
EGS-zs8-2	7.48	25.12 ± 0.05	0.96 ± 0.17	Roberts-Borsani et al. (2016), Stark et al. (2017)
GN-108036	7.21	25.17 ± 0.07	0.58 ± 0.18	Ono et al. (2012)
COSY	7.15	25.06 ± 0.06	1.03 ± 0.06	Stark et al. (2017)
B14-65666	7.15	$24.60^{+0.3}_{-0.2}$	>0.5	Hashimoto et al. (2019)

with and without emission lines), where the red colour is primarily due to a Balmer break.

Considering next how the templates and synthesis models match the IRAC colours of 13 spectroscopically confirmed $z > 7$ galaxies, we can see that line emission in both Pégase3 and the EGSY8p7 (line emitting) template are required to explain the strong dip in [3.6]−[4.5] colour at $z \simeq 6.6$–6.8 as indicated by Smit et al. (2018); as expected the JD1 template with masked emission lines has no dip. However, at higher redshift, where the [O III]+H β lines enter the 4.5 μm band, the red colour is initially more easily reproduced by cases with strong line emission. For SEDs with a flat continuum, as assumed by Roberts-Borsani et al. (2016), such a colour remains relatively constant until the lines leave the band at $z = 9$. However, in the case of a moderate to strong Balmer break and reduced (but not absent) line emission, the IRAC excess *increases* from $z \gtrsim 7.5$

onwards as the Balmer break moves redwards, thereby removing flux from the 3.6 μm band whilst simultaneously providing a relatively constant amount of flux in the 4.5 μm band. In the case where the rest-frame optical continuum is not flat, this impacts the *slope* of the [3.6]−[4.5] colour evolution. The impact of the Balmer break is particularly evident when comparing the evolution of the masked JD1 spectrum to the older aged Pégase3 synthesis models that, despite including (normal) emission lines, display virtually identical [3.6]−[4.5] evolution.

Examining the actual data, one can reasonably securely conclude that the large IRAC excesses at $7 \lesssim z \lesssim 7.5$ (e.g. Ono et al. 2012; Finkelstein et al. 2013; Roberts-Borsani et al. 2016; Hashimoto et al. 2019) are difficult to reproduce without an extreme [O III]+H β contribution, as is clearly the case for IRAC colours for the sources at $z \simeq 6.8$ studied by e.g. Sobral et al. (2015), Smit et al. (2018), Laporte

Table 2. A list of $z > 7$ galaxies from Table 1 with [O III] $\lambda88\,\mu m$ detections and dust continuum constraints (detections and non-detections) from ALMA Band 7 observations. The columns represent, from left to right: the redshift of the galaxy, the assumed magnification factor, the observed wavelength of the [O III] $\lambda88\,\mu m$ detection, its integrated flux, and the measured FWHM. The last two columns are the central observed wavelength of the dust continuum observations and the associated flux. All error bars and upper limits quoted here are 2σ.

ID	z_{spec}	μ	$\lambda_{[O\,III]}$ (μm)	[O III] Integrated flux (Jy km s^{-1})	[O III] FWHM (km s^{-1})	λ_{cont} (μm)	Continuum flux (μJy)
MACS1149_JD1	9.11	10	893.86	0.229 ± 0.050	154 ± 39	917.63	<35.4
A2744_YD4	8.38	2	829.55	0.030 ± 0.008	49.8 ± 4.2	842.70	99.0 ± 23.0
MACS0416_Y1	8.31	1.43	823.32	0.660 ± 0.160	141 ± 21	850.57	137.0 ± 26.0
B14-65666	7.15	–	720.78	1.500 ± 0.180	429 ± 37	733.68	470.0 ± 128.0

et al. (2017b), and Matthee et al. (2019). However, for the sources at $7.5 \lesssim z \leq 9$ (e.g. Watson et al. 2015; Hoag et al. 2017; Laporte et al. 2017a; Tamura et al. 2019 plus GN-z10-3 and EGSY8p7), the paucity of spectroscopic data makes it premature to conclude that the IRAC excess in galaxies beyond $z \simeq 7$ arises entirely from line emission.

3 DISTINGUISHING BETWEEN A BALMER BREAK AND INTENSE LINE EMISSION

3.1 ALMA data and the origin of the [O III] line ratio

In Section 2, we have shown that red *Spitzer*/IRAC [3.6]−[4.5] colours for galaxies lying between $7 < z < 9$ could arise from contributions of both intense nebular line emission and starlight. We now consider whether it is possible to break this degeneracy prior to the use of spectroscopy with JWST. ALMA observations with Band 7 targeting the [O III] $\lambda88\,\mu m$ line and dust continuum may provide a potential way forward. [O III] $\lambda5007\,\text{Å}$ and [O III] $\lambda88\,\mu m$ emission originate from the same star-forming regions and species, and thus constraints from the [O III] $\lambda88\,\mu m$ line should provide valuable limits to the strength of the [O III] $\lambda5007\,\text{Å}$ emission and therefore *Spitzer*/IRAC [4.5] flux contributions. Furthermore, while the precise origin and nature of dust at high-z remains an open debate, constraints from continuum emission can place additional (indirect) constraints on the rate of supernovae explosions and and thus the underlying stellar populations. Using detailed SED modelling with synthetic spectra from Pégase3, we now investigate whether ALMA observations can place constraints on the contribution from young stars via [O III] $\lambda88\,\mu m$ emission and mature stellar populations from the presence of a dust continuum.

Currently, four of the 13 $z > 7$ spectroscopically confirmed galaxies with *Spitzer*/IRAC excesses listed in Table 1 have the appropriate ALMA data: JD1, A2744_YD4, MACS0416_Y1 and B14-65666 (henceforth YD4, Y1, and B14 for convenience), see Table 2. To determine accurate SEDs, for each of the aforementioned galaxies we use the relevant references in Table 1 (and references therein) to compile *HST* ACS+WFC3/IR, (B_{435}, V_{606}, I_{814}, Y_{105}, J_{125}, JH_{140}, and H_{160} bands), VLT/HAWK-I K_s, as well as *Spitzer*/IRAC [3.6] and [4.5] photometry. For B14 we use near-infrared z, Y, J, and H photometry from VISTA, in addition to data in the VLT/HAWK-I K_s and *Spitzer*/IRAC [3.6] and [4.5] bands. All *HST* upper limits and error bars represent 1σ uncertainties, whilst those redwards of these are 2σ.

To evaluate the relative contributions of nebular emission lines and starlight, we create a repertoire of Pégase3 spectra with which to fit the above data for the four spectroscopically confirmed $z > 7$ galaxies. We generate mass-normalized galaxy spectra for a young

component dominated by a recent burst of constant star formation with duration $\tau_{young} = 10\,\text{Myr}$, and for a component with a less recent phase of constant star formation for a range of durations $\tau_{old} = [10, 50, 100, 200, 300, 400, 500]\,\text{Myr}$, where a Balmer break is allowed to form. We then extract spectra at 1 Myr intervals for the young component, and 20 Myr intervals from ages of 1 Myr to the age of the Universe at the redshift of each galaxy for the older component. For simplicity, we assume emission lines arise from the young component only since these come from star-forming regions and are not seen in mature stellar populations. These models are used, sometimes in combination (i.e. recent burst + earlier star formation), with a custom SED-fitting code in a Bayesian framework, to maximize the log-likelihood of the model given the data, including an analytical treatment of upper limits (Sawicki 2012). The free parameters of the code are the mass of the galaxy system, M_{sys}[1] (one for each stellar component) and a multiplicative factor, L_{neb}, to scale the luminosity of the nebular emission lines, whose FWHM are fixed to that of the [O III] $\lambda88\,\mu m$ line presented in Table 2. The adopted priors are log $M_{sys} = [5,15]\,M_\odot$ and $L_{neb} = [0,50]$, allowing for both normal and extreme nebular emission contributions.

Throughout the subsequent analysis, we make the assumption that the [O III] $\lambda5007\,\text{Å}$ line can be constrained from the flux of the [O III] $\lambda88\,\mu m$ line, since these originate from the same star-forming regions. However, the ratio of the two lines is highly dependent on the conditions of the surrounding gas (e.g. electron density and temperature, gas metallicity, the emission rate of ionizing photons from the ionizing star), given that the two species are characterized by different excitation energies and critical densities. We therefore provide a short description of Pégase3's nebular emission treatment (for full details we refer the reader to Fioc & Rocca-Volmerange 2019) and consider the [O III] ratios probed by our young starburst component to ensure they are suitable for a meaningful analysis of the emission line contributions to the *Spitzer*/IRAC channels.

Pégase3 determines the integrated luminosities of a large suite of nebular emission lines by using a pre-computed luminosity grid from `Cloudy` (Ferland et al. 2017) and linking it to the stellar and ISM evolution of the galaxy model. The large emission grid was constructed as a function of ISM metallicity and emission rate of Lyman continuum photons (with ranges {0–0.1} and {10^{46}–10^{53}} s^{-1}, respectively) for spherically symmetric, radiation bound H II regions filled with a constant gas density of $n_H = 10^2\,\text{cm}^{-3}$. Each of the `Cloudy` simulations were run out to a radial distance from the central ionizing star where the free proton density drops below $10^{-2}\,\text{cm}^{-3}$. For each star cluster in a given simple stellar population with age t, Pégase3 computes the ionizing output and gas-

---

[1]For a proper definition of how this translates in the Pégase3 formalism to stellar mass, see Fioc & Rocca-Volmerange 2019.
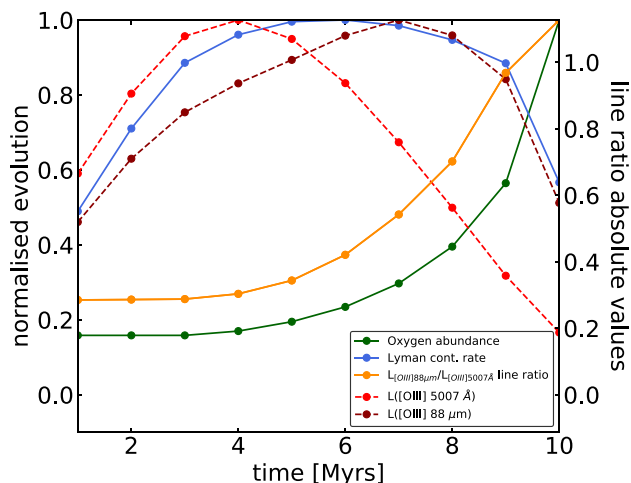
Figure 4. The evolution of the [O III] $\lambda 88$ μm / [O III] $\lambda 5007$ Å line ratio as a function of time for our young burst model, with associated galaxy properties (Oxygen abundance, Lyman continuum emission rate, and the luminosities of both lines). Each quantity is normalized by its maximum value and a clear trend is seen between the line ratio and the galaxy metallicity, due to the cooling effects and the large difference in excitation potential between the two lines. Our low ratios (i) are tied to the underlying stellar and gas evolution of the starburst model and (ii) probe a suitable range of ratios that favour intense [O III] $\lambda 5007$ Å nebular emission capable of producing up to ~ 1.4 mag *Spitzer*/IRAC [3.6]−[4.5] colours.

phase metallicity prior to determining the integrated line luminosity through interpolatation of the Cloudy grid. The emission lines from Pégase3 are thus computed in a self-consistent manner and directly linked to the underlying stellar and ISM evolution. As such, no prior assumptions on either the electron density or temperature need be supplied.

With the above in mind, we plot in Fig. 4 the normalized line ratio as a function of time for the young starburst model, in addition to several relevant galaxy quantities. Our [O III] $\lambda 88$ μm / [O III] $\lambda 5007$ Å values span a reasonably large range of $\{0.28-1.13\}$. Using rest-frame optical line ratios from our spectra as electron temperature (T_e) and density (n_e) diagnostics (Proxauf, Öttl & Kimeswenger 2014), we find our nebular emission also probe relatively low values of $T_e \sim 8500-11\,000$ K and $n_e \lesssim 10^2$ cm^{-3}, respectively. Given such values remain well below the critical densities of the two [O III] lines (510 cm^{-3} and 6.8×10^5 cm^{-3} for the 88 μm and 5007 Å line, respectively; Osterbrock & Ferland 2006), collisional de-excitation is unlikely to play a major role in regulating the luminosities of the lines. Instead, the large difference between the excitation potentials of the two lines (0.014 and 2.476 eV for the 88 μm and 5007 Å line, respectively) suggests temperature fluctuations primarily due to the gas-phase metallicity acts as the primary regulator of the ratio, as is clearly illustrated in the figure. At very low metallicities, the increase in Lyman continuum photon production rate coincides with a rise in temperature and thus the number of collisions with energetic electrons and population of the [O III] $\lambda 5007$ Å upper energy level. Given the low excitation potential of the [O III] $\lambda 88$ μm line, however, its luminosity remains largely unaffected and thus does not vary significantly, maintaining a low line ratio. Once the production of metals becomes significant, however, cooling effects begin to take place and the luminosity of the [O III] $\lambda 5007$ Å line drops dramatically, thereby rapidly increasing the line ratio towards unity.

Prior to the launch of JWST, observations of the [O III] $\lambda 5007$ Å line at $z > 6$ are not possible, while the number of low-z analogues with matched observations of both lines is limited. Since metallicity plays a key role in regulating the evolution of the line ratio, however, we compare the Oxygen abundances for each of our young starburst spectra (and therefore line ratios) to those of the local, low-metallicity dwarf galaxies in the Dwarf Galaxy Survey (DGS; Madden et al. 2013). To be consistent with the metallicity derivations of the DGS, we use the 'R23' ([O II] $\lambda\lambda 3727,3729$ + [O III] $\lambda\lambda 4959,5007$)/H β parametrizations by Pilyugin & Thuan (2005) with a log([N II] $\lambda 6584$/[O II] $\lambda\lambda 3727,3729$) ~ -1.2 limit to break the R23 degeneracy (Kewley & Ellison 2008). Our derived values range from $12+$log (O/H) $\approx 8.40-8.67$ (0.50–0.93 Z$_\odot$), characteristic of sub-solar abundances and consistent with the high end of the DGS metallicities.

Finally, we find our range of line ratios more than capable of producing extreme rest-frame optical emission, with $z > 7$ *Spitzer*/IRAC colours up to [3.6]−[4.5] ~ 1.4 mag. This is consistent with all of the red colours of $z > 7$ spectroscopically confirmed galaxies in Fig. 3 and much larger than those reported for the four galaxies selected above. Should our ratios have probed larger values only, characteristic of other areas in a simple T_e–n_e grid, subsequent analyses would likely be biased against strong nebular emission. Thus, given the consistency of our line ratios with expected trends and data for low-metallicity systems, as well as the large *Spitzer*/IRAC excess allowed by our models, we deem the nebular emission generated here well suited for the purposes of this paper.

To illustrate how the ALMA observations may differentiate between intense nebular emission and starlight in explaining the IRAC colours, we consider the case of YD4 since, for this source, all photometric points redwards of the Lyman break and the ALMA spectroscopic constraints are robustly measured. First, we fit the data with a single-component young model (permitting a dust contribution), once with *HST*+VLT/HAWK-I+*Spitzer*/IRAC data only and then again incorporating the ALMA Band 7 constraints. The best-fitting SEDs are shown in Fig. 5. The continuum fits to the *HST* and HAWK-I photometry are satisfactory and comparable in each case. However, there is a major difference in the predicted *Spitzer*/IRAC photometry. Ignoring the ALMA constraints, the IRAC excess demands the presence of strong nebular emission lines in which case the [O III] $\lambda 88$ μm line is considerably overpredicted (by a factor $\simeq 10$, not shown) and the continuum dust emission is similarly poorly matched. Additionally, the presence of nebular emission lines – namely the [O II] doublet at a rest frame of ~ 3730 Å – adds non-negligible boosting to the [3.6] band. Including the ALMA constraints, the nebular emission in the IRAC bands is modest, indicating the need for an additional component to fit these data, for example, a Balmer break originating from star formation at earlier times.

3.2 SED-fitting and results

We thus now proceed to fit all of the available data for each of the four galaxies with a two-component model comprising a contribution from young stars with intense nebular emission and an older stellar component. The two-component models are derived from all unique combinations of young and older spectra requiring only that age$_{young}$ < age$_{galaxy}$−τ_{old}, so that the recent burst of star formation from the young component occurs only when star formation in the older component has completed. In these two-component models, for simplicity, we assume the dust contributions arise only from the older component and specifically only if there is a Band 7 continuum
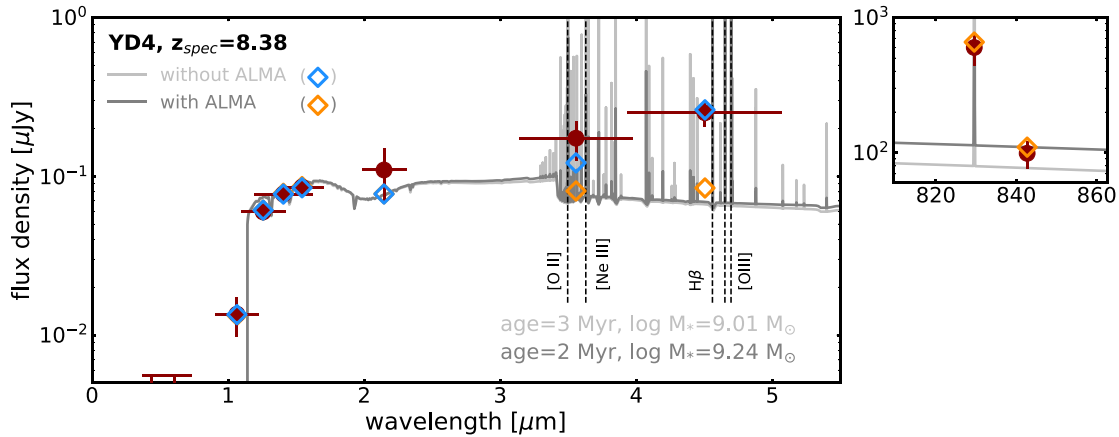
Figure 5. SED fits to the spectroscopically confirmed galaxy, YD4, with *HST*, VLT/HAWK-I K_s and *Spitzer*/IRAC photometry (the red points and the error bars) and single models of dusty 1–10 Myr Pégase3 spectra. The suite of young Pégase3 spectra are fit once without the ALMA Band 7 spectroscopic observations (the light grey line and the blue points) and once with them (the dark grey line and the orange points), both with the strength of nebular emission lines as a free parameter in addition to their stellar masses. The IRAC excess is well fit by contributions from nebular emission lines without inclusion of the ALMA data. However, when such constraints are included, the nebular emission is suppressed and cannot account for the excess flux in the *Spitzer*/IRAC 4.5 μm band. Thus, a secondary component arising from starlight is necessary to match the data.

detection. By comparing these two-component fits to those for a single-component (with dust included following the guidelines indicated above), we can determine, as suggested in the case of YD4, (i) whether the two-component fits are significantly better than the single young component ones and (ii) whether the first indicate the presence of an older, more mature stellar population.

The results of our best-fitting two-component models are presented in Fig. 6, where we find generally good agreement with the observed photometric data sets and ALMA constraints. For two of the four galaxies (JD1 and YD4), the best-fitting model correctly predicts the presence or upper limit of dust mass based on the ALMA continuum detections, whilst simultaneously reproducing the *HST* photometry, *Spitzer*/IRAC excess, and [O III] λ88 μm flux, within the error bars. For Y1 and B14, on the other hand, the *HST* photometry and *Spitzer*/IRAC excess are well reproduced, but the models are unable to simultaneously match both the [O III] λ88 μm flux and dust continuum (in the case of B14, the [O III] λ88 μm flux is matched but the dust continuum is underpredicted, and in the case of Y1 the [O III] λ88 μm flux is underpredicted and dust continuum overpredicted). In comparing these fits to those assuming a single-component only, although the *HST* photometry can be reasonably well reproduced for JD1 and YD4, the single-component fits fail to simultaneously reproduce both the ALMA constraints and the *Spitzer*/IRAC excess. In the cases of Y1 and B14, the ALMA constraints are better matched by the one-component fit, however, the *Spitzer*/IRAC excess is only partially matched.

In our two-component fits, a sizeable contribution to the IRAC flux arises from a more mature stellar component. For JD1, YD4 and Y1, the contribution to the IRAC fluxes from the recent burst of star formation is only ∼10–30 per cent and the older component (characterized by Balmer ratios of ∼2) dominates the flux at ≳70 per cent. This is due primarily to the relative weakness of the [O III]+H β lines, for which we measure an EW, EW([O III]+H β) ≈ 25–106 Å (compared to EW([O III]+H β) ≈ 450–7700 Å for a single-component). The exception to this trend is B14, whose IRAC fluxes remain dominated by the younger component by ∼60–75 per cent (EW([O III]+H β) ≈ 200 Å for the two-component model and

EW([O III]+H β) ≈ 3460 Å for the one-component model). The total stellar mass (corrected for lensing for those lensed sources) for all of these fits range from log $M_* = 8.88$–10.19 $M_\odot$, with virtually all of the total stellar mass also coming from the earlier period of star formation and the most recent burst contributing primarily through the presence of weak-to-moderate nebular emission.

Furthermore, we also note a difference in the ages of the galaxies, as determined by the onset of the earlier burst in the two-component model or the age of the single-component model. We find an increase in age for each of the four galaxies when multiple components are used, with ages of 80, 260, 140, and 20 Myr characterizing the two-component fits of JD1, YD4, Y1, and B14, and such ages decreasing down to 1, 2, 1, and 3 Myr assuming on a single burst of recent star formation. In the case of Y1, the age estimate from the two-component fit is in fact a lower limit, since we only have an upper limit for the *Spitzer*/IRAC 3.6 μm photometry. We note here that our preferred $\tau_{old} = 10$ Myr value and the galaxy age estimate for JD1 are somewhat lower than those estimated by Hashimoto et al. (2018), whose best fit comprises an older stellar population with an episode of star formation lasting $\tau_{old} = 100$ Myr and a galaxy age of 290^{+190}_{-120} Myr. However, as in their analysis, we find considerable statistical similarity with the best two-component fit assuming $\tau_{old} = 100$ Myr, in which case our age estimate increases to 140 Myr, closer to the lower limit of their assumed value.

Finally, to enable a quantitative comparison of the one- and two-component fits we examine the log-likelihoods and find, with the exception of Y1, that the two-component fits of each of the galaxies is considered a better fit. However, given the obvious danger of concluding better fits with an additional component with further free parameters, we compare the goodness-of-fit via a comparison of their Bayesian information criteria (BIC), which uses the log-likelihoods whilst penalizing for additional free parameters. With this consideration, the additional free parameters in the fit to the B14 data are sufficiently penalised to justify only the one-component fit. Whilst such comparisons no doubt gloss over the complexities of defining the birth of such galaxies, they serve as an important illustration of the potential consequences from overlooking the
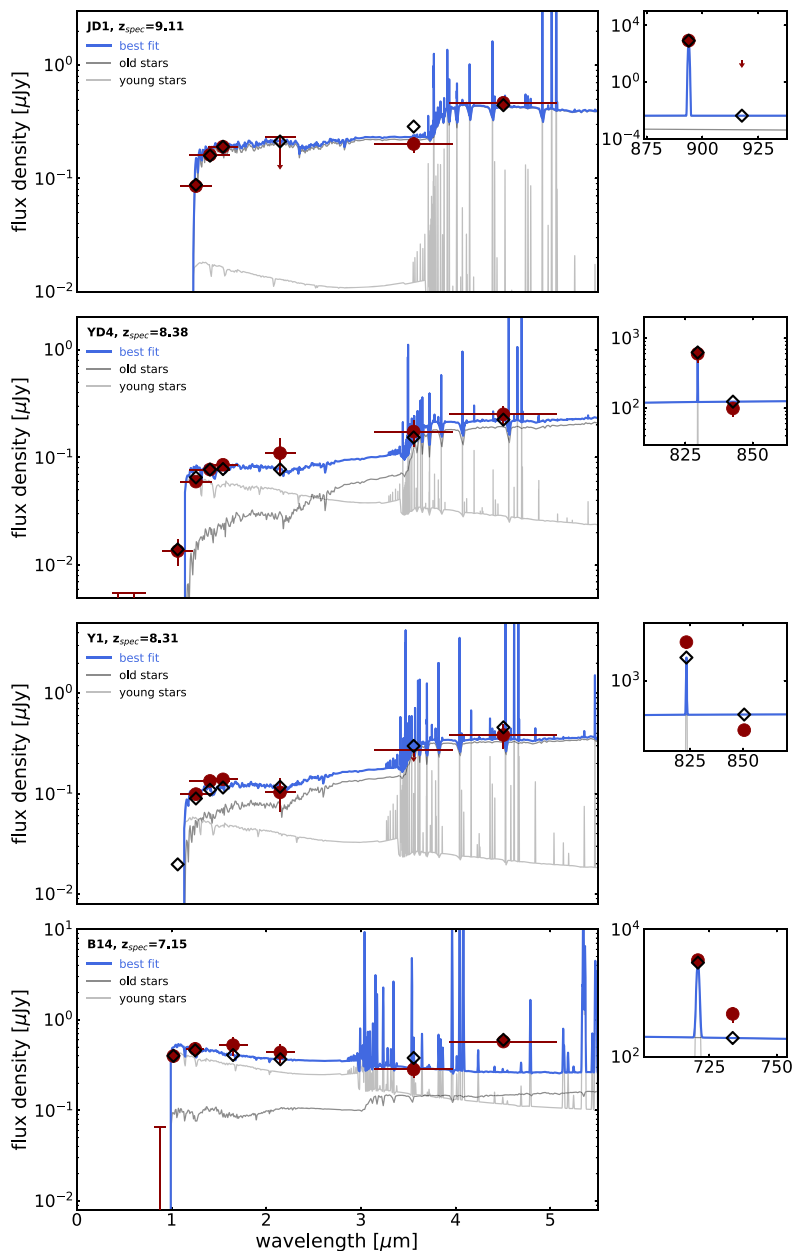
Figure 6. The best-fitting two-component SED models (the blue lines and the black points) to observed data (the dark red circles and the error bars) from *HST*/VISTA + VLT/HAWK-I K_s + *Spitzer*/IRAC photometry (left) and ALMA spectroscopy (upper right inset) for galaxies at $z > 7$ with red *Spitzer*/IRAC colours (JD1, YD4, Y1, and B14). The total fit consists of contributions from both old (darker grey) and young (lighter grey) stars.

consideration of multiple episodes of star formation. We provide a summary of the above comparison and the main properties of our fits in Table 3.

3.3 Possible AGN contributions

Throughout the SED analyses described above, we assumed only thermal contributions to the heating of the metal-rich gas. However, given recent observations suggesting possible active galactic nuclei (AGN) contributions at high redshift (e.g. Tilvi et al. 2016; Laporte et al. 2017b; Mainali et al. 2018), it is natural to consider how our results might change should there be non-thermal radiation components in our four galaxies.

3.3.1 Influence on the [O III] ratio

To quantify differences in the FIR/optical [O III] ratio arising from excitation by stellar or AGN contributions, we run Cloudy using v2.1 of the Binary Population and Spectral Synthesis (BPASS; Eldridge et al. 2017) models as well as the code's own AGN option (based on the model of Mathews & Ferland 1987). For both stellar and AGN cases, the [O III] line ratio is computed over a range of electron densities n_e = [3, 10, 100, 1000] (cm^{-3}) and gas-phase metallicities Z = [0.05, 0.2, 0.4, 1] ($Z_\odot$) for a fixed ionization parameter log U = -2.5. We show the results of these simulations in Fig. 7, which demonstrate the AGN case makes little difference at fixed electron density and gas-phase metallicity (a result, we note, is unaffected by a change in the ionization parameter). The similarity between the

Table 3. Summary of the main properties and parameters (uncorrected for any lensing of the objects) of the favoured one- and two-component SED fits determined by our SED fitting code with Pégase3 spectra. The tabulated flux percentages for the two-component models in the middle section of the table are the relative contributions from the (young, old) components to the total flux measured in that band.

|	JD1		YD4		Y1		B14	
	1-comp	2-comp	1-comp	2-comp	1-comp	2-comp	1-comp	2-comp
τ_{old} (Myr)	–	10	–	200	–	100	–	10
Galaxy age (Myr)	1	80	2	260	1	140	3	20
$M_{*, young}/M_{*, old}$	–	2×10^{-3}	–	3.2×10^{-3}	–	1.9×10^{-3}	–	8×10^{-2}
log $M_{*, total}$ ($M_\odot$)	8.23 ± 0.01	9.88 ± 0.01	8.71 ± 0.02	10.26 ± 0.06	8.93 ± 0.01	10.35 ± 0.03	9.25 ± 0.04	9.49 ± 0.99
EW([O III]+Hβ) (Å)	7668 ± 793	106 ± 43	448 ± 133	25 ± 9	7307 ± 1211	51 ± 11	3462 ± 720	198 ± 260
D_n4000	0.55 ± 0.01	1.94 ± 0.02	0.77 ± 0.00	2.09 ± 0.07	0.81 ± 0.01	1.99 ± 0.02	0.82 ± 0.01	0.97 ± 0.35
3.6 μm flux contribution	–	(9.8, 90.2) per cent	–	(25.4, 74.6) per cent	–	(23.9, 76.1) per cent	–	(62.4, 37.6) per cent
4.5 μm flux contribution	–	(8.2, 91.8) per cent	–	(18.4, 81.6) per cent	–	(30.8, 69.2) per cent	–	(75.5, 24.5) per cent
log-likelihood	−188.81	−12.22	−9.44	−3.02	−62.47	−66.92	−6.17	−4.86
BIC	403.21	62.84	44.02	43.76	150.33	171.94	38.10	48.37
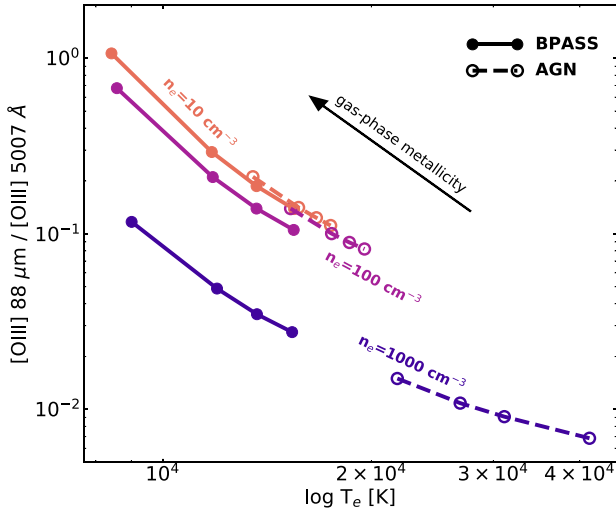
Figure 7. The [O III] 88 μm/[O III] 5007 Å line ratio computed over a grid of electron densities and gas-phase metallicities for a stellar radiation field using BPASS (the filled points and the solid lines) and an AGN (the empty points and the dashed lines) models using Cloudy. The curves and points are colour-coded according to their constant electron density. For clarity, the case with $n_e = 3 \, cm^{-3}$ is not shown. From left to right, the first, second, third, and fourth scatter points in each of the curves represent metallicites of 1, 0.4, 0.2, and 0.05 $Z_\odot$. AGN contributions do not significantly alter the allowed [O III] line ratio for a given electron density and metallicity, except at very high densities and temperatures.

stellar and AGN results is particularly evident for electron densities $n_e \lesssim 1000 \, cm^{-3}$, where relatively large ratios ([O III] 88 μm / [O III] 5007 Å $\approx$ 0.1–1) are seen that increase with metallicity. Such a trend is consistent with the results in Section 3.1 and Fig. 4 and arises because of relatively low electron temperatures ($T_e \lesssim 15\,000$–$20\,000$ K) where, for low-density regimes, the line ratio is influenced primarily by temperature fluctuations rather than by collisional de-excitation. As the electron density increases to $n_e \gtrsim 1000 \, cm^{-3}$, however, order-of-magnitude differences appear: while both the stellar and AGN cases probe line ratios <0.1 favouring particularly strong [O III] 5007 Å emission, collisional de-excitation of the [O III] 88 μm line can occur and, combined with the higher temperatures

($T_e > 20\,000$ K) possible in the AGN case, this increases excitation of [O III] 5007 Å whereas [O III] 88 μm is quenched. As a result, the ratio is driven down to low values of ∼0.01. For our sample of four sources, the measured electron densities and temperatures (see Section 3.1) probe regions of the parameter space where AGN do not have a significant effect and thus our [O III] line ratios are unlikely to be strong affected by non-thermal radiation.

3.3.2 Observational evidence for AGN at $z > 7$?

Although we have found that the [O III] line ratio is only affected by an AGN contribution in regions of very high temperature and electron densities unlikely to be representative of the general $z > 7$ population, it is none the less informative to consider whether our sample of four galaxies and those in the extended list of IRAC-excess sources in Table 1 show any observational evidence for AGN activity, and if so at what level. At $z > 7$, the most useful diagnostics for distinguishing between thermal and non-thermal contributions are rest-frame UV emission lines and their ratios, in particular N V $\lambda1240,1243$ Å, C IV $\lambda1549$ Å, He II $\lambda1640$ Å, and C III] $\lambda\lambda1907,1909$ Å (Feltre, Charlot & Gutkin 2016; Laporte et al. 2017b; Stark et al. 2017; Mainali et al. 2018). With an ionizing energy of nearly 80 eV, the detection of N V alone is a strong indicator of non-thermal radiation (Feltre et al. 2016). Each of our four *Spitzer*/IRAC- and ALMA-selected objects have been targeted with near-infrared (NIR) spectroscopy where the aforementioned emission lines reside at $z > 7$. 7.5–10 h of VLT/X-Shooter observations revealed only Ly α in JD1 (Hashimoto et al. 2018) and YD4 (Laporte et al. 2017a) and no rest-frame UV lines were seen in Y1 (Tamura et al. 2019). Furthermore, 4 h of Subaru/FOCAS observations revealed only Ly α in B14 (Furusawa et al. 2016). Of the remaining sources compiled in Table 1, all but GN-z10-3, MACS1423-z7p64, and GN-108036 have relevant NIR observations, albeit at varying depths. Only four of these reveal one or more detections of the relevant lines: EGSY8p7 (N V $\lambda1243$ Å; Mainali et al. 2018), EGS-zs8-1 (both components of the C III] $\lambda\lambda1907,1909$ Å doublet; Stark et al. 2017), z8_GND_5296 (C III] $\lambda1907$ Å; Hutchison et al. 2019), and COSY (N V $\lambda1240$ Å and He II $\lambda1640$ Å; Laporte et al. 2017b). Given their large ionization potentials, the detection of any of these lines is indicative of extreme radiation fields (Feltre et al. 2016; Stark et al. 2017). However, to place quantitative constraints on the powering mechanism, accurate measures of both C III] $\lambda1909$ Å and He II $\lambda1640$ Å lines are neces-
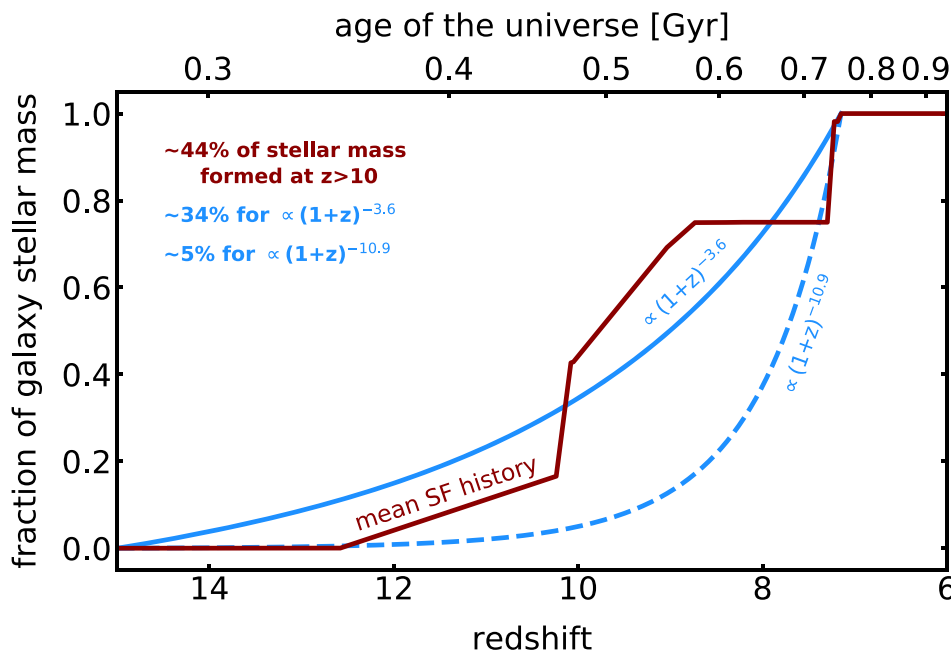
Figure 8. The fractional stellar mass assembly history averaged over the four galaxies in Table 2 (dark red), adopting the best two-component fits in Fig. 6. The solid and dashed blue lines represent the equivalent fractional histories for the two cosmic SFR density relations presented in fig. 9 of Oesch et al. (2014).

sary. As shown in fig. 7 of Laporte et al. (2017b), while the ratio N V $\lambda 1240\,\text{Å}$/He II $\lambda 1640\,\text{Å}$ constrains the strength of the radiation field, only the C III] $\lambda 1909\,\text{Å}$/He II $\lambda 1640\,\text{Å}$ ratio can robustly distinguish a metal-poor stellar radiation field from one involving an AGN. With the available spectroscopic data, even upper flux limits on these lines do not usefully constrain the nature of the radiation field. Thus, while it is not yet possible to rule out a modest non-thermal contribution, specifically for the four ALMA sources under consideration here, given the absence of any detected high ionization emission lines, it seems reasonable to conclude any AGN contribution cannot be dominant.

4 DISCUSSION

We have shown that a Balmer break, arising from a mature stellar population, may be a significant contributor to the IRAC excess seen in spectroscopically confirmed $z > 7$ star-forming galaxies. While our analysis does not rule out the possibility that much of this excess arises, as has been conventionally assumed, from intense [O III] emission, using ALMA [O III] $\lambda 88\,\mu\text{m}$ emission and dust mass measures, we have examined whether we can constrain the relative contributions of starlight and line emission.

The distinction between intense line emission, attributed to recent episodes of star formation from a young ($\leq 10\,\text{Myr}$) stellar population, and a prominent Balmer break consistent with more mature stars, is important in considerations of the early assembly history of galaxies. Both the stellar masses and earlier star formation histories will differ depending on the relative contributions and this, in turn, will affect the inferred star formation activity beyond the current *HST* redshift horizon of $z \simeq 10$. This was first demonstrated for the $z = 9.11$ galaxy JD1 by Hashimoto et al. (2018), where the IRAC excess must arise primarily from starlight, leading to a stellar mass of $4.2 \pm 1.0 \times 10^9\,\text{M}_\odot$ (lens-corrected for the

preferred gravitational magnification) only $\simeq 550\,\text{Myr}$ after the big bang with an implied epoch of first star formation as early as $z \sim 15$.

As an illustration, if we adopt the significant contribution to the IRAC excess from starlight for those $z > 7$ sources in Table 2 for which we fit two-components, their stellar masses increase by an average factor of $\simeq 30$ compared to contributions from young starlight alone. Although a single-component fit is probably an extreme comparison in this context, none the less our two-component fits with an earlier period of star formation must imply an assembly history beyond $z \simeq 10$, as discussed by Hashimoto et al. (2018), with interesting consequences for the interpretation of 21 cm experiments (Bowman et al. 2018) and the timing of 'cosmic dawn'. In Fig. 8, we plot the fractional stellar mass assembly history, averaged over our four galaxies, up to the epoch of observation in which it can be seen that ~ 44 per cent of the stellar mass was produced before a redshift $z \simeq 10$. Although clearly a modest sample restricted largely to the brightest studied sources at $z > 7$, we can compare this fractional mass assembly history with the prediction of two star formation histories for galaxies with SFRs $> 0.7\,\text{M}_\odot\,\text{yr}^{-1}$ discussed by Oesch et al. (2014) similarly normalized at the mean redshift of our galaxies. Whilst the uncertainties are large due to small number statistics, if our galaxies are representative this would indicate a more gradual decline in the star formation history beyond $z \simeq 8$ than Oesch et al. (2014) prefer (see McLeod et al. 2016). Whilst clearly a simplistic comparison, it serves to emphasize the importance of determining the true origin of the IRAC excess in $z > 7$ galaxies. Ultimately NIRSpec on JWST will be well placed to resolve the ambiguities explored in this paper via direct spectroscopy of a large sample of $7 < z < 9$ galaxies securing not only the strength of rest-frame optical lines such as [O III] $\lambda 5007\,\text{Å}$ but also absorption line measures such as H δ, which is a further indicator of stellar ages.

ACKNOWLEDGEMENTS

DATA AVAILABILITY

The data presented in this paper are available upon email request to the first author.

REFERENCES

Atek H. et al., 2015, ApJ, 800, 18
Bingelli C. et al., 2019, MNRAS, 489, 3827
Bouwens R. J. et al., 2011, ApJ, 737, 90
Bouwens R. J. et al., 2015, ApJ, 803, 34
Bowman J. D. et al., 2018, Nature, 555, 67
Bradley L. D. et al., 2014, ApJ, 792, 76
Chabrier G., 2003, PASP, 115, 763
Coe D. et al., 2019, ApJ, 884, 85
Eldridge J. J. et al., 2017, Publ. Astron. Soc. Aust., 34, e058
Ellis R. S. et al., 2013, ApJ, 763, L7
Feltre A., Charlot S., Gutkin J., 2016, MNRAS, 456, 3354
Ferland G. J. et al., 2017, Rev. Mex. Astron. Astrofís., 53, 385
Finkelstein S. et al., 2013, Nature, 502, 524
Fioc M., Rocca-Volmerange B., 2019, A&A, 623, A143
Furusawa H. et al., 2016, ApJ, 822, 46
Grogin N. A. et al., 2011, ApJS, 197, 35
Hashimoto T. et al., 2018, Nature, 557, 392
Hashimoto T. et al., 2019, PASJ, 71, 71
Hoag A. et al., 2017, Nat. Astron., 1, 0091
Hutchison T. A. et al., 2019, ApJ, 879, 70
Katz H. et al., 2019, MNRAS, 484, 4054
Kewley L. J., Ellison S. L., 2008, ApJ, 681, 1183
Koekemoer A. M. et al., 2011, ApJS, 197, 36
Labbé I. et al., 2013, ApJ, 777, L19
Laporte N. et al., 2017a, ApJ, 837, L21
Laporte N. et al., 2017b, ApJ, 851, 40
Lotz J. M. et al., 2017, ApJ, 837, 97
Madden S. C. et al., 2013, PASP, 125, 600
Mainali R. et al., 2018, MNRAS, 479, 1180
Maseda M. V. et al., 2014, ApJ, 791, 17
Mathews W. G., Ferland G. J., 1987, ApJ, 323, 456
Matthee J. et al., 2017, ApJ, 851, 145
Matthee J. et al., 2019, ApJ, 881, 124
McLeod D. J., McLure R. J., Dunlop J. S., 2016, MNRAS, 459, 3812
Oesch P. A. et al., 2012, ApJ, 759, 135
Oesch P. A. et al., 2014, ApJ, 786, 108
Oesch P. A. et al., 2015, ApJ, 804, L30
Oke J. B., Gunn J. E., 1983, ApJ, 266, 713
Ono Y. et al., 2012, ApJ, 744, 83
Osterbrock D. E., Ferland G. J., 2006, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei, University Science Books, Sausalito, CA
Pentericci L. et al., 2016, ApJ, 829, L11
Pilyugin L. S., Thuan T. X., 2005, ApJ, 631, 231
Proxauf B., Öttl S., Kimeswenger S., 2014, A&A, 561, A10
Roberts-Borsani G. W. et al., 2016, ApJ, 823, 143
Salmon B. et al., 2018, ApJ, 864, L22
Sawicki M., 2012, PASP, 124, 1208
Smit R. et al., 2015, ApJ, 801, 122
Smit R. et al., 2018, Nature, 553, 178
Sobral D. et al., 2015, ApJ, 808, 139
Stark D. P., 2016, ARA&A, 54, 761
Stark D. P. et al., 2017, MNRAS, 464, 469
Tamura Y. et al., 2019, ApJ, 874, 27
Tilvi V. et al., 2016, ApJ, 827, L14
Watson D., Christensen L., Knudsen K. K., Johan R., Anna G., Michał Jerzy M., 2015, Nature, 519, 327
Zheng W. et al., 2017, ApJ, 836, 210
Zitrin A., et al., 2015, ApJ, 810, L12

This paper has been typeset from a TEX/LATEX file prepared by the author.