Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **468,** 3071–3108 (2017)
Advance Access publication 2017 March 14

doi:10.1093/mnras/stx605

ZFIRE: using Hα equivalent widths to investigate the *in situ* initial mass function at $z \sim 2$

Themiya Nanayakkara,[1⋆] Karl Glazebrook,[1⋆] Glenn G. Kacprzak,[1] Tiantian Yuan,[2] David Fisher,[1] Kim-Vy Tran,[3] Lisa J. Kewley,[2] Lee Spitler,[4,5] Leo Alcorn,[3] Michael Cowley,[4,5] Ivo Labbe,[6] Caroline Straatman[6,7] and Adam Tomczak[8]

[1]*Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, VIC 3122, Australia*
[2]*Research School of Astronomy and Astrophysics, the Australian National University, Cotter Road, Weston Creek, ACT 2611, Australia*
[3]*Department of Physics and Astronomy, George P. and Cynthia W. Mitchell Institute for Fundamental Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA*
[4]*Department of Physics and Astronomy, Macquarie University, Sydney, NSW 2109, Australia*
[5]*Australian Astronomical Observatory, PO Box 915, North Ryde, NSW 1670, Australia*
[6]*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*
[7]*Max-Planck Institute for Astronomy, Königstuh 17, D-69117 Heidelberg, Germany*
[8]*Department of Physics, University of California Davis, One Shields Avenue, Davis, CA 95616, USA*

Accepted 2017 March 9. Received 2017 February 6; in original form 2016 October 30

ABSTRACT

We use the ZFIRE (http://zfire.swinburne.edu.au) survey to investigate the high-mass slope of the initial mass function (IMF) for a mass-complete ($\log_{10}(M_*/\mathrm{M}_\odot) \sim 9.3$) sample of 102 star-forming galaxies at $z \sim 2$ using their Hα equivalent widths (Hα EWs) and rest-frame optical colours. We compare dust-corrected Hα EW distributions with predictions of star formation histories (SFHs) from PEGASE.2 and STARBURST99 synthetic stellar population models. We find an excess of high Hα EW galaxies that are up to 0.3–0.5 dex above the model-predicted Salpeter IMF locus and the Hα EW distribution is much broader (10–500 Å) than can easily be explained by a simple monotonic SFH with a standard Salpeter-slope IMF. Though this discrepancy is somewhat alleviated when it is assumed that there is no relative attenuation difference between stars and nebular lines, the result is robust against observational biases, and no single IMF (i.e. non-Salpeter slope) can reproduce the data. We show using both spectral stacking and Monte Carlo simulations that starbursts cannot explain the EW distribution. We investigate other physical mechanisms including models with variations in stellar rotation, binary star evolution, metallicity and the IMF upper-mass cut-off. IMF variations and/or highly rotating extreme metal-poor stars ($Z \sim 0.1\ \mathrm{Z}_\odot$) with binary interactions are the most plausible explanations for our data. If the IMF varies, then the highest Hα EWs would require very shallow slopes ($\Gamma > -1.0$) with no one slope able to reproduce the data. Thus, the IMF would have to vary stochastically. We conclude that the stellar populations at $z \gtrsim 2$ show distinct differences from local populations and there is no simple physical model to explain the large variation in Hα EWs at $z \sim 2$.

Key words: dust, extinction – galaxies: abundances – galaxies: fundamental parameters – galaxies: high-redshift – galaxies: star formation.

1 INTRODUCTION

The initial mass function (IMF) or the mass distribution of stars formed in a volume of space at a given time is one of the most fundamental empirically derived relations in astrophysics (Salpeter 1955; Miller & Scalo 1979; Kennicutt 1983; Scalo 1986a; Kroupa 2001; Baldry & Glazebrook 2003; Chabrier 2003). Since the mass of a star is the primary factor that governs its evolutionary path, the collective evolution of a galaxy is driven strongly by its distribution of stellar masses (Bastian, Covey & Meyer 2010). Therefore, understanding and quantifying the IMF is of paramount importance since it affects galactic star formation rates (SFRs), galactic chemical evolution, formation and evolution of stellar clusters, stellar remnant populations, galactic supernova rates, the energetics and phase

⋆ E-mail: themiyananayakkara@gmail.com (TN); kglazebrook@swin.edu.au (KG)

balance of the interstellar medium (ISM), mass-to-light ratios, galactic dark matter content and how we model galaxy formation and evolution (Kennicutt 1998a; Hoversten 2007).

The IMF of stellar populations (SPs) can be investigated either via direct (studies that count individual stars to infer stellar ages to compute an IMF, e.g. Bruzzese et al. 2015) or indirect methods (studies that model the integrated light from SPs to infer an IMF, e.g. Baldry & Glazebrook 2003). Due to current observational constraints, the number of extragalactic IMF measurements that utilizes direct measures of the IMF is limited (Leitherer 1998). Therefore, most studies employ indirect measures of the IMF, which are affected by numerous systematic uncertainties and limitations.

Indirect IMF measures can be insensitive to low-mass SPs since bright O, B and red supergiant stars may outshine low-mass stars. In contrast at the highest mass end, there can be an insufficient number massive stars to make a significant contribution to the detected light (Leitherer 1998; Hoversten & Glazebrook 2008). In addition, degeneracies in SP models play a significant role in the uncertainties of the derived IMFs, especially since stellar age, stellar metallicity, galactic dust, galactic star formation history (SFH) and stellar IMF cannot be easily disentangled (Conroy 2013). Furthermore, indirect IMF results may depend strongly on more sophisticated features of SP models [mainly stellar rotation, binary evolution of O and B stars and the treatment of Wolf–Rayet (W-R) stars (Murdin 2000)] and dark matter profiles of galaxies. Smith (2014) showed that galaxy by galaxy comparisons of inferred IMF mass factors via dynamical and spectroscopic fitting techniques can lead to inconsistent results due to our limited understanding of element abundance ratios, dark matter contributions and/or more sophisticated shape of the IMFs (McDermid et al. 2014; Smith 2014).

The concept of an IMF was first introduced by Salpeter (1955) as a logarithmic slope Γ defined by

$$\Phi(\log\ m) = \mathrm{d}N/\mathrm{d}(\log\ m) \propto m^{\Gamma}, \quad (1)$$

where m is the mass of a star, N is the number of stars within a logarithmic mass bin and $\Gamma = -1.35$ is the slope of the IMF. Historic studies of the IMF slope at the high-mass end ($M \gtrsim 1$ M$_{\odot}$) showed no statistically significant differences from the value derived by Salpeter giving rise to the concept of IMF universality (Scalo 1986b; Gilmore 2001). Theoretical studies attempt to explain the concept of universal IMF by invoking mechanisms such as fragmentation of molecular clouds (Larson 1973) or feedback from the ISM (Klishin & Chilingarian 2016). However, there is no definitive theoretical model that can predict a given universal IMF from first principles, which limits our theoretical understanding of the fundamental physics that govern the IMF.

1.1 Should the IMF vary?

We expect the IMF to vary since a galaxy's metallicity, SFRs and environment can change dramatically with time (e.g. Schwarzschild & Spitzer 1953; Larson 1998, 2005; Weidner et al. 2013; Chattopadhyay et al. 2015; Ferreras et al. 2015; Lacey et al. 2016). Lower metallicities, higher SFRs and high cloud surface densities prominent at high redshift can favour the formation of high-mass stars (Chattopadhyay et al. 2015) while interactions between gas clumps in dense environments can suppress the formation of low-mass stars (Krumholz et al. 2010). Furthermore, physically motivated models of early-type galaxies (ETGs) suggest scenarios in which star formation occurs in different periods giving rise to variability in the mass of the stars formed (Vazdekis et al. 1996; Weidner et al. 2013; Ferreras et al. 2015).

Following from theoretical predictions, recent observational studies have started showing increasing evidence for a non-universal IMF (Hoversten & Glazebrook 2008; van Dokkum & Conroy 2010; Gunawardhana et al. 2011; Meurer 2011; Cappellari et al. 2012, 2013; Conroy et al. 2013; Ferreras et al. 2013; La Barbera et al. 2013; Martín-Navarro et al. 2015a,b,c,d). These studies investigate both ETGs and late-type galaxies in different physical and environmental conditions and use different techniques to probe the IMF at the lower and upper mass end.

IMF studies of ETGs in the local universe infers/has shown a high abundance of low-mass stars (van Dokkum & Conroy 2010; Ferreras et al. 2013; La Barbera et al. 2013) with strong evidence for IMF variations as a function of galaxy velocity dispersion (Cappellari et al. 2012, 2013; van Dokkum & Conroy 2012; Conroy et al. 2013), metallicity (Martín-Navarro et al. 2015d) and radial distance within a galaxy (Martín-Navarro et al. 2015a). These results suggest that the IMF of ETGs are most likely to depend on the physical conditions of the galaxy when it formed bulk of its stars. Local star-forming galaxies show evidence for IMF variation as a function of galaxy luminosity (Schombert et al. 1990; Lee et al. 2004; Hoversten & Glazebrook 2008; Meurer et al. 2009; Meurer 2011), metallicity (Rigby & Rieke 2004), and SFR (Gunawardhana et al. 2011). Comparisons between H I-selected low and high surface brightness galaxies have shown the need for a systematic variation of the upper mass cut-off and/or the slope of the IMF to model the far-ultraviolet and Hα luminosities (Meurer et al. 2009; Meurer 2011). Hα EW and optical colour analysis of the Sloan Digital Sky Survey (SDSS) (York et al. 2000) data showed that low-luminous galaxies may be deficient in high-mass stars (Hoversten & Glazebrook 2008), while a similar analysis on the Galaxy And Mass Assembly (GAMA) survey (Driver et al. 2009) showed an excess of high-mass stars in high star-forming galaxies (Gunawardhana et al. 2011), both compared to expectations from a Salpeter IMF.

In spite of IMF being fundamental to galaxy evolution, our understanding of it at higher redshifts ($z \gtrsim 2$) is extremely limited. IMF studies of strong gravitational lenses at $z > 1$ have shown no deviation from Salpeter IMF (Pettini et al. 2000; Steidel et al. 2004; Quider et al. 2009), but quiescent galaxies at $z < 1.5$ have shown systematic trends for the IMF with stellar mass (Martín-Navarro et al. 2015c). Using local analogues to $z \sim 2$ galaxies, Martín-Navarro et al. (2015b) find evidence for an abundance of low-mass stars in the early universe.

Understanding the *relic* IMF at high redshift requires populations of quiescent galaxies which are relatively rare at high redshift, extremely long integration times to obtain absorption line/kinematic features and complicated modelling of stellar absorption line features. Since IMF defines the mass distribution of formed stars at a given time, in the context of understanding the role of IMF in galaxy evolution, it should be investigated *in situ* at an era when most galaxies are in their star-forming phase and evolving rapidly to produce large elliptical galaxies found locally. Furthermore, high-mass stars are absent in ETGs and therefore star-forming galaxies are imperative to study the high-mass end of the IMF. Simulations have shown that $z \sim 2$ universe is ideal for such studies (Hopkins & Beacom 2006). Rest-frame optical spectra of high-redshift galaxies are dominated by strong emission lines produced by nebulae associated with high-mass stars (>15 M$_{\odot}$) and therefore provide a direct tracer of the high-mass end of the IMF (Bastian et al. 2010). Due to the recent development of sensitive near-infrared (NIR) imagers and multiplexed spectrographs that take advantage of the Y, J, H and K atmospheric windows,

the $z \sim 2$ universe is ideal to study rest-frame optical features of galaxies.

1.2 Investigating the IMF with Hα emission

The total flux of a galaxy at Hα emission wavelength is the sum of the Hα emission flux, the continuum level at the same wavelength minus the Hα absorption. Hα absorption for galaxies at $z \sim 2$ is $\leq$3 per cent of its flux level (Reddy et al. 2015) and therefore can be ignored. In case B recombination (Brocklehurst 1971), following the Zanstra principle (Zanstra 1927), the Hα flux of a galaxy is directly related to the number of Lyman continuum photons emitted by massive young O and B stars with masses >10 $M_\odot$. The continuum flux at the same wavelength is dominated by red giant stars with masses between 0.7 and 3.0 $M_\odot$. Therefore, Hα EW that is the ratio of the strength of the emission line to the continuum level can be considered as the ratio of massive O and B stars to ∼1 $M_\odot$ stars present in a galaxy.

The rest-frame optical colours of a galaxy is tightly correlated with its Hα emission. The Hα flux probes the specific SFR (sSFR) of the shorter lived massive stars, while the optical colours probe the sSFR of the longer lived less massive stars. Therefore, in a smooth exponentially declining SFH, the optical colour of a galaxy will transit from bluer to redder colours with time due to the increased abundance of older less massive red stars. Similarly, with declining SFR the Hα flux will decrease and the continuum contribution of the older redder stars will increase, which will act to decrease the Hα EW in a similar SFH. The Hα EW and optical colour parameter space is degenerated in such a way that the slope of the function is equivalent to lowering the highest mass stars that are formed and/or increasing the fraction of intermediate-mass stars.

Multiple studies have investigated possibilities for IMF variation in galaxies using Balmer line flux in the context of probing SFHs (Meurer et al. 2009; Weisz et al. 2012; Zeimann et al. 2014; Guo et al. 2016; Smit et al. 2016). Modelling effects of IMF variation using Hα or Hβ to UV flux ratios have strong dependence on the assumed SFH and dust extinction of the galaxies and is only sensitive to the upper end of the high-mass IMF. Apart from IMF variation (Boselli et al. 2009; Meurer et al. 2009; Pflamm-Altenburg, Weidner & Kroupa 2009), stochasticity in SFH (Boselli et al. 2009; Fumagalli, da Silva & Krumholz 2011; Guo et al. 2016), non-constant SFHs (Weisz et al. 2012) and Lyman leakage (Relaño et al. 2012) can provide viable explanations to describe offsets between expected Balmer line to UV flux ratios and observed values.

Kauffmann (2014) used SFRs derived via multiple nebular emission line analysis with the 4000 Å break and $H\delta_A$ absorption to probe the recent SFHs of SDSS galaxies with $\log_{10}(M_*/M_\odot) < 10$ and infer possibilities for IMF variation. They did not find conclusive evidence for IMF variation, with contradictions in the 4000 Å features with Bruzual & Charlot (2003) stellar templates being attributed to errors in the spectro-photometric calibration. However, using absorption line analysis to probe possible IMF variations in actively star-forming galaxies suffers from strong Balmer line emissions that dominate and fill the absorption features. Furthermore, absorption lines probe older SPs, and linking them with current star formation requires further assumptions about the SFH.

Smit et al. (2016) used Spectral Energy Distribution (SED) fitting techniques to probe discrepancies between Hα to UV SFRs ratios of $z \sim$ 4–5 galaxies and local galaxies. They inferred an excess of ionizing photons in the $z \sim$ 4–5 galaxies but the origin could not be distinguished between a shallow high-mass IMF scenario and a metallicity dependent ionizing spectrum. Using broad-band imaging and SED fitting techniques to infer Hα flux has underlying uncertainties from the contamination of other emission lines that fall within the same observed filter (e.g. [N II], [S II]) and assumptions of IMF, SFH, metallicity and dust law of the SED templates.

In this paper, we use the MOSFIRE NIR spectra of galaxies obtained as a part of the ZFIRE survey (Nanayakkara et al. 2016) along with multiband photometric data from the ZFOURGE survey (Straatman et al. 2016) to study the high-mass IMF of a mass complete sample of star-forming galaxies at $z \sim 2$. MOSFIRE spectra has higher resolution and is able to probe into redder wavelengths (up to *K* band) compared to *Hubble Space Telescope* (*HST*) grism spectra (e.g. Zeimann et al. 2014), thus allows us to detect clear Hα nebular emissions up to $z \sim 2.7$. Galaxies follow nearly the same locus in Hα EW, optical colour as long as their SFHs are smoothly increasing or decreasing (Kennicutt 1983) and we show this in later figures. The change in galaxy evolutionary tracks due to IMF is largely orthogonal to the changes in tracks due to effects of dust extinction. Therefore, our method allows stronger constraints to be made on the high-mass IMF compared to the studies that probe Balmer line to UV flux ratios and is an improvement of the recipe that was first implemented by Kennicutt (1983) and subsequently used by Hoversten & Glazebrook (2008) and Gunawardhana et al. (2011) to study the IMF at $z \sim 0$.

The paper is organized in the following way. Section 2 describes the sample selection, the continuum fitting procedure, Hα EW calculation, optical colour derivation and the completeness of our selected sample. Section 3 shows how the synthetic stellar population (SSP) models were computed. In Section 3.2, we show the first results of our IMF study and identify SP effects that could describe the distribution of our sample. We discuss effects of dust to our analysis in Section 4, observational bias in Section 4.5 and starbursts in Section 5. In Section 6, we discuss the effects of other properties such as stellar rotation, binary stars, metallicity and the high-mass cut-off of the stellar models to our analysis. Section 7 investigates dependences of our parameters with other observables. Section 8 gives a through discussion of our results investigating the change of IMF and other possible scenarios. We conclude the paper in Section 9 by describing further improvements needed in the field to determine the IMF of the galaxies in the high-redshift universe. Throughout the paper, we refer to the IMF slope of $\Gamma = -1.35$ computed by Salpeter (1955) as the Salpeter IMF. We assume various IMFs and a cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_\Lambda = 0.7$ and $\Omega_m = 0.3$. All magnitudes are expressed using the AB system.

2 OBSERVATIONS AND DATA

2.1 Galaxy sample selection

The sample used in this study was selected from the ZFIRE (Nanayakkara et al. 2016) spectroscopic survey, which also consists of photometric data from the ZFOURGE survey (Straatman et al. 2016). In this section, we describe the sample selection process from the ZFIRE survey for our analysis.

ZFIRE is a spectroscopic redshift survey of star-forming galaxies at $1.5 < z < 2.5$, which utilized the MOSFIRE instrument (McLean et al. 2012) on Keck-I to primarily study galaxy properties in rich environments. ZFIRE has observed ∼300 galaxy redshifts with typical absolute accuracy of ∼15 km s^{-1} and derived basic galaxy properties using multiple emission line diagnostics. Nanayakkara et al. (2016) give full details on the ZFIRE survey. In this study, we use the subset of ZFIRE galaxies observed in the COSMOS field (Scoville et al. 2007) based on a stellar mass limited sample reaching

up to 5σ emission line flux limits of $\sim 3 \times 10^{-18}$ erg s^{-1} cm^{-2} selected from deep NIR data $K_{AB} < 25$ obtained by the ZFOURGE survey.

ZFOURGE[1] (PI I. Labbé) is a K_s-selected deep 45 night photometric legacy survey carried out using the purpose built FourStar imager (Persson et al. 2013) in the 6.5-m Magellan Telescopes located at Las Campanas observatory in Chile. The survey covers 121 arcmin^2 in each of the COSMOS, UDS (Beckwith et al. 2006) and CDFS (Giacconi et al. 2001) legacy fields. Deep FourStar medium band imaging (5σ depth of $K_s \leq 25.3$ AB) and the wealth of public multiwavelength photometric data (UV to far infrared) available in these fields were used to derive photometric redshifts with accuracies $\lesssim 1.5$ per cent using EAZY (Brammer, van Dokkum & Coppi 2008). Galaxy masses, ages, SFRs and dust properties were derived using FAST (Kriek et al. 2009) with a Chabrier (2003) IMF, exponentially declining SFHs and Calzetti et al. (2000) dust law. At $z \sim 2$, the public ZFOURGE catalogues are 80 per cent mass complete to $\sim 10^9$ M$_\odot$ (Nanayakkara et al. 2016). Refer to Straatman et al. (2016) for further details on the ZFOURGE survey.

ZFIRE and ZFOURGE are ideal surveys to use in this study since both provide mass complete samples. The total ZFIRE sample in the COSMOS field contains 142 Hα-detected [$>5\sigma$, redshift quality flag (Q_z)=3] star-forming galaxies that is mass complete down to $\log_{10}(M_*/M_\odot) > 9.30$ (at 80 per cent for $K_s = 24.11$). Thus, our Hα-selected sample contains no significant systematic biases towards SFH, stellar mass and magnitude. Furthermore, ZFIRE contains a large cluster at $z = 2$ containing 51 members with 5σ Hα detections (Yuan et al. 2014) and therefore we are able to examine if the IMF is affected by the local environment of galaxies.

For this study, we apply the following additional selection criteria to the 142 Hα-detected galaxies.

(i) We remove active galactic nucleus (AGN) using photometric (Cowley et al. 2016) and emission line [$\log_{10}$(f([N II])/f(Hα))>-0.5; Coil et al. (2015)] criteria resulting in identifying 26 AGN with our revised sample containing $N = 116$ galaxies. We note that all galaxies selected as AGN from ZFOURGE photometry by Cowley et al. (2016) are flagged as AGN by the Coil et al. (2015) selection. We further discuss contamination to Hα from subdominant AGN in Appendix A1.

(ii) Galaxies must have a matching ZFOURGE counterpart such that we can obtain galaxy properties, resulting in $N = 109$ galaxies.

(iii) We compute the total spectroscopic flux for these galaxies and remove four galaxies with negative fluxes resulting in $N = 105$ galaxies. We perform stringent Hα emission quality cuts to the spectra for these 105 galaxies and remove two galaxies due to strong sky line subtraction issues. We further remove one galaxy due to an overlap of the galaxy spectra with a secondary object that falls within the same slit.

Our final sample of galaxies used for the IMF analysis in this paper comprise of 102 galaxies. Henceforth, we refer to this sample of galaxies as the ZFIRE SP sample.

The redshift distribution for the ZFIRE-SP sample is shown by Fig. 1. The ZFIRE-SP sample is divided into continuum-detected and non-detected galaxies as described in Section 2.3. Galaxies in our sample lie within redshifts of $1.97 < z < 2.46$ corresponding to a $\Delta t \sim 650$ Myr.

[1] http://zfourge.tamu.edu

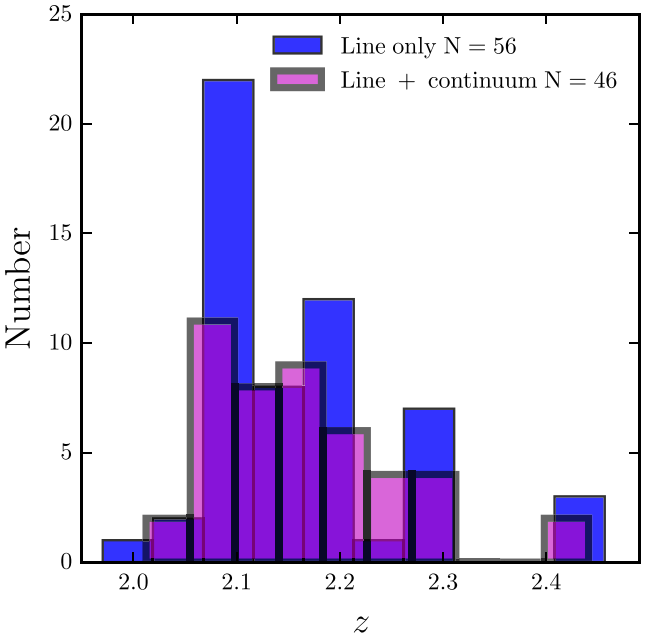

Figure 1. The redshift distribution of the ZFIRE-SP sample. Galaxies with line+continuum detection are shown by magenta and galaxies only with Hα line detection are shown by blue.

2.2 Completeness

In order to determine any significant detection biases in our ZFIRE-SP sample, we evaluate the completeness of the galaxies selected in this analysis. We define a redshift window for analysis between $1.90 < z < 2.66$ (~8.6 Gpc), which corresponds to the redshifts that Hα emission will fall within the MOSFIRE K band. Note that here we discuss galaxies with Hα detections and $Q_z > 1$, while in Section 2.1 we discussed the $Q_z = 3$ Hα-detected sample.

In the ZFOURGE catalogues used for the ZFIRE sample selection (see Nanayakkara et al. 2016 for details), there were 1159 galaxies (including star-forming and quiescent galaxies) in the COSMOS field with photometric redshifts ($z_{\rm photo}$) within $1.90 < z_{\rm photo} < 2.66$. 160 of these galaxies with $1.90 < z_{\rm photo} < 2.66$ were targeted in K band out of which 128[2] were detected with at least one emission line with SNR >5. None of the Hα-detected galaxies had spectroscopic redshifts outside the considered redshift interval. However, three additional galaxies (one object with $Q_z = 2$, two objects with $Q_z = 3$) fell within $1.90 < z < 2.66$ due to inaccurate photometric redshifts. There were eight galaxies targeted in K band that did not have Hα detections but do have other emission line detections (i.e. No Hα but have [N II], [O III], Hβ etc.). Furthermore, there were no galaxies that were targeted in K band expecting Hα but resulted in other emission line detections.

There were 151 objects within $1.90 < z < 2.66$ with Hα detections ($Q_z > 1$) and 26 of them were flagged as AGN following selection criteria from Coil et al. (2015) and Cowley et al. (2016). In the remaining 125 galaxies, 8 galaxies did not have matching ZFOURGE counterparts and 8 galaxies had low confidence for redshift detection ($Q_z = 2$) from Nanayakkara et al. (2016). We

[2] Note that this is different from the 142 galaxies mentioned in Section 2.1 because the sample of 142 galaxies has a $Q_z = 3$, includes galaxies with no ZFOURGE counterparts (see Nanayakkara et al. 2016 for further details) and galaxies with non-optimal ZFOURGE photometry (see Straatman et al. 2016 for further details).

Table 1. Galaxies selected for the IMF study.

M_* [a]	$N_{\rm ZFOURGE}$	$N_{\rm ZFIRE}$	$N_{\rm detections}$	$N_{\rm outliers}$	$N_{\rm AGN}$ [b]	$N_{\rm sky}$	$N_{\rm selected}$	$N_{\rm line\_only}$	$N_{\rm null\_detection}$
<9.5	568	59	44	0	2	1	41	34	9
9.5–10.0	318	47	39	0	1	0	35	16	6
10.0 <	273	54	45	0	20	1	26	6	5
Total	1159	160	128	0	23	2	102	56	20

Notes. The columns keys are as follows:
M_*: the Mass bin of the galaxies in $\log_{10}(M_\odot)$.
$N_{\rm ZFOURGE}$: number of ZFOURGE galaxies with photometric redshifts within $1.90 < z < 2.66$.
$N_{\rm ZFIRE}$: number of ZFIRE-targeted galaxies in K band with photometric redshifts within $1.90 < z < 2.66$.
$N_{\rm detections}$: number of ZFIRE-detected galaxies in K band ($Q_z > 1$) with spectroscopic redshifts within $1.90 < z < 2.66$.
$N_{\rm outliers}$: number of ZFIRE-detected galaxies with spectroscopic redshifts outside $1.90 < z < 2.66$.
$N_{\rm AGN}$: number of ZFIRE-detected galaxies identified as AGN with spectroscopic redshifts within $1.90 < z < 2.66$.
$N_{\rm sky}$: number of ZFIRE-detected galaxies with spectroscopic redshifts within $1.90 < z < 2.66$ removed from ZFIRE-SP sample due to sky line interference.
$N_{\rm selected}$: number of ZFIRE-detected galaxies selected for the IMF study with spectroscopic redshifts within $1.90 < z < 2.66$.
$N_{\rm line\_only}$: number of galaxies selected for the IMF study which shows no continuum detection with spectroscopic redshifts within $1.90 < z < 2.66$.
$N_{\rm null\_detection}$: number of ZFIRE K-band-targeted galaxies with photometric redshifts within $1.90 < z < 2.66$ and no Hα detection.
[a]Where applicable spectroscopic redshifts have been used to calculate the stellar masses from FAST.
[b]One galaxy flagged as an AGN does not have a matching ZFOURGE counterpart.

removed those 16 galaxies from the sample. Out of the 109 remaining galaxies, 7 are removed due to the following reasons: 4 galaxies due to negative spectroscopic flux, 1 galaxy due to multiple objects overlapping in the spectra and 2 galaxies due to extreme sky line interference.

Our sample constitutes of the remaining 102 galaxies out of which, 46 have continuum detections (see Section 2.3). Furthermore, 38 (out of which 16 are continuum detected) galaxies are confirmed cluster members (Yuan et al. 2014) and the remaining 64 (out of which 30 are continuum detected) galaxies comprise of field galaxies. 32 galaxies targeted with photometric redshifts between $1.90 < z < 2.66$ show no Hα emission detection. We divide our sample into three mass bins with masses between $\log_{10}(M_\odot) < 9.5$, $9.5 \leq \log_{10}(M_\odot) \leq 10.0$, $10.0 < \log_{10}(M_\odot)$ and show the corresponding data as described above in Table 1.

We define observing completeness as the percentage of detected galaxies ($Q_z > 1$) with photometric redshifts between $1.90 < z < 2.66$ and calculate it to be ∼80 per cent. However, it is possible that the 32 null detections with photometric redshifts within $1.90 < z < 2.66$ to have been detected if the ZFIRE survey was more sensitive. We stack the photometric redshift likelihood functions ($P(z)$) of the ZFIRE targeted galaxies within this redshift range, to compute the expectation of detections based of photometric redshift accuracies (see the Nanayakkara et al. 2016 section 3.2 for further details on how $P(z)$ stacking is performed). The calculated expectation for Hα to be detected within K band is ∼80 per cent, which is extremely similar to the observed completeness. Therefore, non-detections rate is consistent with uncertainties in the photometric redshifts. To further account for any detection bias, we employ a stacking technique of the non-detected spectra in order to calculate a lower limit to the stacked EW values. This is further discussed in Section 5.2.

2.3 Continuum fitting and Hα EW calculation

In this section, we describe our continuum fitting method for our 102 Hα-detected galaxies selected from the ZFIRE survey. Fitting a robust continuum level to a spectrum requires nebular emission lines and sky line residuals to be masked. Furthermore, the wavelength interval used for the continuum fit should be sufficient enough to perform an effective fit but should be smaller enough to not to be influenced by the intrinsic SED shape. After extensive testing of various measures used to fit a continuum level, we find the method outlined below to be the most effective to fit a continuum level for our sample.

By visual inspection and spectroscopic redshift of the galaxies in our sample, we mask out the Hα and [N II] emission line regions in the spectra. We further mask all known sky lines by selecting a region × 2 the spectral resolution (±5.5 Å) of MOSFIRE K band. We then use the `astropy` (Astropy Collaboration et al. 2013) sigma-clipping algorithm to mask out remaining strong features in the spectra. These spectra are then used to fit an inverse variance weighted constant level fit, which we consider as the continuum level of the galaxy. Three objects fail to give a positive continuum level using this method and for these we perform a 3σ clip with two iterations without masking nebular emission lines and sky lines. Using this method we are able to fit positive continuum levels to all galaxies in our sample. We further investigate the robustness of our measures continuum levels in Appendix A using ZFOURGE photometric data and conclude that our measured continuum level is consistent (or in agreement) with the photometry.

We use two approaches to calculate the Hα line flux: (1) direct flux measurement and (2) Gaussian fit to sky-line blended and kinematically unresolved emission lines. Our two methods provide consistent results for emission lines that are not blended with sky lines (see Appendix A). By visual inspection, we selected kinematically resolved (due to galaxy rotation etc.) Hα emission lines that were not blended with sky lines and computed the EW by integrating the line flux. Within the defined emission line region, we calculated the Hα flux by subtracting the flux at each pixel (F_i) by the corresponding continuum level of the pixel ($F_{{\rm cont}_i}$). For the remaining sample, which comprises of galaxies with no strong velocity structure and galaxies with Hα emission with little velocity structure and/or Hα contaminated by sky lines, we perform Gaussian fits to the emission lines, to calculate the Hα flux values. We then subtract the continuum level from the computed Hα line flux.

Next, we use the calculated Hα flux along with the fitted continuum level to calculate the observed Hα EW (H$\alpha_{\rm EW_{obs}}$) as follows:

$$\mathrm{H}\,\alpha\ \mathrm{EW_{obs}} = \sum_i \left(1 - \frac{F_i - F_{\mathrm{cont}_i}}{F_{\mathrm{cont}_i}}\right) \times \Delta\lambda_i, \tag{2a}$$

where $\Delta\lambda_i$ is the increment of wavelength per pixel. Finally, using the spectroscopic redshift (z) we calculate the rest-frame Hα EW ($\alpha_{\mathrm{EW_{rest}}}$), which we use throughout the paper:

$$\mathrm{H}\alpha\ \mathrm{EW} = \frac{\mathrm{H}\alpha\ \mathrm{EW_{obs}}}{1+z}. \quad (2b)$$

We calculate EW errors by bootstrap re-sampling the flux of each spectra randomly within limits specified by its error spectrum. We re-calculate the EW iteratively 1000 times and use the 16th and 84th percentile of the distribution of the EW measurements as the lower and upper limits for the EW error, respectively. Since the main uncertainty arises from the continuum fitting, we do not consider the error of the Hα flux calculation in our bootstrap process.

The robustness of an EW measurement relies on the clear identification of the nebular emission line and the underlying continuum level. The latter becomes increasingly hard to quantify at high redshift for faint star-forming sources due to the continuum not being detected. Therefore, we derive continuum detection limits to identify robustly measured continua from non-detections.

In order to establish the limit to which our method can reliably measure the continuum, we select 14 2D slits with no continuum or nebular emission line detections to extract 1D spectra. We define extraction apertures using a representative profile of the flux monitor star and perform multiple extractions per slit depending on their spatial size. A total of 93 1D sky spectra are extracted and their continuum level is measured by masking out sky lines and performing a sigma-clipping algorithm. The error of the sky continuum fit is calculated by bootstrap re-sampling of the sky fluxes 1000 times. We consider the 1σ scatter of the bootstrapped continuum values to be the error of the sky continuum fit and 1σ scatter of the flux values used for the continuum fit as the rms of the flux.

The comparison between the flux continuum levels for the sky spectra with the ZFIRE-SP sample spectra are shown in Fig. 2. The median and the 2σ standard deviation for the continuum levels of the sky spectra are -2.3×10^{-21} erg s^{-1} cm^{-2} Å^{-1} and 5.4×10^{-20} erg s^{-1} cm^{-2} Å^{-1}, respectively. We consider the horizontal blue dashed line in Fig. 2, which is 2σ above the median sky level, to be our lower limit for the continuum detections in our sample. The 46 galaxies in our ZFIRE-SP sample with continuum levels above this flux level detections are considered to have a robust continuum detection. For the remaining 56 galaxies, we consider the continuum measurement as a limit and use it to calculate a lower limit to the Hα EW values. The redshift distribution of these galaxies is shown by Fig. 1.

2.4 Calculating optical colours

Rest-frame optical colours for the ZFIRE-SP sample are computed using an updated version of EAZY[3] (Brammer et al. 2008), which derives the best-fitting SEDs for galaxies using high-quality ZFOURGE photometry to compute the colours. We investigate the robustness of the rest-frame colour calculation of EAZY in Appendix B3. The main analysis of our sample is carried out using optical colours derived using two idealized, synthetic box-car filters, which probes the bluer and redder regions of the rest-frame SEDs. We select these filters to avoid regions of strong nebular emission lines as explained in Section 3 and Appendix B.

In order to allow direct comparison between ZFIRE $z \sim 2$ galaxies with $z = 0.1$ SDSS galaxies from HG08, we further calculate optical colours for the ZFIRE-SP sample at $z = 0.1$ using blueshifted SDSS g and r filters. Blueshifting the filters simplifies the $(g-r)$ colour calculation at $z = 0.1$ ($(g-r)_{0.1}$) by avoiding additional uncertainties, which may arise due to K-corrections if we redshift the galaxy spectra to $z = 0.1$ from $z = 0$.

[3] Development version: https://github.com/gbrammer/eazy-photoz/

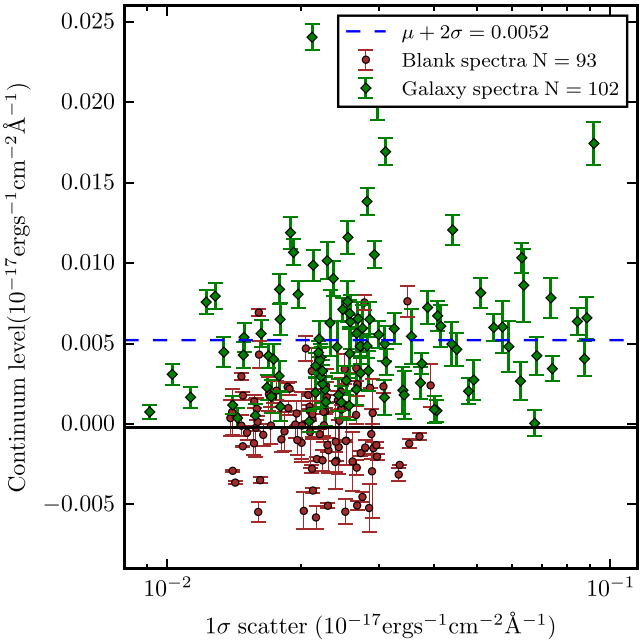

Figure 2. The figure illustrates the continuum detection levels for the ZFIRE-SP sample. The measured continuum level is plotted against the 1σ scatter of the flux values used to fit the continuum level. The brown circles represent the continuum levels calculated for the blank slits and the green diamonds represent the continuum level calculated for the IMF sample. The blue horizontal line is the 2σ scatter above the median ($\sim$0) for the blank sky regions. Any continua detected above this level of 5.2×10^{-20} erg s^{-1} cm^{-2} Å^{-1} are considered as detected continuum levels.

3 GALAXY SPECTRAL MODELS

In this section, we describe the theoretical galaxy stellar spectral models employed to investigate the effect of IMF, SFHs and other fundamental galaxy properties in Hα EW versus optical colour parameter space. We use PEGASE.2 detailed in Fioc & Rocca-Volmerange (1997) as our primary spectral synthesis code to perform our analysis and further employ STARBURST99 (S99; Leitherer et al. 1999) and BPASSV2 (Eldridge & Stanway 2016) models to investigate the effects of other exotic stellar features.

PEGASE is a publicly available spectral synthesis code developed by the Institut d'Astrophysique de Paris. Once the input parameters are provided, PEGASE produces galaxy spectra for varying time-steps, which can be used to evaluate the evolution of fundamental galaxy properties over cosmic time.

3.1 Model parameters

In this paper, we primarily focus on the effect of varying the IMF, SFH and metallicity on Hα EW and optical colour of galaxies. A thorough description of the behaviour of PEGASE models in this parameter space can be found in Hoversten & Glazebrook (2008). The parameters we vary are as follows.

(i) The IMF : we follow HG08 and use an IMF with a single power law as shown by equation (1). Models were calculated with varying IMF slopes (Γ) ranging between −0.5 and −2.0 in logarithmic space. The lower and upper mass cut-offs were set to 0.5 and 120 $M_\odot$, respectively. The IMF indication method used in this analysis is dependant on the ability of a star with a specific mass to influence the Hα emission and the optical continuum level. Stars below 1 $M_\odot$ cannot strongly influence the optical continuum and hence this method is not sensitive to probe the IMF below 1 $M_\odot$. Furthermore stars below ∼0.5 $M_\odot$ gives no significant variation to the parameters investigated by this method (Hoversten & Glazebrook 2008). Therefore, we leave the lower mass cut-off at 0.5 $M_\odot$. The higher mass cut-off of 120 $M_\odot$ used is the maximum mass allowed by PEGASE. Though the high-mass end has a stronger influence on the IMF identified using this method we justify the 120 $M_\odot$ cut-off due to the ambiguity of the stellar evolution models above this mass. Varying the upper mass cut-off has a strong effect on Hα EW and optical colours. As HG08 showed this is strongly degenerated with changing Γ. In this work, we focus on Γ parametrization, noting that changing the cut-off could produce similar effects. We further discuss the degeneracy between the high-mass cut-off and the Hα EW versus optical colours slope in Section 6.4.

(ii) The SFH : exponentially increasing/declining SFHs, constant SFHs and starbursts are used. Exponentially declining SFHs are in the form of $\mathrm{SFR}(t) = p_2 \exp(-t/p_1)/p_1$, with p_1 varying from 500 to 1500 Myr. Starbursts are used on top of constant SFHs with varying burst strength and time-scales. Further details are provided in Section 5.3.

(iii) Metallicity : models with consistent metallicity evolution and models with fixed metallicity of 0.02 are used.

The other parameters we use for the PEGASE models are as follows. We use Supernova ejecta model B from Woosley & Weaver (1995) with the stellar wind option turned on. The fraction of close binary systems are left at 0.05 and the initial metallicity of the ISM is set at 0. We turn off the galactic in-fall function and the mass fractions of the substellar objects with star formation are kept at 0, Galactic winds are turned off, nebular emissions are turned on and we do not consider extinction as we extinction correct our data.

As a comparison with HG08, in Fig. 3 we show the evolution of four model galaxies from PEGASE in the Hα EW versus $(g-r)_{0.1}$ colour space. The models computed with exponentially declining SFHs with $p_1 = 1000$ Myr, varying IMFs and nebular emission lines agree well with the SDSS data. However, the evolution of the $(g-r)_{0.1}$ colour shows strong dependence on the nebular emission contribution, especially for shallower IMFs. HG08 never considered the effect of emission lines in $(g-r)_{0.1}$ colours and the significant effect at younger ages/bluer colours are likely to be important for $z \sim 2$ galaxies.

Fig. 4 shows an example of a synthetic galaxy spectra generated by PEGASE. The galaxy is modelled to have an exponentially declining SFH with $p_1 = 1000$ Myr and a $\Gamma = -1.35$ IMF. Due to the declining nature of the SFR, the stellar and nebular contribution of the galaxy spectra decreases with cosmic time. We overlay the filter response functions of the $g_{z=0.1}$ and $r_{z=0.1}$ filters used in the analysis by HG08. As evident from the spectra, this spectral region covered by the $g_{z=0.1}$ and $r_{z=0.1}$ filters includes strong emission lines such as [O III] and Hβ. Therefore, the computed $(g-r)_{0.1}$ colours will have a strong dependence on photoionization properties of the galaxies.

To mitigate uncertainties in photoionization models in our analysis, we employ synthetic filters specifically designed to avoid regions with strong nebular emission lines. We design two box-car filters centred at 3400 and 5500 Å with a width of 450 Å. The rest-frame wavelength coverage of these filters corresponds to a similar region covered by the FourStar J_1 and H_{long} filters in the observed frame for galaxies at $z = 2.1$ and therefore requires negligible K-corrections. Further details on this filter choice is provided in Appendix B1. Henceforth, we refer to the blue filter as [340], the redder filter as [550] and the colour of blue filter − red filter as [340]–[550]. The [340]–[550] colour evolution of a galaxy is independent of the nebular emission lines.

We also compare results using S99 (Leitherer et al. 1999) models in Appendix C. We find that PEGASE and S99 models show similar evolution and find that our choice of SSP model (PEGASE or S99) to interpret the IMF of the ZFIRE-SP sample at $z \sim 2$ to be largely independent to our conclusions. However, stellar libraries that

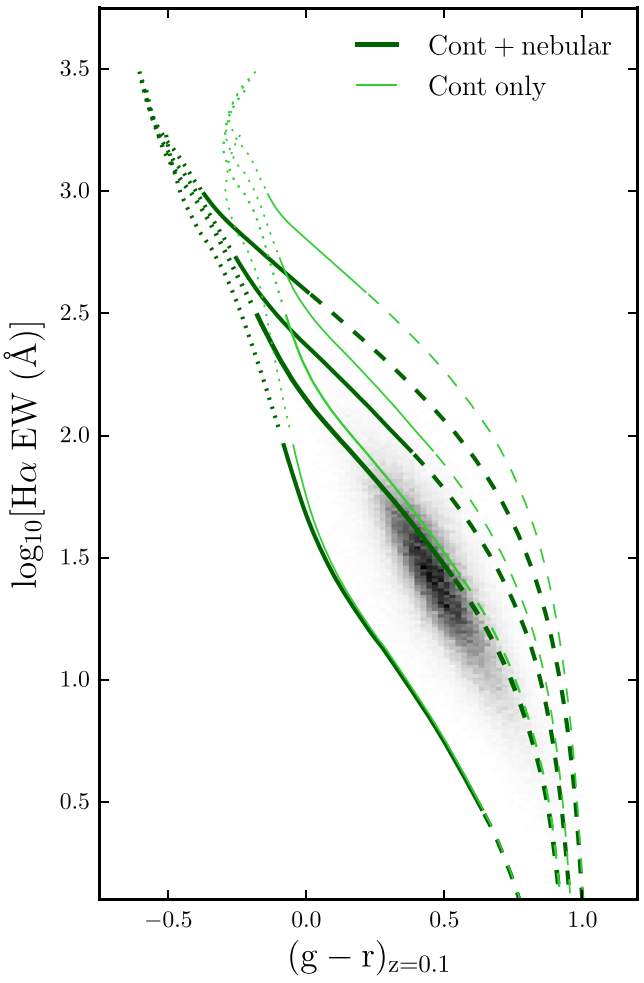

Figure 3. The evolution of PEGASE SSP galaxies in the Hα EW versus $(g-r)_{0.1}$ colour space. We show four galaxy models with exponentially declining SFHs computed using identical parameters but varying IMFs. The thick dark green tracks show from top to bottom galaxies with Γ values of −0.5, −1.0, −1.35 and −2.0, respectively. The thin light green tracks follow the same evolution as the thick ones, but the nebular line contribution is not considered for the $(g-r)_{0.1}$ colour calculation. All tracks commence at the top left of the figure and are divided into three time bins. The dotted section of the track corresponds to the first 100 Myr of evolution of the galaxy. The solid section of the tracks show the evolution between 100 and 3100 Myr ($z \sim 2$) and the final dashed section shows evolution of the galaxy up to 13 100 Myr ($z \sim 0$). The distribution of the galaxies from the SDSS HG08 sample is shown by 2D histogram.

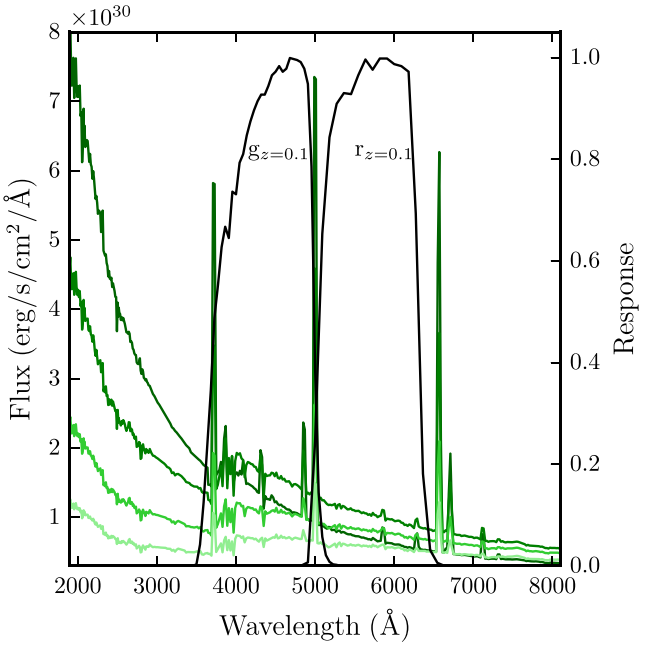

Figure 4. An example of a model galaxy spectrum generated by PEGASE. Here, we show the evolution of the optical wavelength of a galaxy spectra with an exponentially declining SFH and a $\Gamma = -1.35$ with no metallicity evolution. The time-steps of the models from top to bottom are: 100 Myr (dark green), 1100 Myr (green), 2100 Myr (lime green) and 3100 Myr (light green). The $g_{z=0.1}$ and $r_{z=0.1}$ filter response functions are overlaid on the figure.

introduce rotational and/or binary stars used in these models do have an influence of the Hα EWs and [340]–[550] colours, which we discuss in detail in Section 8.4.

3.2 Comparison to Hα EW and optical colours at $z \sim 2$

We explore the IMF of $z \sim 2$ star-forming galaxies using Hα EW values from ZFIRE spectra and rest-frame optical colours from ZFOURGE photometry. Our observed sample used in our analysis is shown in Fig. 5. The left-hand panel shows the distribution of Hα EW and [340]–[550] colours of the ZFIRE-SP sample before dust corrections are applied. We overlay model galaxy tracks generated by PEGASE for various IMFs. All models are computed using an exponentially declining SFH, but with varying time constants (p_1) as shown in the figure caption. For a given IMF, smoothly varying monotonic SFHs have very similar loci in this parameter space. The thick set of models (third from top) shows a slope with $\Gamma = -1.35$, which is similar to the Salpeter slope. Galaxies above these tracks are expected to contain a higher fraction of higher mass stars in comparison to the mass distribution expected following a Salpeter IMF. Similarly, galaxies below these tracks are expected to contain a lower fraction of high-mass stars. Galaxies have a large spread in this parameter space but we expect this scatter to decrease when dust corrections are applied to the data as outlined in Section 4.1.

We note the large scatter of the Hα EW values with respect to the Salpeter IMF, especially the large number of high-EW objects ($\gtrsim$0.5 dex above the Salpeter locus). Could this simply be due to the ZFIRE-SP sample only detecting Hα emissions in bright objects, i.e. a sample bias? First, we note our high completeness of ∼80 per cent for Hα detections (Section 2.2). Secondly, our Hα flux limits are actually quite faint. To show this explicitly, we define Hα flux detection limits for our sample using 1σ detection thresholds for each galaxy parametrized by the integration of the error spectrum within the same width as the emission line. Fig. 5 (right-hand panel) shows the Hα EW calculated using Hα flux detection limits, which illustrates the distribution of the ZFIRE-SP sample if the Hα flux was barely detected. The Hα EW of the continuum-detected galaxies decrease by ∼1 dex which suggest that our EW detection threshold is not biased towards higher Hα EW values.

Similar to IMF, there are a number of effects that may account for the clear disagreement between the observed data and models. In subsequent sections, we explore effects from

(i) dust (Section 4),
(ii) observational bias (Section 4.5),
(iii) starbursts (Section 5),
(iv) stellar rotation (Section 6.1),
(v) binary stellar systems (Section 6.2),
(vi) metallicity (Section 6.3) and
(vii) high-mass cut-off (Section 6.4)

in SSP models to explain the distribution of Hα EW versus optical colours of the ZFIRE-SP sample without invoking IMF change.

4 IS DUST THE REASON?

As summarized by Kennicutt (1983), the dust vector is nearly orthogonal to IMF change vector and therefore we expect the tracks in the Hα EW versus optical colour parameter space to be independent of galaxy dust properties. In this section, we describe galaxy dust properties. We explain how dust corrections were applied to the data and their IMF dependence and explore the difference in reddening between stellar and nebular emission line regions as quantified by Calzetti et al. (2000) for $z \sim 0$ star-forming galaxies.

We use FAST (Kriek et al. 2009) with ZFIRE spectroscopic redshifts from Nanayakkara et al. (2016) and multiwavelength photometric data from ZFOURGE (Straatman et al. 2016) to generate estimates for stellar attenuation (A_v) and stellar mass for our galaxies. FAST uses SSP models from Bruzual & Charlot (2003) and a χ^2 fitting algorithm to derive ages, star formation time-scales and dust content of the galaxies. All FAST SED templates have been calculated assuming solar metallicity, Chabrier (2003) IMF and Calzetti et al. (2000) dust law. We refer the reader to Straatman et al. (2016) for further information on the use of FAST to derive SP properties in the ZFOURGE survey.

4.1 Applying SED-derived dust corrections to data

We use stellar attenuation values calculated by FAST to perform dust corrections to our data. First, we consider the dust corrections for rest-frame Hα EWs and then we correct the [340]–[550] colours.

By using Cardelli, Clayton & Mathis (1989) and Calzetti et al. (2000) attenuation laws to correct nebular and continuum emission lines, respectively, we derive the following equation to obtain dust-corrected Hα EW (EW_i) values:

$$\log_{10}(\mathrm{EW}_i) = \log_{10}(\mathrm{EW}_{\mathrm{obs}}) + 0.4 A_{\mathrm{c}}(V)(0.62 f - 0.82), \quad (3)$$

where $\mathrm{EW}_{\mathrm{obs}}$ is the observed EW, A_{c} is the SED-derived continuum attenuation and f is the difference in reddening between continuum and nebular emission lines.

Calzetti et al. (2000) found a $f \sim 2$ for $z \sim 0$ star-forming galaxies, which we use for our analysis under the assumption that the actively star-forming galaxies at $z \sim 0$ are analogues to star-forming galaxies at $z \sim 2$. Henceforth, for convenience we refer to $f = 1/0.44$ Calzetti et al. (2000) value as $f = 2$. We further show key plots in this

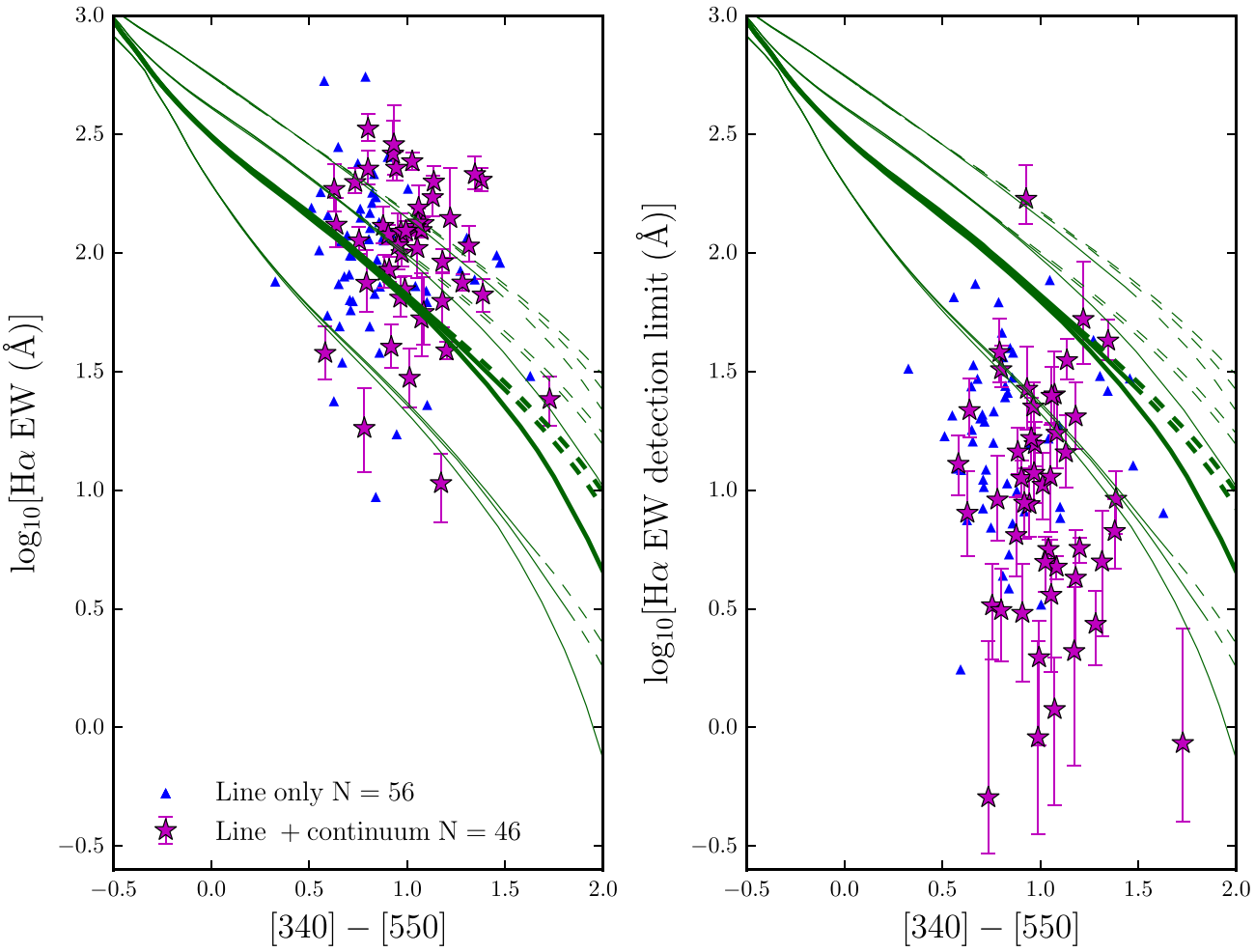

Figure 5. The Hα EW versus [340]–[550] colour distribution of the ZFIRE-SP sample. No dust corrections have been applied to the observed data. Galaxies with Hα and continuum detections are shown by magenta stars while galaxies only with Hα detections (and continuum from 1σ upper limits) are shown as 1σ lower limits on EW by blue triangles. The errors for the continuum-detected galaxies are from bootstrap re-sampling. The solid ($t < 3200$ Myr) and dashed lines ($t > 3200$ Myr) are SSP models computed from PEGASE. Similar to Fig. 3, we compute models for four IMFs with Γ values of −0.5, −1.0, −1.35 (this is the thick set of tracks which is similar to the IMF slope inferred by Salpeter) and −2.0. Each set of tracks from top to bottom represents these IMF in order. For each IMF, we compute three models with exponentially declining SFHs with varying p_1 values. From top to bottom, for each IMF these tracks represent p_1 values of 1500, 1000 and 500 Myr. *Left:* the Hα EW versus [340]–[550] colours of the ZFIRE-SP sample. *Right:* similar to the left figure but the Hα EW has been calculated using 1σ detection limits of the Hα flux values to demonstrate the sensitivity limits of our EW measurements.

analysis using a dust correction of $f = 1$ to consider equal dust extinction between stellar and ionized gas regions. This is driven by the assumption that A and G stars that contribute to the continuum of $z \sim 2$ star-forming galaxies are still associated within their original birthplaces similar to O and B stars due to insufficient time for the stars to move away from the parent birth clouds within the <3 Gyr time-scale.

Similarly, using Calzetti et al. (2000) attenuation law we obtain dust-corrected fluxes for the [340] and [550] filters as follows:

$$f([340]) = f([340]_{\rm obs}) \times 10^{0.4 \times 1.56 A_{\rm c}(V)} \quad (4a)$$

$$f([550]) = f([550]_{\rm obs}) \times 10^{0.4 \times 1.00 A_{\rm c}(V)}. \quad (4b)$$

A complete derivation of the dust corrections presented here are shown in Appendix D2.

Fig. 6 shows the distribution of our sample before and after dust corrections are applied. In the left-hand panels, we show our sample before any dust corrections are applied, with arrows in cyan denoting dust vectors for varying f values. It is evident from the figure that the galaxies in this parameter space are very dependent on the f value used. For f values of 1 and 2, the effect of dust is orthogonal to IMF change, while values above 2 may influence the interpretation of the IMF. We note that $f > 2$ makes the problem of high Hα EW objects worse, so we do not consider such values further.

Fig. 6 right-hand panels show the dust corrections applied to both Hα EW and the [340]–[550] colours for the ZFIRE-SP sample. Without the effect of dust, we expect the young star-forming galaxies to show similar bluer colours and therefore the narrower [340]–[550] colour space occupied by our dust-corrected sample is expected. With a dust correction of $f = 1$, majority of the galaxies lie below the $\Gamma = -1.35$ IMF track with only ∼1/5th of galaxies showing higher Hα EWs. However, with $f \sim 2$ dust correction, there is a significant presence of galaxies with extremely high Hα EW values for a given [340]–[550] colour inferred from a $\Gamma = -1.35$ IMF and ∼60 per cent of the galaxies lie above this IMF track.

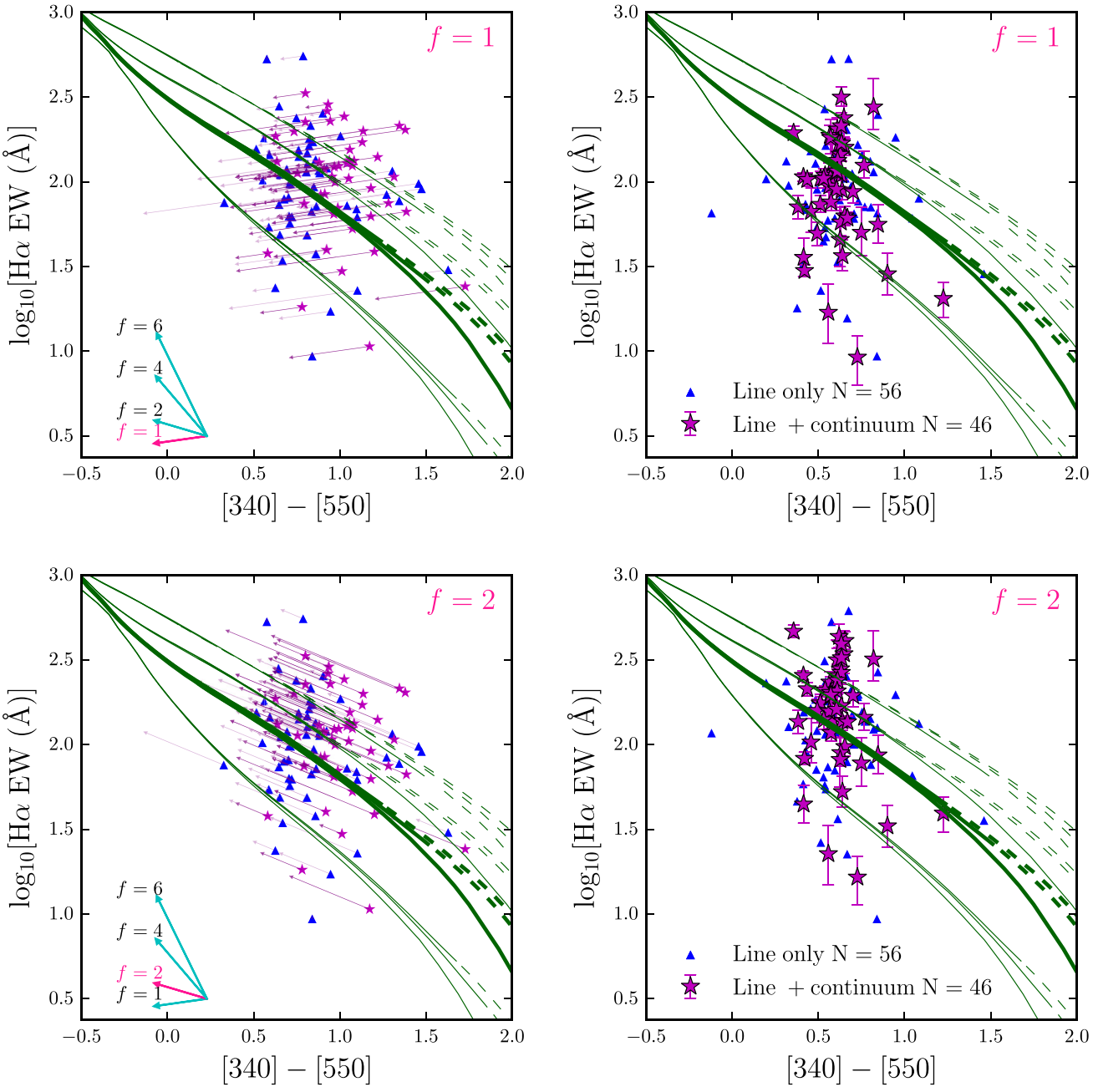

Figure 6. The dust correction process of the ZFIRE-SP sample. This figure is similar to Fig. 5 but shows the intermediate and final step of the dust correction process. *Top left:* here, we show the dust correction vector for each galaxy in our sample, computed following the prescriptions explained in Section 4.1. In summary, we use Calzetti et al. (2000) attenuation law to correct the continuum levels and the optical [340]–[550] colours. We use Cardelli et al. (1989) attenuation law to dust correct the nebular emission lines. We use attenuation values calculated by FAST and apply equal amount of extinction to continuum and nebular emission line regions. The purple arrows denote the dust vector for the individual galaxies. Galaxies with no arrows have 0 extinction. The arrows in the bottom left corner show the dust vector for a galaxy with $A_v = 0.5$ but with varying Calzetti, Kinney & Storchi-Bergmann (1994) factors, which is shown as f next to each arrow. *Top right:* the final Hα EW versus [340]–[550] colour distribution of the dust-corrected ZFIRE-SP sample with $f = 1$. Most galaxies lie at ([340] − [550]) ∼ 0.6, which corresponds to ∼850 Myr of age following the Salpeter IMF tracks. *Bottom left:* similar to top left panel, but with a higher amount of extinction to nebular emission line regions compared (∼×2) to the continuum levels. *Bottom right:* the final Hα EW versus [340]–[550] colour distribution of the dust-corrected ZFIRE-SP sample with $f = 2.27$.

Even ∼×2 larger errors for the individual Hα EW measurements cannot account for the galaxies with the largest deviations from the Salpeter tracks. The change of f from 2⇒1 decreases the median Hα EW value by ∼0.2 dex. However, galaxies still show a large scatter in Hα EW versus [340]–[550] colour parameter space with points lying well above the Salpeter IMF track.

The form of the attenuation law of galaxies at $z > 2$ shows conflicting results between studies. Observations from the Atacama Large Millimeter Array have indicated the presence of galaxies with low infrared (IR) luminosities suggesting galaxies with attenuation similar to the Small Magellanic Cloud (SMC, Capak et al. 2015; Bouwens et al. 2016). Reddy et al. (2015) showed an SMC-like

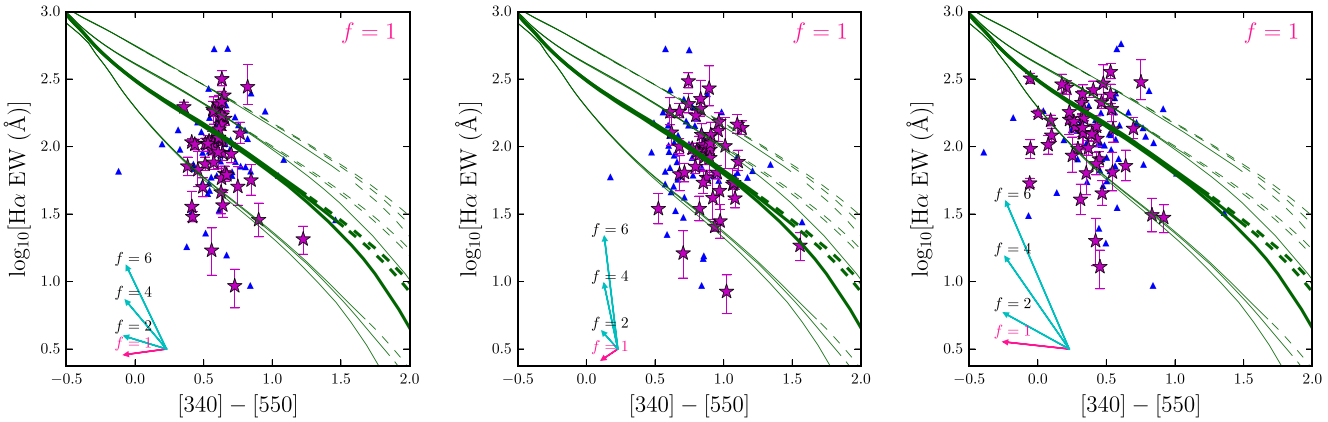

Figure 7. Here, we show the distribution of the ZFIRE-SP sample in the Hα EW versus [340]–[550] colour parameter space with the Hα continuum and optical colours dust corrections applied following *left:* Calzetti et al. (2000) attenuation law, *centre:* Pei (1992) SMC attenuation law and *right:* Reddy et al. (2015) attenuation law. In all panels, Cardelli et al. (1989) attenuation law has been used to dust correct the nebular emission lines with equal amount of extinction applied to continuum and nebular emission line regions ($f = 1$). The arrows in the bottom left corner show the dust vector for a galaxy with $A_v = 0.5$ but with varying Calzetti et al. (1994) factors, which is shown as f next to each arrow.

attenuation curve for $z \sim 2$ galaxies at $\lambda \gtrsim 2500$ Å and a Calzetti et al. (2000)-like attenuation curve for the shorter wavelengths. However, *HST* grism and SED fitting analysis of galaxies at $z \sim 2$–6 has shown no deviation in the attenuation law derived by Calzetti et al. (2000) for local star-forming galaxies. Such conflicts are also apparent in simulation studies, where Mancini et al. (2016) showed evidence for SMC-like attenuation with clumpy dust regions while Cullen et al. (2017) have shown that galaxies contain similar dust properties as inferred by Calzetti et al. (2000).

In order to understand the role of dust laws in the Hα EW versus [340]–[550] colour parameter space, we compare the results using other dust laws such as Pei (1992) SMC dust law and Reddy et al. (2015) $z \sim 2$ dust law to correct the stellar contributions (Hα continuum and optical colours). A comparison between the distribution of galaxies obtained with different dust laws for a given f is shown by Fig. 7. The fraction of galaxies with ΔEW $>2\sigma$ from the $\Gamma = -1.35$ IMF track with $f = 1$ ($f = 2$) dust corrections are ∼20 per cent(∼45 per cent), ∼35 per cent(∼75 per cent) and ∼15 per cent(∼55 per cent) for Calzetti et al. (2000), Pei (1992) SMC and Reddy et al. (2015) dust laws, respectively. However, we refrain from interpreting the differences in the distributions of the sample between the considered dust laws because the attenuation values used in the ZFIRE/ZFOURGE surveys have been derived from SED fitting by FAST using a Calzetti et al. (2000) dust law. Compared to the adopted dust law, the change in the value of f has a stronger influence on the galaxies in our parameter space and can significantly affect the EW values, which is discussed further in Section 4.4.

To investigate differences between our $z \sim 2$ sample with HG08 $z \sim 0$ sample, we derive dust corrections to the $(g - r)_{0.1}$ colours. Using the following equations to apply dust corrections to $g_{0.1}$ and $r_{0.1}$ fluxes we recalculate the $(g - r)_{0.1}$ colours for the ZFIRE-SP sample.

$$f(g_i)_{0.1} = f(g_{\rm obs})_{0.1} \times 10^{0.4 \times 1.25 A_c(V)} \quad (5a)$$

$$f(r_i)_{0.1} = f(r_{\rm obs})_{0.1} \times 10^{0.4 \times 0.96 A_c(V)}. \quad (5b)$$

We show the Hα EW versus $(g - r)_{0.1}$ colour comparison between ZFIRE and SDSS samples in Fig. 8. The dust corrections for the ZFIRE-SP sample have been performed using an $f = 1$ and an $f = 2$. Similar to the [340]–[550] colour relationship, there is a significant presence of galaxies with extremely high Hα EW values and ∼60 per cent of the galaxies lie above the Salpeter IMF track when dust corrections are applied with an $f = 2$. Furthermore, the $z \sim 2$ sample shows much bluer colours compared to HG08 sample, which we attribute to the younger ages (∼850 Myr inferred from tracks with a Salpeter IMF) and the higher SFRs of galaxies at $z \sim 2$.

In Fig. 9, we use the $\Gamma = -1.35$ IMF tracks to compute the deviation of observed Hα EW values from a canonical Salpeter like IMF. For each $(g - r)_{0.1}$ galaxy colour, we calculate the expected Hα EW using the standard PEGASE model computed using an exponential decaying SFH with a $p_1 = 1000$ Myr. We then calculate the deviation between the observed values to the expected values. Only the $f = 2$ scenario is considered here to be consistent with the dust corrections applied by HG08. Our results suggest that the ZFIRE sample exhibits a lognormal distribution with a mean and a standard deviation of 0.090 and 0.321 units, respectively. Similarly for the HG08 sample, the values are distributed with a mean and a standard deviation of −0.032 and 0.250 units. Compared to HG08, the ZFIRE-SP sample shows a larger scatter and favours higher Hα EW values for a given Salpeter-like IMF. A simple two sample K-S test for the ZFIRE-SP sample and HG08 gives a K_s statistic of 0.37 and a P value of 1.32×10^{-12}, which suggests that the two samples are distinctively different from each other. In subsequent sections, we further explore whether the differences between the $z \sim 0$ and $z \sim 2$ populations are driven by IMF change or other SP parameters.

4.2 IMF dependence of extinction values

Dust corrections applied to the ZFIRE-SP sample, as explained in Section 4.1, are derived from FAST (Kriek et al. 2009) using best-fitting model SEDs to ZFOURGE photometric data. FAST uses a grid of SED template models to fit galaxy photometric data to derive the best-fitting redshift, metallicity, SFR, age and A_v values for the galaxies via a χ^2 fitting technique. Even though these derived properties may show degeneracy with each other (see Conroy 2013 for a review), in general FAST successfully describes observed galaxy properties of deep photometric redshift surveys (Whitaker et al. 2011; Skelton et al. 2014; Straatman et al. 2016). FAST has a limited variety of stellar templates and therefore we

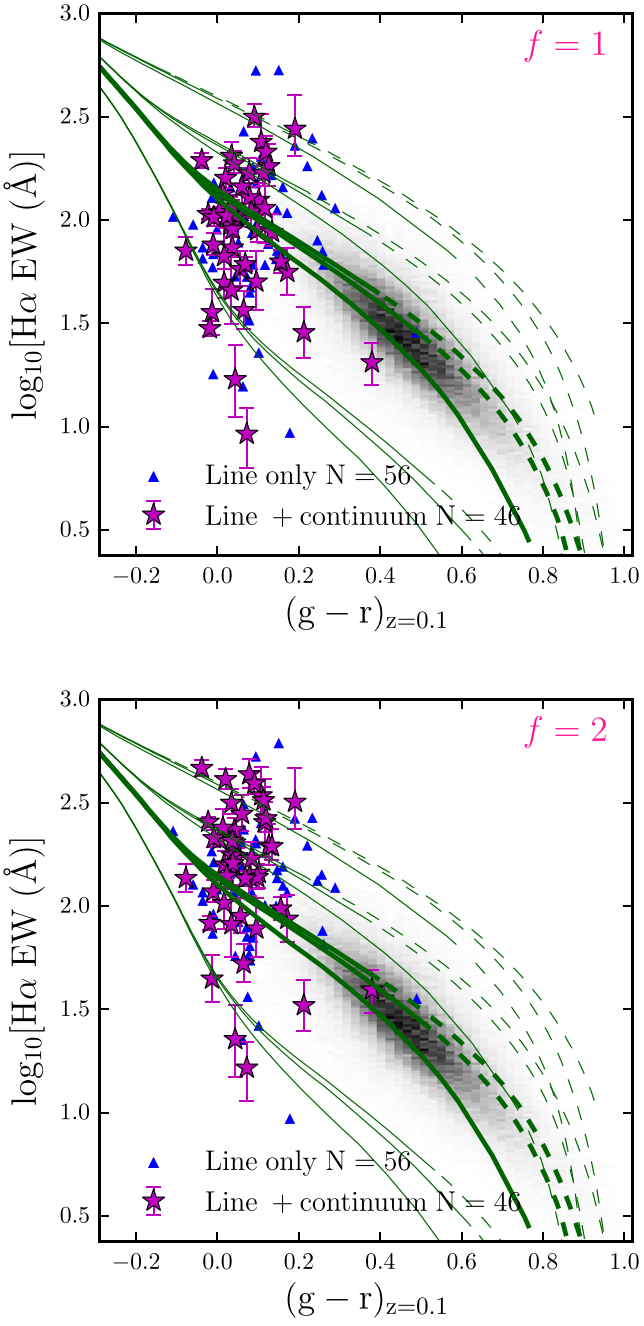

Figure 8. Comparison of the Hα EW and $(g-r)_{0.1}$ colours of the $z\sim2$ ZFIRE-SP sample with the HG08 $z\sim0$ sample. The HG08 sample is shown by the 2D grey histogram. The PEGASE models shown correspond to varying IMFs: from top to bottom $\Gamma=-0.5$, -1.0, -1.35 and -2.0. Similar to Fig. 5, for each IMF we show three models with exponentially declining SFHs with varying p_1 values (from top to bottom $p_1=1500$, 1000 and 500 Myr). Model tracks at $t>3200$ Myr are shown by the dashed lines. *Top:* ZFIRE-SP sample with dust corrections applied with an $f=1$. *Bottom:* ZFIRE-SP sample with dust corrections applied with an $f=2$. Note that HG08 uses an $f=2$ in their dust corrections.

cannot explore the effect of varying IMFs on the FAST-derived extinction values.

In order to examine the role of IMF on derived extinction values, we compare the distribution of ZFIRE rest-frame UV and optical colours with PEGASE model galaxies. Following the same procedure used to derive the [340] and [550] filters, we design two box-car

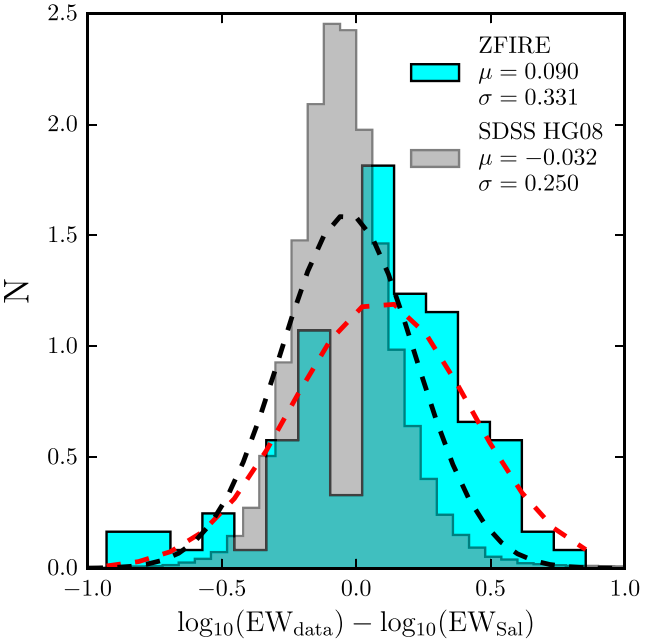

Figure 9. Deviations of the observed Hα from the canonical Salpeter like IMF tracks in the Hα EW versus $(g-r)_{0.1}$ colour space. We show the $z\sim0$ SDSS HG08 sample (grey–black) and the $z\sim2$ ZFIRE-SP sample (cyan–red). Both histograms are normalized to an integral sum of 1 and the best-fitting Gaussian functions are overlaid. The parameters of the Gaussian functions are shown in the legend.

filters centred at 1500 Å ([150]) and 2600 Å ([260]) with a length of 675 Å. The wavelength regime covered by these two filters approximately correspond to the *B* and *I* filters in the observed frame for galaxies at $z\sim2$ (further information is provided in Appendix B2). Therefore, *K*-corrections are small and the computed values are robust.

By binning galaxies in stellar mass, we find massive galaxies to be dustier than their less massive counterparts. We show the distribution of our sample in the rest-frame UV versus rest-frame optical parameter space in Fig. 10 (left-hand panel). PEGASE model galaxies with $\Gamma=-1.35$ and varying SFHs are shown by the solid model tracks. When we apply an $A_v=1$ extinction, the models show a strong diagonal shift due to reddening of the colours in both axis. For each set of tracks, we perform a best-fitting line to the varying SFH models. The dust vector (shown by the arrow) joins the two best-fitting lines drawn to the models with $A_v=0$ and $A_v=1$ at time *t*. We define A_v(ZF) to be the correction needed for each individual galaxy to be brought down parallel to the dust vector to the best-fitting line with $A_v=0$, and is parametrized by the following equation:

$$A_v(\mathrm{ZF}) = -0.503\times([340]-[550]) + 1.914\times([150]-[260]) + 0.607. \quad (6)$$

Our simple method of dust parametrization is similar to the technique used by FAST, which fits SED templates to the UV continuum to derive the extinction values. The [150]–[260] colour probes the UV continuum slope, which is ultrasensitive to dust, while the [340]–[550] probes the optical continuum slope which is less sensitive to dust. Hα emission does not fall within these filters, and hence, is not strongly sensitive to the SFR of the galaxies.

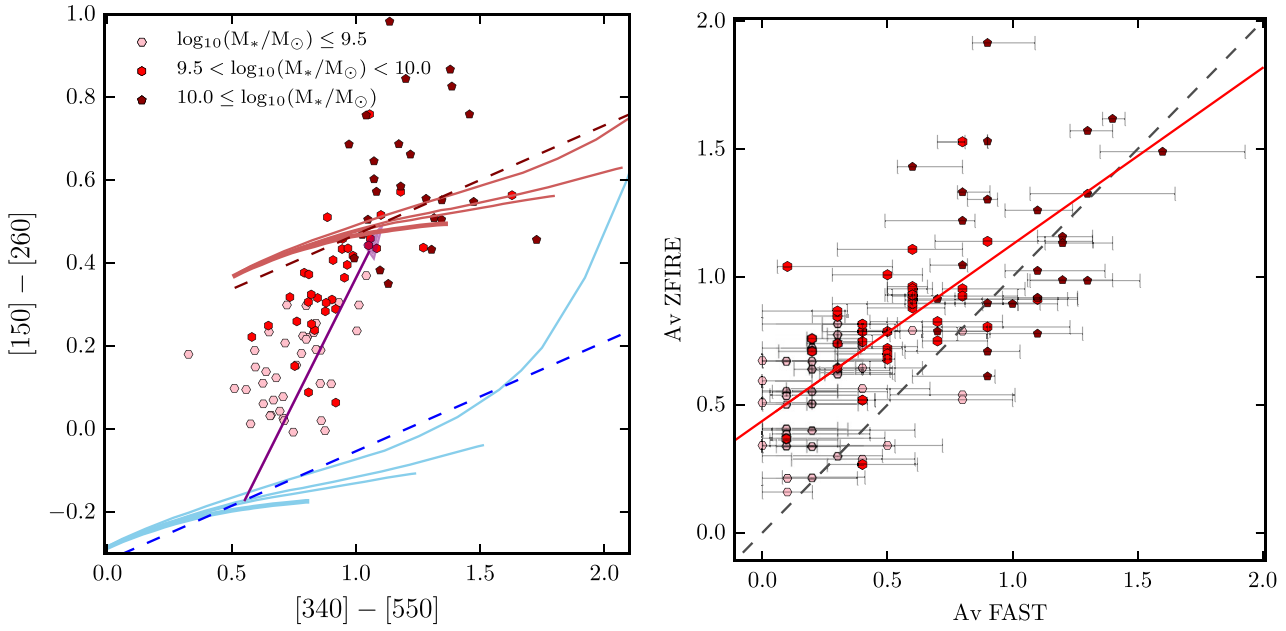

Figure 10. Dust parametrizations to investigate IMF dependences of dust extinction values. Galaxies are divided into three mass bins. *Left:* the dust content of the galaxies are parametrized using UV/optical colours. The cyan solid lines are PEGASE models for $\Gamma = -1.35$ IMF with varying SFHs. The dashed blue line is the best-fitting line for these models. The pink lines are similar to the cyan lines, but with an extinction of 1 mag. The magenta dashed line is the best-fitting line for these models with $A_v = 1$. The purple arrow denotes the direction of the dust vector and connects the two best-fitting lines at time *t*. *Right:* comparison between the extinction derived by equation (6) with the extinction values derived by FAST. The error bars are the upper and lower 68th percentile of the A_v values compiled by FAST and the diagonal red line is the error-weighted least-squares fit to the data. The diagonal dashed line is the A_v(SED) = A_v(ZF) line.

In the right-hand panel of Fig. 10, we compare the derived extinction values from our method [A_v(ZF)] with the extinction values derived by FAST [A_v(SED)]. Since Chabrier (2003) IMF at $M_* > 1\,M_\odot$ is similar to the slope of Salpeter IMF ($\Gamma = -1.35$), the comparison is largely independent of the IMF. The median and σ_{NMAD} scatter of the A_v values derived via FAST and our method is $\sim$−0.3 and $\sim$−0.3, respectively. Therefore, the values agree within 1σ. There is a systematic bias for A_v(ZF) to overestimate the extinction at lower A_v(SED) values and underestimate at higher A_v(SED) values. We attribute this residual pattern to age metallicity degeneracy, which is not considered in the derivation of A_v(ZF).

The choice of IMF will affect dust corrections derived from UV photometry (using FAST or our empirical method) as there is a modest dependence of the rest-frame UV continuum slope on IMF for star-forming populations. We are primarily interested in IMF slopes shallower than Salpeter slope ($\Gamma > -1.35$) to explain our population of high Hα EW galaxies. For $\Gamma = -0.5$, we find the best-fitting model line in the left-hand panel of Fig. 10 shifts down by $\sim$0.1 mag. This increases the magnitude of dust corrections and extends the arrows in Fig. 6 to bluer colours and higher EWs and does not explain the presence of high Hα EW objects. For the purpose of comparing with our default hypothesis (Universal IMF with $\Gamma = -1.35$), we adopt the FAST-derived dust corrections.

4.3 Balmer decrements

Stellar attenuation values computed by fitting a slope to galaxy SEDs in UV, estimates the extinction of old SPs that primarily contributes to the galaxy continuum. Nebular emission lines originate from hot ionized gas around young and short-lived O and B stars. Given their short lifetime ($\sim$10–20 Myr), O and B stars are not expected to move far from their birthplace (dusty clouds), thus, the nebular emission lines are expected to have high levels of extinction. Next, we investigate the dust properties of the stars in different star-forming environments using the luminosity ratios of nebular hydrogen lines and observed UV colours.

Luminosity ratios of nebular hydrogen lines are insensitive to the underlying SP and IMF parameters for a fixed electron temperature (Osterbrock 1989). These line ratios are governed by quantum mechanics and therefore can be used to probe the reddening of nebular emission lines and dust geometry under the assumption that ionized gas attenuation resembles that of the underlying SP.

With the recent development of sensitive NIR imagers and multi-object spectrographs, studies have now started to investigate the properties of dust at $z \sim 2$ (Reddy et al. 2015; Shivaei et al. 2015; de Barros, Reddy & Shivaei 2016). These studies show conflicting results on the fraction of stellar to nebular attenuation of galaxies at $z \sim 2$. Here, we show Balmer decrement results for a subsample of our ZFIRE-SP sample which shows SNR >5 detections for both Hα and Hβ. The data presented herein are a combination of data released by the ZFIRE data release Nanayakkara et al. (2016) and additional MOSFIRE observations carried out during 2016 January. Our sample comprises of 42 galaxies with both Hα and Hβ emission line detections with an SNR >5 and 35 galaxies are part of the ZFIRE-SP sample. Further details on Hβ detection properties are explained in Appendix D1.

We show the Hβ flux versus Hα flux for our total ZFIRE galaxies Fig. 11 (left-hand panel). The diagonal dashed line of the left-hand panel shows the Balmer decrement for Case B recombination models with Hα/Hβ= 2.86 (Osterbrock 1989). Galaxies that fall below this criteria are expected to have realistic dust models. In Fig. 11 (right-hand panel), we show the comparison between extinction computed for stars by FAST with the extinction computed for ionized gas regions using the Balmer decrement.

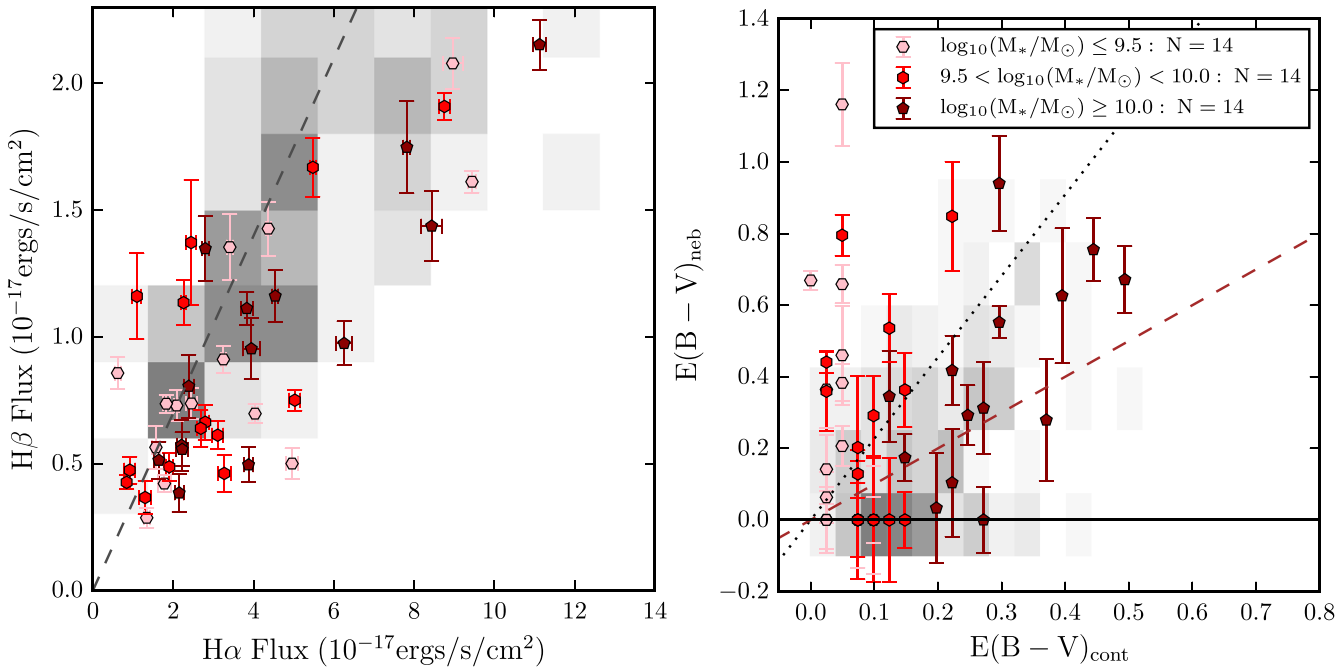

Figure 11. Balmer decrement properties of the ZFIRE sample. Here, we show the subset of Hα- and Hβ-detected galaxies in our ZFIRE-SP sample. The 2D density histogram shows the distribution of values from Reddy et al. (2015) *Left:* Hβ flux versus Hα flux measurements for the ZFIRE-SP sample. The individual galaxies are divided into three mass bins. The diagonal dashed line is the $f(\mathrm{H}\alpha) = 2.86 \times f(\mathrm{H}\beta)$ line which denotes the Balmer decrement for case B recombination. *Right:* comparison between SED-derived extinction values with extinction computed from the Balmer decrement. The black dotted line is the $E(B-V)_{\rm neb} = E(B-V)_{\rm cont}/0.44$ relationship expected from Calzetti et al. (2000) to compute the extra extinction for nebular emission line regions. The brown dashed line is the $E(B-V)_{\rm neb} = E(B-V)_{\rm cont}$ line and the horizontal solid black line is the $E(B-V)_{\rm neb} = 0$ line. Galaxies with $E(B-V)_{\rm neb} < 0$ have been assigned a value of 0.

The colour excess is computed from the Balmer decrement using

$$E(B-V)_{\rm neb} = \frac{2.5}{1.163} \times \log_{10}\left\{\frac{f(\mathrm{H}\alpha)}{2.86 \times f(\mathrm{H}\beta)}\right\}. \quad (7)$$

The distribution of our sample in these panels is similar to Reddy et al. (2015) results as shown by the 2D density histogram and therefore highlights the complicated dust properties of $z \sim 2$ star-forming galaxy populations.

4.4 The difference in extinction between stellar and nebular regions

In this section, we investigate how the differences in dust properties between stellar and ionized gas regions can affect the distribution of our galaxies in Hα EW versus [340]–[550] colour space. Calzetti et al. (2000) showed that, for $z \sim 0$ star-forming galaxies, the nebular lines are $\sim\times2$ more attenuated than the continuum regions, but at $z \sim 2$ studies show conflicting results (Reddy et al. 2015; Shivaei et al. 2015; de Barros et al. 2016). Using ZFIRE data in Fig. 11, we show that galaxies occupy a large range of f values in our Hβ-detected sample. We attribute the scatter in extinction to the properties of sight-lines of the nebular line regions.

Since the universe is only $\sim$3 Gyr old at $z \sim 2$, the dense molecular clouds collapsed to form stars would only have had limited time to evolve into homogeneous structures within galaxies. This can give rise to differences in the dust geometries within ionizing clouds resulting in non-uniform dust sight-lines for galaxies at $z \sim 2$. By varying the value of f as a free parameter, we calculate the f values required for our galaxies to be consistent with a universal IMF with slope $\Gamma = -1.35$. For each dust-corrected [340]–[550] colour, we compute the Hα EW of the PEGASE $\Gamma = -1.35$ IMF track with $p_1 = 1000$ Myr. We then use the observed and required Hα EW values to compute the f as follows:

$$f = \left\{\frac{\log_{10}(\mathrm{H}\alpha\ \mathrm{EW})_{\Gamma=-1.35} - \log_{10}(\mathrm{H}\alpha\ \mathrm{EW})_{\rm obs}}{0.44 \times 0.62 \times A_{\rm c}(v)} + \frac{0.82}{0.62}\right\}, \quad (8)$$

where $\log_{10}(\mathrm{H}\alpha\,\mathrm{EW})_{\Gamma=-1.35}$ is the Hα EW of the PEGASE model galaxy for dust-corrected [340]–[550] colours of our sample and $\log_{10}(\mathrm{H}\alpha\,\mathrm{EW})_{\rm obs}$ is the observed Hα EW.

In Fig. 12, we show the f values required for our galaxies to agree with a universal IMF with a slope of $\Gamma = -1.35$. For the 46 continuum-detected galaxies, $\sim$30 per cent show $f < 1$. It is extremely unlikely that galaxies at $z \sim 2$ would have $f < 1$, which suggests that ionizing dust clouds where the nebular emission lines originate from are less dustier than regions with old SPs. Furthermore, galaxies that lie above the Salpeter track [$\log_{10}$(Hα EW) > 2.2] requires $f < 1$ and therefore even a varying f hypothesis cannot account for the high-EW galaxies. $\sim$17 per cent of continuum detections have $f < 0$ which is not physically feasible since it requires dust to have the opposite effect to attenuation. Therefore, we reject the hypothesis that varying f values could explain the high Hα EWs of our galaxies.

4.5 Observational bias

The ZFIRE-SP sample spans a large range of Hα EWs, suggesting a considerable variation in the sSFRs of the ZFIRE galaxies at $z \sim 2$.

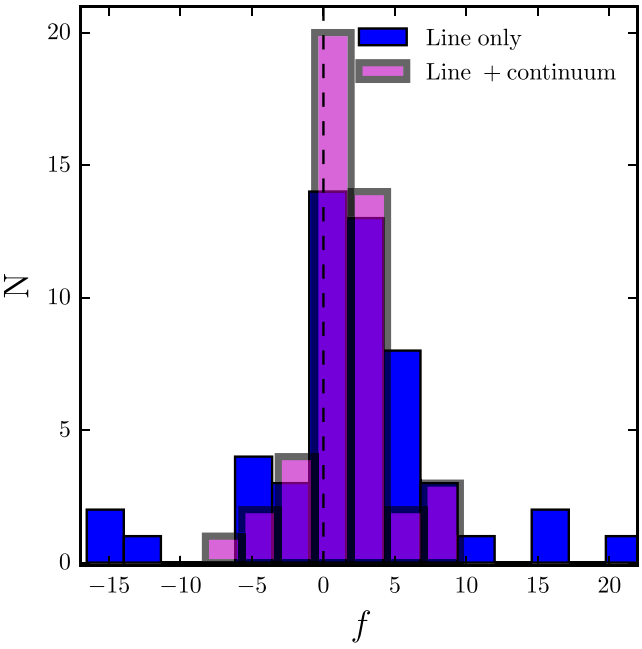

Figure 12. The distribution of f values required for galaxies in the ZFIRE-SP sample to agree with a $\Gamma = -1.35$ Salpeter like IMF. For each dust-corrected [340]–[550] colour of the ZFIRE-SP sample galaxies, we use the corresponding Hα EW of the $\Gamma = -1.35$ PEGASE track with an exponentially declining SFH with a $p_1 = 1000$ Myr to compute the f value required for the observed Hα EW to agree with the $\Gamma = -1.35$ IMF. The vertical dashed line is the $f = 0$ line.

High Hα EW can result due to two reasons:

(i) High line flux: suggests a higher SFR in time-scales of $\sim$10 Myr.
(ii) Lower continuum level: suggests lower stellar mass for galaxies.

These two scenarios should be considered together: i.e. a higher line flux with lower continuum level would suggest the galaxy to be going through an extreme star formation phase. We investigate any detection bias that could explain our distribution of Hα EWs.

In Nanayakkara et al. (2016), we show that the ZFIRE COSMOS K-band detections are mass complete to $\log_{10}(M_*/\mathrm{M}_\odot) \sim 9.3$. In Fig. 13, we show the distribution of the Hα flux and continuum levels of our sample. It is evident from Fig. 13 (left-hand panel) that our galaxies evenly sample the star-forming main sequence described by Tomczak et al. (2014) without significant bias towards extreme Hα flux values. Therefore, we conclude that the Hα fluxes we detect are typical of star-forming galaxies at $z = 2.1$.

In Fig. 13 (right-hand panel), we compare our Hα flux values with the derived continuum levels. Continuum-detected galaxies show continuum levels that are in the order of $\sim$2 mag smaller compared to the Hα fluxes and therefore the higher Hα EWs in our sample are primarily driven by the low continua. Note that our continuum detection level is ~ -2.3 log flux units. Therefore, for galaxies with only line detection, the difference between Hα flux and continuum level is much higher, which suggests much larger Hα EWs.

Several studies investigated the Hα EW of galaxies at higher redshifts ($z \geq 1.5$) (Erb et al. 2006; Shim et al. 2011; Fumagalli et al. 2012; Kashino et al. 2013; Stark et al. 2013; Masters et al. 2014; Sobral et al. 2014; Speagle et al. 2014; Marmol-Queralto et al. 2016; Rasappu et al. 2016) using SED fitting techniques and/or grism spectra. We find that our Hα EWs show good agreement with EWs expected at $z \sim 2$ (Marmol-Queralto et al. 2016) and conclude that our observed Hα EW values are typical of $z \sim 2$ galaxies.

However, there are no studies that use high-quality spectra to study the Hα EW at $z \sim 2$. Even though our Hα fluxes and EWs are typical of $z \sim 2$ galaxies, in the Hα EW versus [340]–[550] colour space, a large fraction of our galaxies show high EWs for a given [340]–[550] colour compared to the expectation by a Salpeter-like IMF. Our high EWs are driven by lower continuum levels, for which we consider two possible explanations.

(i) Most galaxies have quenched their starburst phase in a time-scale of $\sim$10 Myr. Therefore, the old SPs are still being built up explaining the lack of continuum level from the older stars.
(ii) Stars are being formed continuously at $z \sim 2$ with a higher fraction of high-mass stars.

In Section 5, we investigate the effects of starbursts on our study to examine how probable it is for $\sim$1/3rd of our galaxies to have quenched their star formation within a time-scale of $\lesssim$10 Myr.

5 CAN STARBURSTS EXPLAIN THE HIGH Hα EWS?

Galaxies at $z \sim 2$ are at the peak of their SFH (Hopkins & Beacom 2006). We expect these galaxies to be rapidly evolving with multiple stochastic star formation scenarios within their SPs. If our sample consists of a significant population of starburst galaxies, it may cause significant systemic biases to our IMF analysis.

In this section, we investigate the effects of bursts on the SFHs of the galaxies. We study how the distribution of galaxies in Hα EW versus [340]–[550] colour space may be affected by such bursts and how we can mitigate their effects. We demonstrate that our final conclusions are not affected by starbursts.

5.1 Effects of starbursts

A starburst event would abruptly increase the Hα EW of a galaxy within a very short time-scale ($\lesssim$5 Myr). The increase in ionizing photons is driven by the extra presence of O and B stars during a starburst which increases the amount of Lyman continuum photons. Assuming that a constant factor of Lyman continuum photons get converted to Hα photons via multiple scattering events, we expect the number of Hα photons to increase as a proportion to the number of O and B stars. Furthermore, the increase of the O and B stars would drive the galaxy to be bluer causing the [340]–[550] colours to decrease.

The ability of a starburst to drive the points away from the monotonic Salpeter track is limited. The deviations are driven by the burst fraction, which we define as the burst strength divided by the length of the starburst. If the burst fraction is small, it has a small effect. However, if it is very large it dominates both the Hα and the optical light, the older population is 'masked', and it heads back towards the track albeit at a younger age, i.e. one is seeing the monotonic history of the burst component. The maximum deviation in our study occurs for burst mass fractions of 20–30 per cent occurring in time-scales of 100–200 Myr or fractions of thereof, which can cause excursions of up to $\sim$1 dex. However, as we will see this only occurs for a short time.

We show the effect of a starburst on a PEGASE model galaxy with a monotonic SFH in Fig. 14. A starburst with a time-scale of $\tau_b = 100$ Myr and a strength $f_\mathrm{m} = 0.2$ (fraction of mass at

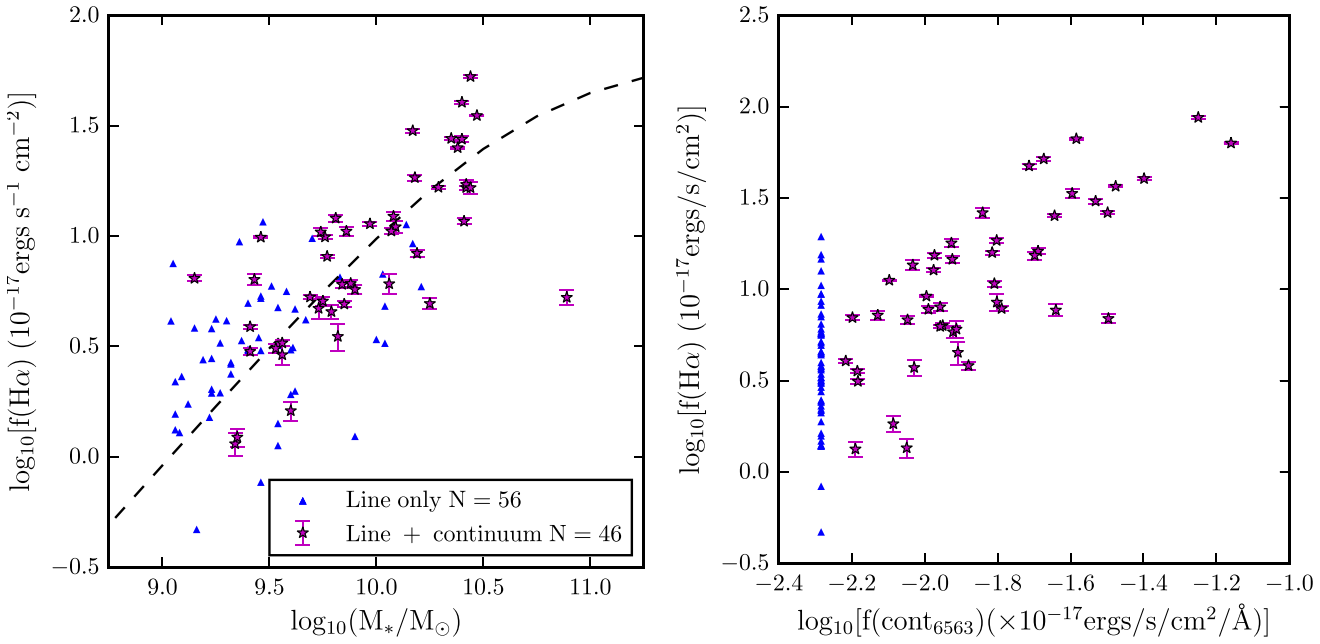

Figure 13. Here, we investigate which observable parameter/s drive the high Hα EW values compared to $\Gamma = -1.35$ Salpeter-like IMF expectations in Hα EW versus [340]–[550] colour space. *Left:* the logarithmic Hα flux of the ZFIRE-SP sample as a function of stellar mass. The Hα flux has been dust corrected following equation (D6a) with $f = 1/0.44$. The black line is derived from the star formation main sequence from Tomczak et al. (2014), converted to Hα flux using Kennicutt (1998b) Hα star formation law at $z = 2.1$. *Right:* the Hα flux versus the continuum level at 6563 Å. Both parameters [Hα flux as above, continuum level following equation (D6b)] have been corrected for dust extinction and are plotted in logarithmic space. The Hα fluxes are ∼2 orders of magnitude brighter than the continuum levels.

∼3000 Myr generated during the burst) is overlaid on the constant SFH model at time = 1500 Myr. The starburst drives the increase of Hα EW which occurs in a very short time-scale. In Fig. 14, the galaxy deviates from the constant SFH track as soon as the burst occurs and reaches a maximum Hα EW within 4 Myr. At this point, the extremely high mass stars made during the burst will start to leave the main sequence. This will increase the number of red giant stars resulting in higher continuum level around the Hα emission line. Therefore, the Hα EW starts to decrease slowly after ∼4 Myr. Once the burst stops, the Hα EW drops rapidly to values lower than pre-burst levels. The galaxy track will eventually join the $\Gamma = -1.35$ smooth SFH track at a later time than what is expected by a smooth SFH model.

We further investigate the effect of starbursts with smaller time-scales ($t_b < 20$ Myr) and find that the evolution of Hα EW in the aftermath of the burst to be more extreme for similar $f_{\rm m}$ values. This is driven by more intense star formation required to generate the same amount of mass within ∼1/10th of the time-scale. Since both Hα EW and [340]–[550] colours are a measure of sSFR, we expect the evolution to strongly dependent on $f_{\rm m}$ and τ_b of the burst and to be correlated with each other.

To consider effects of bursts, we adopt two complimentary approaches. First, we stack the data in stellar mass and [340]–[550] colour bins. SPs are approximately additive and by stacking we smooth the stochastic SFHs in individual galaxies and also account the effect from galaxies with no Hα detections. Secondly, we use PEGASE to model starbursts to generate Monte Carlo simulations to predict the distribution of the galaxies in Hα EW versus [340]–[550] colour space. Using the simulations, we investigate whether it is likely that the observed discrepancy is driven by starbursts and also double check whether the stacking of galaxies would generate smooth SFH models.

5.2 stacking

In order to remove effects from stochastic SFHs of individual galaxies in our sample, we employ a spectral stacking technique. We first divide the galaxies into three mass and dust-corrected [340]–[550] colour bins as follows.

(i) Mass bins: $\log_{10}(M_*/{\rm M}_\odot) \leq 9.5$, $9.5 < \log_{10}(M_*/{\rm M}_\odot) < 10$, $\log_{10}(M_*/{\rm M}_\odot) \geq 10$
(ii) [340]–[550] colour bins: ([340]–[550]) ≤ 0.56, 0.56 < ([340]–[550]) < 0.65, ([340]–[550]) ≥ 0.65

We select a wavelength interval of ∼1500 Å centred around the Hα emission for each spectra and mask out the sky lines with approximately 2× the spectral resolution. In order to avoid systematic biases arisen from narrowing down the sampled wavelength region in the rest frame, we instead redshift all spectra to a common $z = 2.1$ around which most of the galaxies reside. We sum all the spectra at this redshift, in their respective bins. The error spectra are stacked in quadrature following standard error propagation techniques.

We mask out the nebular emission line regions of the stacked spectra and use a sigma-clipping algorithm to fit a continuum (c1). The error in the continuum is assigned as the standard deviation of the continuum values of 1000 bootstrap iterations.

We visually inspect the stacked spectra to identify the Hα emission line profiles to calculate the integrated flux. Stacked Hα EW is calculated following equations (2a) and (2b).

To estimate the error on the stack due to the stochastic variations between galaxies, we use a bootstrapping technique to calculate the error of the stacked Hα EW values. We bootstrap galaxies with replacement in each bin to produce 1000 stacked spectra for each of which we calculate the Hα EW. The standard deviation of the logarithmic EW values for each bin is considered as the error of

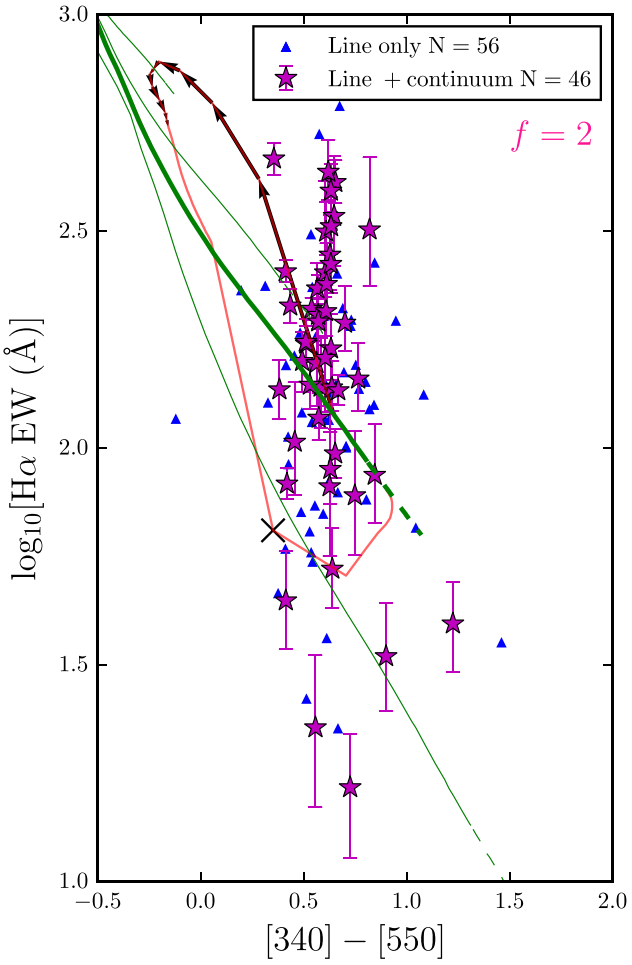

Figure 14. The effect of a starburst on a PEGASE model galaxy track. The green tracks are computed with constant SFHs but with different IMFs. From top to bottom they have Γ values of, respectively, −0.5, −1.0, −1.35 and −2.0. All solid tracks end ∼3.1 Gyr and the continuation up to 13 Gyr is shown by the dashed lines. The track in red follows the constant SFH model with an IMF slope Γ = −1.35 (thick green track) up to $t = 1500$ Myr. A burst with a duration of $t_b = 100$ Myr is superimposed on the track at $t = 1500$ Myr. The burst generates 30 per cent of the galaxies' total mass at ∼3000 Myr. The black arrows superimposed on the tracks show the direction of the burst in the first 10 Myr and are plotted every 1 Myr to distinctively demonstrate the short time-scales in Hα EW evolution on the aftermath of a starburst. The cross denotes $t = 1600$ Myr signalling the end of the burst. The ZFIRE-SP sample is dust corrected with an $f = 2$ following prescriptions outlined in Section 4.1. The magenta stars show galaxies with continuum detections while solid blue triangles show galaxies only with Hα emission lines.

the Hα EW of the stacked spectra. We expect the bootstrap errors to include stochastic variations in the SFHs between galaxies. If our sample comprises of galaxies undergoing extreme starbursts, the effects from such bursts should be quantified within these error limits.

We stack the individual [340] and [550] fluxes of the galaxies in similar mass and [340]–[550] colour bins. The average extinction value of the galaxies in each bin is considered as the extinction of the stacked spectra. We use this extinction value to dust correct the Hα EW and [340]–[550] colours of the stacked spectra, following recipes explained in Section 4.1.

Fig. 15 shows the distribution of the stacked spectra in Hα EW versus [340]–[550] colour space before and after dust corrections are applied. We consider dust corrections with $f = 1$ and $f = 2$. For $f = 1$ dust correction scenario, the bluest colour bin and the medium and high mass stacked data points agree with the Γ = −1.35 track. However, the redder colours bins and lowest mass galaxies on average prefer shallower IMFs. With $f = 2$ dust correction, the distribution of stacked data points even with bootstrap errors suggest values ∼0.1–0.4 dex above the Salpeter IMF, if interpreted as an IMF variation. In both scenarios, redder galaxies show larger deviation from the canonical Salpeter IMF. If starbursts drive the distribution of the galaxies, we expect the bluer galaxies on average to show larger deviation from the Γ = −1.35 tracks.

To further account for any detection bias arisen from Hα undetected galaxies, we use the ZFIRE COSMOS field *K* band targeted galaxies with no Hα emission detections to compute a continuum (c2) contribution to the stacked spectra. We use the photometric redshifts to select 37 galaxies within $1.90 < z < 2.66$, which is the redshift interval the Hα emission line falls within the MOSFIRE *K* band. Next, we perform a cut to select galaxies with similar stellar masses ($9.04 < \log_{10}(M_{\odot}) < 10.90$) and [340]–[550] colours ($-0.12 <$ ([340]–[550]) < 1.46) to the galaxies in the ZFIRE-SP sample. The final sample comprises of 21 galaxies which we use to stack the 1D spectra in mass and [340]–[550] colour bins.

In order to stack the spectra, first we mask out the sky regions and assume that all galaxies are at a common $z = 2.1$. We mask out Hα and [N II] emission line regions and fit a continuum similar to how c1 was derived. We add c1+c2 to re-calculate the Hα EW for each of the mass and colour bins. Since the continuum level is increased by the addition of c2, the Hα EW of the spectra reduces. We note that the highest mass bin contains no Hα undetected galaxies. Fig. 16 shows the change in stacked data points when the Hα non-detected continuum is considered with a dust correction of $f = 1$ and $f = 2$. The maximum deviation of the stacked Hα EW values is ∼0.2 dex and the lowest mass and the reddest [340]–[550] colour bins show the largest deviation. This is driven by the higher number of lower mass redder galaxies which have been targeted but not detected by the ZFIRE survey. The magnitude of the deviations are independent of the *f* value used for the dust corrections and for both $f = 1$ and $f = 2$, the galaxies that show an excess of Hα EW compared to Γ = 1.35 tracks still show an excess when the added c2 continuum contribution is considered. For $f = 2$ dust corrections, even with considering the effect of non-detected continuum levels, majority of the stacked galaxies in our sample are significantly offset from the canonical Salpeter like IMF value.

5.3 Simulations of starbursts

By employing spectral stacking and bootstrap techniques, we showed in Section 5.2 that our galaxies on average favour shallower IMFs than the universal Salpeter IMF. In this section, we use PEGASE SSP models to generate simulations with starbursts to calculate the likelihood for half of the ZFIRE-SP sample to be undergoing simultaneous starbursts at $z \sim 2$. Furthermore, by randomly selecting galaxies from the simulation at random times, we stack the galaxies in mass and [340]–[550] colour bins to make comparisons with the stack properties of the ZFIRE-SP sample.

We empirically tune our burst parameters to produce the largest number of galaxies above the Salpeter track. A single starburst with time-scales of $t_b \sim 100$–300 Myr and with $f_m \sim 0.1$–0.3 are overlaid on constant SFH models with the starburst occurring at any time between 0 and 3250 Myr of the galaxies' lifetime. Simulation

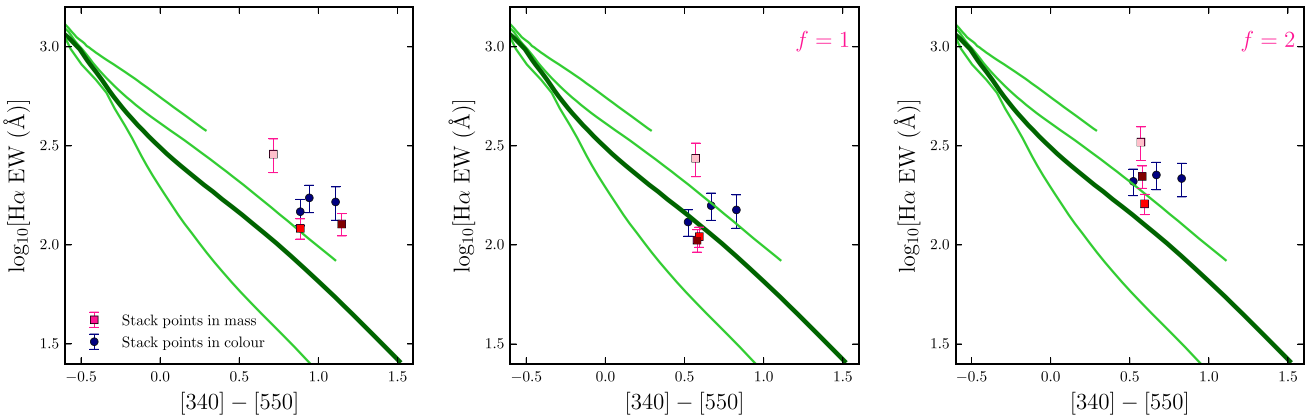

Figure 15. The Hα EW versus [340]–[550] colour distribution of the stacked ZFIRE-SP sample. The spectra are stacked in mass (squares) and [340]–[550] colour (circles) bins. The errors are from bootstrap re-sampling of the objects in each bin. The mass bins are colour coded where the higher masses have darker colours. The tracks are SSP models computed from PEGASE with exponentially decaying SFHs of $p_1 = 1000$ Myr and varying IMFs. From top to bottom the tracks have Γ values of, respectively, -0.5, -1.0, -1.35 (thick dark green line) and -2.0. All tracks end $\sim$3.1 Gyr ($z \sim 2$). *Left:* stacked galaxies before any dust corrections are applied. *Centre:* stacked galaxies after dust corrections are applied following recipes outlined in Section 4.1 with an $f = 1$. *Right:* similar to the centre panel but with an $f = 2$.

properties and the evolution of Hα EW and [340]–[550] colours during starbursts are further discussed in Appendix E.

Our final simulation grid contains 8337 possible time-steps, which we use to randomly select galaxies within $2.0 < z < 2.5$ (similar to the time window where our observed sample lies) to perform a density distribution study and a stacking technique similar to the method described in Section 5.2.

To quantify the probability of starbursts dominating the scatter in the Hα EW versus [340]–[550] colour space, we select 10 000 galaxies randomly from the simulated sample to calculate the relative probability galaxies occupy in Hα EW versus [340]–[550] colour space. Fig. 17 (top panels) shows the density distribution of the selected galaxy sample. The relative probability is calculated by normalizing the highest density bin to 100 per cent. To generate real values for the logarithmic densities, we shift the distribution by 0.01 units. As evident from the figure, for both $f = 1$ and $f = 2$ dust corrections, there is a higher probability for galaxies to be sampled during the pre- or post-burst phase due to the very short time-scale the tracks take to reach a maximum Hα EW value during a starburst. $\sim$90 per cent of the galaxies in the ZFIRE-SP sample lie in regions with $\lesssim$0.1 per cent probability. Therefore, we conclude that it is extremely unlikely that $\sim$1/5th ($f = 1$) and 1/3rd ($f = 2$) of the galaxies in the ZFIRE-SP sample to be undergoing a starburst simultaneously and rule out the hypothesis that starbursts could explain the distribution of the ZFIRE-SP sample in the Hα EW versus [340]–[550] colour parameter space.

In Fig. 18, we use our burst machinery to validate our stacking method. We select 100 galaxies randomly with replacement from the parent population of 100 galaxies. For each galaxy, we select a random time to extract the galaxy spectra at the closest sampled time to retrieve SP parameters. We then stack the selected galaxies in stellar mass and [340]–[550] colour bins. The bins are generated in such a way that the selected galaxies are distributed evenly across the bins. We repeat the galaxy selection and stacking process 100 times to calculate bootstrap errors for the stacked data points. Galaxies containing SFHs with bursts stacked in mass and [340]–[550] colour show similar distribution to galaxy tracks with constant SFHs. Even with large t_b values, the time-scale the tracks deviate significantly above the $\Gamma = -1.35$ is in the order of 1–5 Myr and therefore it is extremely unlikely ($\lesssim$5 selected in the stacked sample of 100 galaxies) to preferentially select a large number of galaxies during this phase. Furthermore, stacked errors from bootstrap re-sampling do not deviate significantly from the $\Gamma = -1.35$ tracks. This further strengthens the point that repetitive sampling of galaxies does not yield stacks with higher Hα EW values for a given [340]–[550] colour.

5.3.1 *Smaller bursts*

Galaxies at $z \sim 2$ appear to be clumpy (e.g. Tadaki et al. 2013). Therefore, it is possible for a single clump to have a high SFR that will add a significant contribution to the ionizing flux of a galaxy resulting in higher Hα EW values. To account for such scenarios, we perform the starburst simulations with smaller burst time-scales ($t_b \sim$ 10–30 Myr) and large burst strengths ($f_m \sim$ 0.1–0.3) and allow the galaxies to commence their constant SFH at a random time between 0 and 2500 Myr. We further constrain the bursts to redshifts between $2.0 < z < 2.5$ corresponding to a $\Delta t \sim 650$ Myr, which is similar to the redshift distribution of our galaxies.

As described previously, we randomly select 10 000 galaxies from our simulated sample, but constraining the selection to the redshift window of $2.0 < z < 2.5$. We show the density distribution of our randomly selected sample with the observed ZFIRE-SP sample in Fig. 17 (bottom panels). A large fraction of galaxies are now selected during the post-burst phase, thus with lower Hα EWs compared to the reference IMF track, specially with $f = 1$ dust corrections. Since the starbursts are now short lived but have to generate the same fraction of stellar masses as the longer lived bursts, the fraction of mass generated by the burst per unit time is extremely high. Therefore, changes in Hα EW and optical colours are much more drastic as a function of time and makes it further unlikely to select galaxies with high Hα EWs. We further test scenarios including smaller bursts within short time-scales (see Table E1) and find the distribution of selected galaxies to be similar to 17 (top panels). We conclude that even limiting the starbursts to a narrow redshift

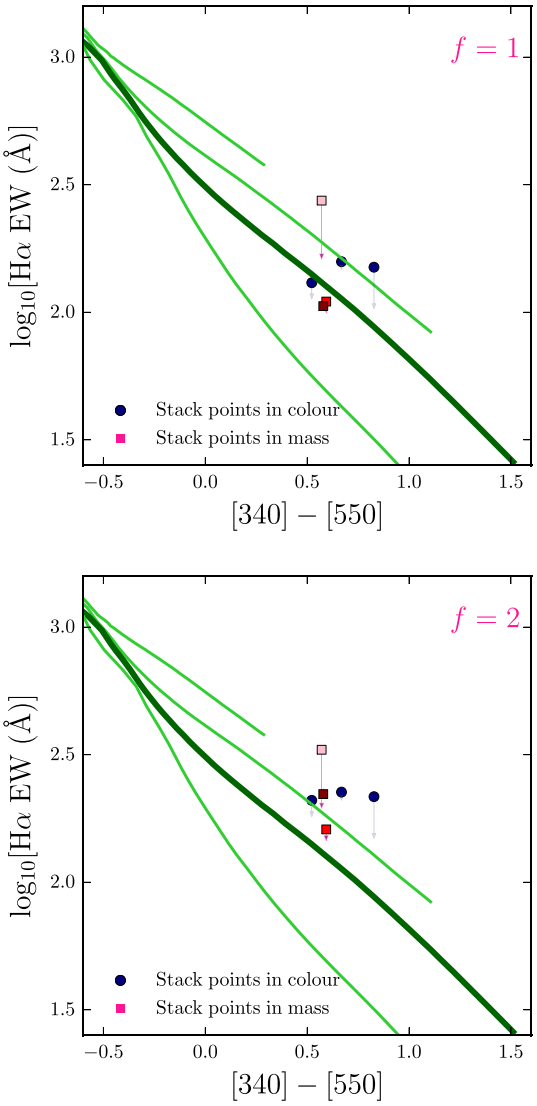

Figure 16. Here, we show the effect of adding the continuum contribution (c2) of galaxies with no Hα detections to the ZFIRE-SP sample stacks. Galaxies are stacked in mass and colour bins and the arrows show the change in EW when the contribution from c2 is added to the data. All tracks shown are similar to Fig. 15. *Top:* dust corrections applied with $f = 1$. The errors for the data points are similar to Fig. 15 (centre panel). *Bottom:* dust corrections applied with $f = 2$. The errors for the data points are similar to Fig. 15 (right-hand panel).

window does not yield a distribution of galaxies that would explain our high Hα EW sample.

6 CONSIDERING OTHER EXOTICA

In the previous sections, we have shown that the distribution of the ZFIRE-SP sample galaxies in the Hα EW versus [340]–[550] colour cannot solely be described by dust or starbursts within a universal IMF framework. In this section, we investigate whether other exotic parameters related to SSP models such as stellar rotations, binary stars, metallicity and the high-mass cut-off could influence the distribution of the galaxies in our parameters to impersonate a varying IMF.

6.1 Stellar rotation

First, we consider effects of implementing stellar rotation in SSP models. Rotating stellar models produce harder ionizing spectra with higher amounts of photons that are capable of ionizing hydrogen. This is driven by rotationally induced larger Helium surface abundances and high luminosity of stars (Leitherer & Ekström 2012), which results in ∼×5 higher ionizing photon output by massive O stars at solar metallicity (Leitherer et al. 2014) and can be $\gtrsim$×10 towards the end of the main-sequence evolution (Szécsi et al. 2015). The minimum initial mass necessary to form W-R stars is also lowered by stellar rotation resulting in longer lived W-R stars (Georgy et al. 2012) increasing the number of ionizing photons. Therefore, stars with rotation shows higher Hα fluxes compared to systems with no rotation, resulting in higher Hα EW values. Additionally, stellar rotation also leads to higher mass-loss in stars, which results in bluer stars in the red supergiant phase (Salasnich, Bressan & Chiosi 1999). Furthermore, stellar models with rotation results in longer lifetimes by 10–20 per cent (e.g. Levesque et al. 2012; Leitherer et al. 2014). This allows a larger build-up of short lives O and B stars compared with similar IMF and SFH models with no rotation resulting in higher Hα flux values and bluer SPs.

S99 supports stellar tracks from the Geneva group (explained in detail in Ekström et al. 2012 and Georgy et al. 2013 and references therein), which allows the user to compute models with and without invoking stellar rotation. Models with stellar rotation assumes an initial stellar rotation velocity ($v_{\rm ini}$) of 40 per cent of the break-up velocity of the zero-age main sequence ($v_{\rm crit}$).

Leitherer et al. (2014) note that $v_{\rm ini} = 0.4 v_{\rm crit}$ for stellar systems is of extreme nature and should be considered as an upper boundary for initial stellar rotation values. $v_{\rm ini}$ is defined as the rotational velocity the star possess when it enter the zero-age main sequence. Depending on stellar properties and interactions with other stars (e.g. de Mink et al. 2013), the initial rotational velocity of the star will be regulated with time (see fig. 12 of Szécsi et al. 2015, where the evolution of stellar rotation has been investigated as a function of time for models with different stellar masses and initial velocities).

A realistic SP will contain a distribution of $v_{\rm ini}/v_{\rm crit}$ values. Levesque et al. (2012) investigated galaxy models with 70 per cent of stars with stellar rotation following $v_{\rm ini} = 0.4 v_{\rm crit}$ and 30 per cent with no stellar rotation, thus allowing more realistic conditions. They found that such models show ∼0.5 dex less hydrogen ionizing photons compared to an SP with all stars with $v_{\rm ini} = 0.4 v_{\rm crit}$ stellar rotation and ∼1.4 dex higher number of hydrogen ionizing photons compared to an SP with no stellar rotation.

The extent of stellar rotation required to describe observed properties of SPs is not well understood. Gravitational torques have been shown to prevent stars from rotating >50 per cent of its break-up velocity during formation (Lin, Krumholz & Kratter 2011). Martins & Palacios (2013) showed that Geneva models with stellar rotation do not reproduce the distribution of massive, evolved stars accurately and require less amounts of convective overshooting thus lowering the required $v_{\rm ini}$. However, recent studies demonstrate the requirement for populations of stars with extreme rotation in low-metallicity scenarios to explain the origin of narrow He emission in galaxies (Gräfener & Vink 2015; Szécsi et al. 2015) and long Gamma-ray bursts (Woosley & Heger 2006; Yoon, Langer & Norman 2006). SPs of the Large Magellanic Cloud have shown to be distributed following a two peak rotational velocity distribution with ∼50 per cent of galaxies rotating at ∼20 per cent of their critical velocities while ∼20 per cent of the population having near-critical

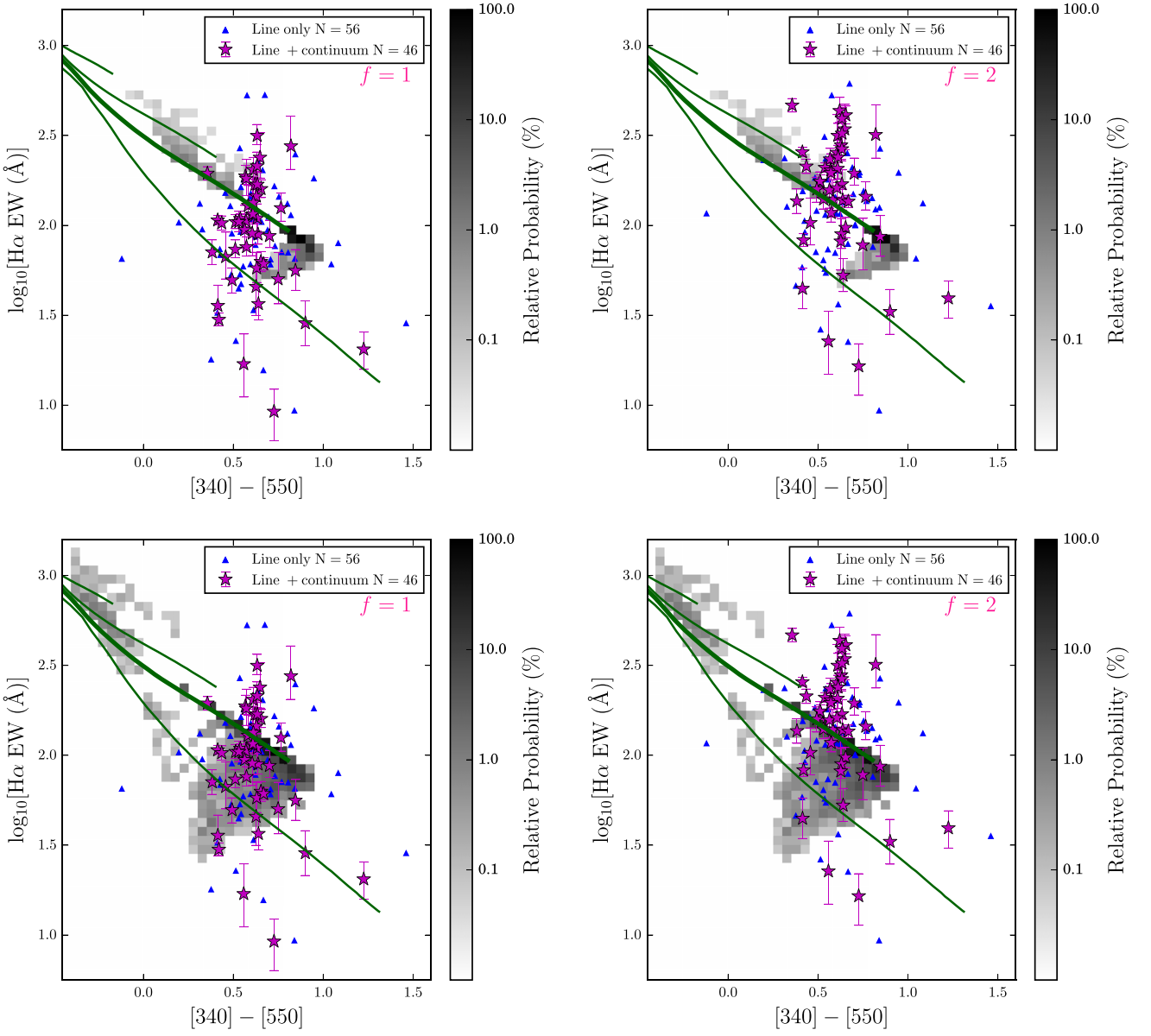

Figure 17. *Top left:* relative probability distribution of galaxies occupying the Hα EW versus [340]–[550] colour space. The density distribution is made from 10 000 galaxies chosen randomly from the simulated galaxies with t_b = 100–300 Myr and f_m = 0.1–0.3. A darker shade suggests a higher probability. The ZFIRE-SP sample is overlaid on the figure with dust corrections applied with an $f = 1$ following prescriptions outlined in Section 4.1. The PEGASE tracks shown are similar to Fig. 15. *Top right:* similar to left but the ZFIRE-SP sample has been dust corrected using an $f = 2$. *Bottom:* similar to the top panels, but galaxies are selected from a simulation with smaller t_b (10–20 Myr) but similar f_m (0.1–0.3) values constrained to a redshift window between $2.0 < z < 2.5$.

velocities (Ramírez-Agudelo et al. 2013). Furthermore, populations of Be stars (Secchi 1866; Rivinius, Carciofi & Martayan 2013), which are near-critically rotating main-sequence B stars observed in local SPs (Lin et al. 2015; Yu et al. 2016; Bastian et al. 2017), have shown evidence for the existence of rapidly rotating stars in massive stellar clusters (e.g. Bastian et al. 2017).

We show the evolution of galaxy properties in the Hα EW versus [340]–[550] colour in Fig. 19 (top panels). Due to limitations in rotational stellar tracks, the metallicity of the stars are kept at $Z = 0.014$, but the stellar atmospheres are kept at $Z = 0.020$. Invoking stellar rotation increases the Hα EW by $\sim$0.1 dex for similar IMFs and shows slightly bluer colours for a given time t. Further analysis of the subcomponents shows us that these changes in the Hα EW versus [340]–[550] colours are driven by the increase in Hα flux and bluer optical colours.

Implementing stellar rotation results in similar effects of a shallower IMF ($\Gamma > -1.35$), but the deviations are not sufficient to explain the $f = 2$ dust-corrected ZFIRE-SP sample within a universal IMF scenario. However, with $f = 1$ dust corrections, only $\sim$5 per cent of the sample lie above the $\Gamma = -1.35$ track with stellar rotation models with a majority of galaxies showing steeper IMFs.

Having a large fraction of stars with extreme rotation will lead to a higher number of ionizing photons and bluer colours and could potentially explain the high-EW objects in our sample. Sustaining such high rotation requires extremely low metallicities, which we further discuss in Section 6.3. Even though we expect the actual

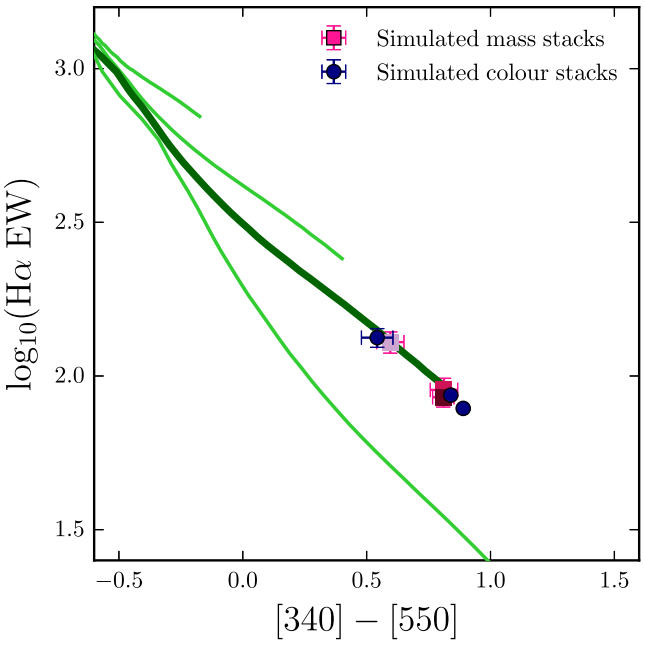

Figure 18. Hα EW versus [340]–[550] colour distribution for stacked PEGASE galaxies used in the simulations in Fig. 17. Here, we stack 100 randomly chosen galaxies from our simulated sample. The selected galaxies have constant SFHs with starbursts with varying strengths (f_m = 0.1–0.3) and durations (τ_b = 100–300 Myr) overlaid at random times (0–3250 Myr) and have been stacked in stellar mass and [340]–[550] colour bins. The errors of the stacks are computed using a bootstrap re-sampling. The PEGASE tracks shown have constant SFHs but varying IMFs. From top to bottom the Γ values of tracks are, respectively, −0.5, −1.0, −1.35 and −2.0. All tracks terminate at t = 3250 Myr.

variation of the Hα EW and [340]–[550] colours due to stellar rotation at near-solar metallicity to be much smaller than what is shown in Fig. 19, we cannot rule out extreme stellar rotation dominant in at low metallicities ($Z \sim 0.002$). Therefore, extreme stellar rotation may provide one explanation independent of the IMF to describe the distribution of our galaxies in the Hα EW and [340]–[550] colour space (see Section 6.3). Furthermore, stellar rotation can introduce fundamental degeneracies to IMF determination which we discuss further in Section 8.4.

6.2 Binary system evolution

We consider the effect of implementing the evolution of binary stellar systems on our study. All SSP models described thus far only considered single SPs, i.e. there were no interactions between stars in an SP. However, recent observational studies in our Galaxy have shown that ∼50 per cent of massive O stars are in binary systems and that the environment may have a strong influence on the dynamical and/or stellar evolution (Langer 2012; Sana et al. 2012, 2013). Only a minority of O stars would have undisturbed evolution leading to supernovae (Leitherer et al. 2014), thus introducing additional complexities to SSP models and strong implications for studies using these models to infer observed stellar properties. Furthermore, Steidel et al. (2016) demonstrated the necessity of invoking models with massive star binaries to fit rest-frame UV and optical features of star-forming galaxies at $z \sim 2.5$.

We use models generated by the Binary Population and Spectral Synthesis v2.0 code (BPASSV2, Stanway, Eldridge & Becker 2016) to investigate the effects of invoking stellar binary evolution in the Hα EW versus [340]–[550] colour space. The computed models have been released by the BPASS team only for a limited set of IMF models. We use IMF models with $Z = 0.02$ and $\Gamma = -1.00$, -1.35 and -1.70 with a lower and upper mass cut-off at 0.5 and 100 $M_\odot$, respectively. The IMF slope for stellar masses between 0.1 and 0.5 $M_\odot$ is kept at $\Gamma = -0.30$ for all the models. We remind the reader that stars with $M_* \lesssim 0.5\ M_\odot$ have negligible effect on the Hα EW versus optical colour parameter space. Fig. 19 (bottom panels) compares the effect of considering stellar binary system evolution in this parameter space. Binary rotation with simple prescriptions for stellar rotation slightly increases the Hα EW (max increase for a given time is ∼0.2 dex) and make galaxies look bluer for a given IMF at a time t. These changes are more prominent for galaxies with steeper IMFs and are driven by the Hα flux and optical colours of the galaxies. Furthermore, unlike effects of rotation, we see a trend on which the steeper IMFs show larger changes (up to ∼×2) in Hα flux and [340]–[550] colours compared to shallower IMFs.

Due to higher ionizing flux and longer lifetimes of massive O type stars in binary systems, galaxies look bluer at an older age compared to what is predicted by single-star models (Eldridge & Stanway 2016). The increase in ionizing flux is driven by transfer of mass between stars causing rejuvenation, generation of massive stars via stellar mergers and stripping of hydrogen envelope to form more hot helium or W-R stars. Mass transfer and mergers between stars also result in larger, bluer stars at later times contributing to the SP to be bluer. The change of Hα EW and [340]–[550] colours due to binary system evolution at $Z = 0.02$ is not sufficient to explain the distribution of the ZFIRE-SP sample galaxies and is significantly smaller than the contribution from stellar rotation.

Note that BPASS single stellar evolutionary models do not consider any form of stellar rotation. BPASS binary models do consider stellar rotation, but only if a secondary star accretes material from a companion. In such scenarios at $Z > 0.004$ the secondary star is spun up, fully mixed, and is rejuvenated resulting it to be a zero-age main-sequence star. However, it is assumed that the star is spun down quickly and stellar rotation is not considered for the rest of its evolution (Eldridge & Stanway 2012; Stanway et al. 2016, J.J. Eldridge., private communication). Since the current version of BPASS binary models does not consider aspects of stellar rotation in the context of reduction in surface gravity and the driving of extra-mixing beyond the expectations from the standard mixing-length theory as discussed in Section 6.1 (also see papers in the series by Meynet & Maeder 2000; Potter, Tout & Eldridge 2012), comparisons between S99 Geneva models and BPASS cannot be performed to constrain the net effect of introducing stellar binary stellar evolution to SSP models that consider stellar rotation.

6.3 Stellar metallicity

Hoversten & Glazebrook (2008) showed that the evolution of galaxies Hα EW versus optical colour parameter space was largely independent of the metallicity of the galaxies. However, these predictions were made using PEGASE models and did not account for the increase in mass-loss via stellar winds and increase in ionizing flux predicted in low-metallicity scenarios by models that consider stellar rotation and binary interactions.

The lack of elements such as Fe that dominate the opacity in radiation-driven stellar winds, stellar interiors, and atmospheres in low metallicity stars results in the generation of higher amount of ionizing photons (e.g. Pauldrach, Puls & Kudritzki 1986; Vink & de Koter 2005; Steidel et al. 2016). Furthermore, at lower metallicities due to weaker stellar winds the mass-loss rate is low, thus most

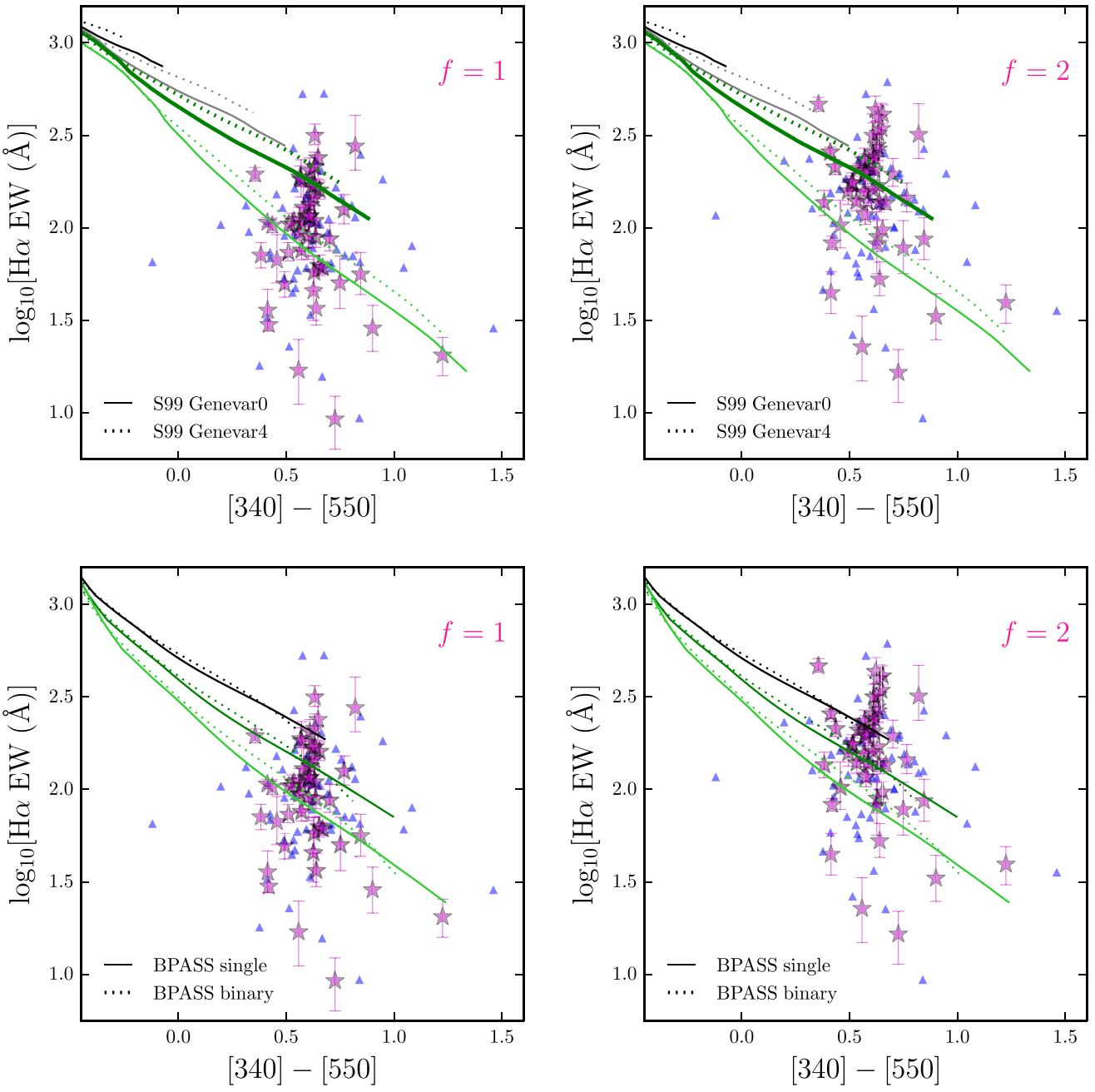

Figure 19. Effects of stellar rotation and evolution of binary stars in the Hα EW versus [340]–[550] colour space. *Top left:* here, we show the evolution of Hα EW versus [340]–[550] colours for S99 SSP models using Geneva tracks with (dotted lines) and without (solid lines) stellar rotation. Each pair of tracks represents the same IMF and the IMFs plot are similar to Fig. 5. The overlaid ZFIRE-SP sample galaxies have been dust corrected using an $f = 1$. *Top right:* similar to the top left panel but the ZFIRE-SP sample has been dust corrected using $f = 2$. *Bottom left:* Here, we show the effects of including the evolution of binary stellar systems in the Hα EW versus [340]–[550] colour space. The models shown here are from the Stanway et al. (2016) BPASSV2 models with $Z = 0.02$. Each pair of tracks with same colour represent the same IMF with (dotted) and without (solid) considering the evolution of binary stellar systems. From top to bottom, the IMFs plot are for Γ values of −1.0, −1.35 (Salpeter IMF), and −1.70. The overlaid ZFIRE-SP sample galaxies have been dust corrected using an $f = 1$. *Bottom right:* similar to the bottom left panel but the ZFIRE-SP sample has been dust corrected using $f = 2$.

massive stars retain their luminosity and continue to shine for an extended time.

When stellar rotation is introduced to single SP models, due to the higher fraction of W-R stars in higher metallicity environments, rotational stellar models with higher metallicity show a larger increment (ΔEW) in ionization flux compared to the increment seen in lower metallicity models (Leitherer et al. 2014). However, at $t < 3100$ Myr, low-metallicity rotating stellar models on average show higher amount of ionizing flux compared to higher metallicities.

When binary interactions are considered, the mass transfer between the binaries results in the increase of angular momentum of the stars causing an increase in stellar rotation (de Mink et al. 2013). Additionally, at $Z \leq 0.004$ if stars with $M_* > 20\,M_\odot$ has accreted

>5 per cent of its original mass, BPASS assumes that the star maintains its rapid rotation throughout its main-sequence lifetime (Stanway et al. 2016). This is driven by weaker stellar winds that allow the stars to maintain their rapid rotations for a prolonged period. Furthermore, rotationally induced mixing of stellar layers causes hydrogen burning to be efficient resulting in rejuvenation of the main-sequence stars. As we show in Section 6.1, stellar rotation increases the production of ionizing photons and therefore lower metallicity systems with binary interactions show higher Hα EW values. Lower cooling efficiencies prominent in lower metallicity environments, also result in the stars to be bluer and brighter. Comparisons between S99 Geneva models with $Z = 0.002$ and $Z = 0.014$ suggest metallicity to have a prominent effect in increasing the Hα EWs compared to stellar rotation. BPASS models also show metallicity effects to be prominent compared to effects by stellar rotation and binary interactions.

Therefore, we conclude that within the scope of current stellar models, metallicity to be the prominent driver in increasing the Hα EWs with stellar rotation and binary interactions contributing to a lesser degree.

In Fig. 20 (top panels), we show the evolution of a $\Gamma = -1.35$ IMF constant SFH stellar tracks from BPASS with varying metallicities. The variation in metallicity in the Hα EW versus [340]–[550] colour is degenerated with IMF variation. Models with lower metallicities favour higher EWs and bluer colours compared to their higher metallicity counterparts.

Next, we explore whether gas-phase metallicities computed for our galaxies (Kacprzak et al. 2015, 2016) suggest sufficiently low stellar metallicities to produce ionizing flux to explain the high-EW galaxies within a $\Gamma = -1.35$ IMF scenario. Converting gas-phase oxygen abundance to stellar iron abundance in high-z galaxies is non-trivial. First, there are considerable systematic uncertainties in the gas-phase metallicities measured using [N II]/Hα ratios. There are uncalibrated interrelations between ionization parameter, electron density and radiation field hardness at $z \sim 2$ (Kewley et al. 2013). For example, at a fixed metallicity, the [N II]/Hα ratio can be enhanced by a lower ionization parameter or the presence of shocks (e.g. Yuan et al. 2012; Morales-Luis et al. 2013) and it is unknown whether the N/O ratio evolves with redshifts (Steidel et al. 2014). From Kacprzak et al. (2015), the gas-phase oxygen abundance of our sample at $\log10(M_*/\mathrm{M}_\odot) = 9.5$ is $\sim$0.5 $\mathrm{Z}_\odot$, however, the systematic uncertainty can be a factor of $\times 2$ because of the unknown calibrations. Because of this, we emphasize that metallicity can be compared reliably in a relative sense, but not yet on an absolute scale (Kewley & Ellison 2008).

Secondly, there is limited knowledge on how iron abundances relative to α-element (e.g. O, Mg, Si, S and Ca) abundance change over cosmic time and in different galactic environments (e.g. Wolfe, Gawiser & Prochaska 2005; Kobayashi et al. 2006; Yuan, Kobayashi & Kewley 2015). In addition, there is a lack of consistency in abundance scale used in stellar atmosphere modelling, stellar evolutionary tracks and nebular models (Nicholls et al. 2016). There are considerable variations in the [O/Fe] ratios that are not well-calibrated at the low-metallicity end. For example, at [Fe/H] < -1.0 the extrapolated [O/Fe] ratio based on Milky Way data is 0.5 (Nicholls et al. 2016), with a $\sim$0.3 dex uncertainty in conversions of individual values (e.g. Stoll et al. 2013). Steidel et al. (2016) argued an average [O/Fe] ratio of 0.74 for $z \sim 2$ UV-selected galaxies at oxygen nebular metallicity of $\sim$0.5 $\mathrm{Z}_\odot$, suggesting a substantially lower stellar metallicity of [Fe/H] ~ -1.0. If we adopt the [O/Fe] ratio of Steidel et al. (2016), then we would reach the same conclusion as Steidel et al. (2016) that our stellar abundance is [Fe/H] ~ -1.0. In this case, we cannot completely rule out extremely low metallicity scenarios to explain the distribution of galaxies in the Hα EW versus [340]–[550] colour space. With $f = 2$ ($f = 1$) dust corrections between BPASS binary models with stellar metallicity of $Z = 0.02$ to $Z = 0.002$, the amount of objects that lie 2σ above the reference $\Gamma = -1.35$ track changes from $\sim$40 per cent ($\sim$9 per cent) to $\sim$9 per cent ($\sim$2 per cent).

Given all the uncertainties mentioned above, we think it is premature to convert our gas-phase oxygen abundance to stellar iron abundance and draw meaningful conclusions. We further note that there are significant uncertainties in massive star evolution in SSP codes and the treatment of stellar rotation and binary stars, which we discuss further in Section 8.4.

6.4 High-mass cut-off

Hoversten & Glazebrook (2008) showed that the high-mass cut-off is degenerated with IMF in the Hα EW versus [340]–[550] colour. In Fig. 20 (bottom panels), we show various IMF slopes with constant SFHs computed for varying values of high-mass cut-off. The deviation between tracks with 80 and 120 $\mathrm{M}_\odot$ high-mass cut-off varies as a function of IMF slope. Shallower IMFs will have a larger effect when the high-mass cut-off is increased due to the high number of stars that will populate the high-mass regions.

The maximum deviation for the high-mass cut-offs for the $\Gamma = -1.35$ tracks is 0.17 dex, which cannot describe the scatter we notice in Hα EWs of our sample. Furthermore, at $z \sim 2$ we expect the molecular clouds forming the stars to be of low metallicity (Kacprzak et al. 2016), which favours the formation of high-mass stars. Therefore, we require the high-mass cut-off to increase, but we are limited by the maximum individual stellar mass allowed by PEGASE. BPASSV2 does allow stars up to 300 $\mathrm{M}_\odot$, however, we do not employ such high-mass limits due to our poor understanding of evolution of massive stars. We conclude that it is extremely unlikely that the high-mass cut-off to have a strong influence on the distribution of the ZFIRE-SP sample galaxies in the Hα EW versus [340]–[550] colour parameter space.

7 DEPENDENCES WITH OTHER OBSERVABLES

In this section, we investigate if the Hα EW versus [340]–[550] colour distribution shows any relationship with environment, stellar mass, SFR and metallicity of the galaxies.

ZFIRE surveyed the Spitler et al. (2012) and Yuan et al. (2014) structure to great detail to probe the effects on environment on galaxy evolution. To date, there are 51 spectroscopically confirmed cluster candidates with ZFOURGE counterparts out of which 38 galaxies are included in our IMF analysis. The other 13 galaxies are removed from our analysis due to the following reasons: eight galaxies are flagged as AGN, two galaxies do not meet the Q_z quality cut for our study, two galaxies give negative spectroscopic flux values and one object due to extreme sky line interference. We perform a 2-sample K-S test on the Hα EW values and [340]–[550] colours for the continuum-detected cluster and field galaxies in our ZFIRE-SP sample and find the cluster and field samples to have similar parent properties. Therefore, we conclude that there are no strong environmental effects on the distribution of galaxies in our parameter space.

For the 22 continuum-detected galaxies in common between Kacprzak et al. (2015) sample and ZFIRE-SP sample, we find no

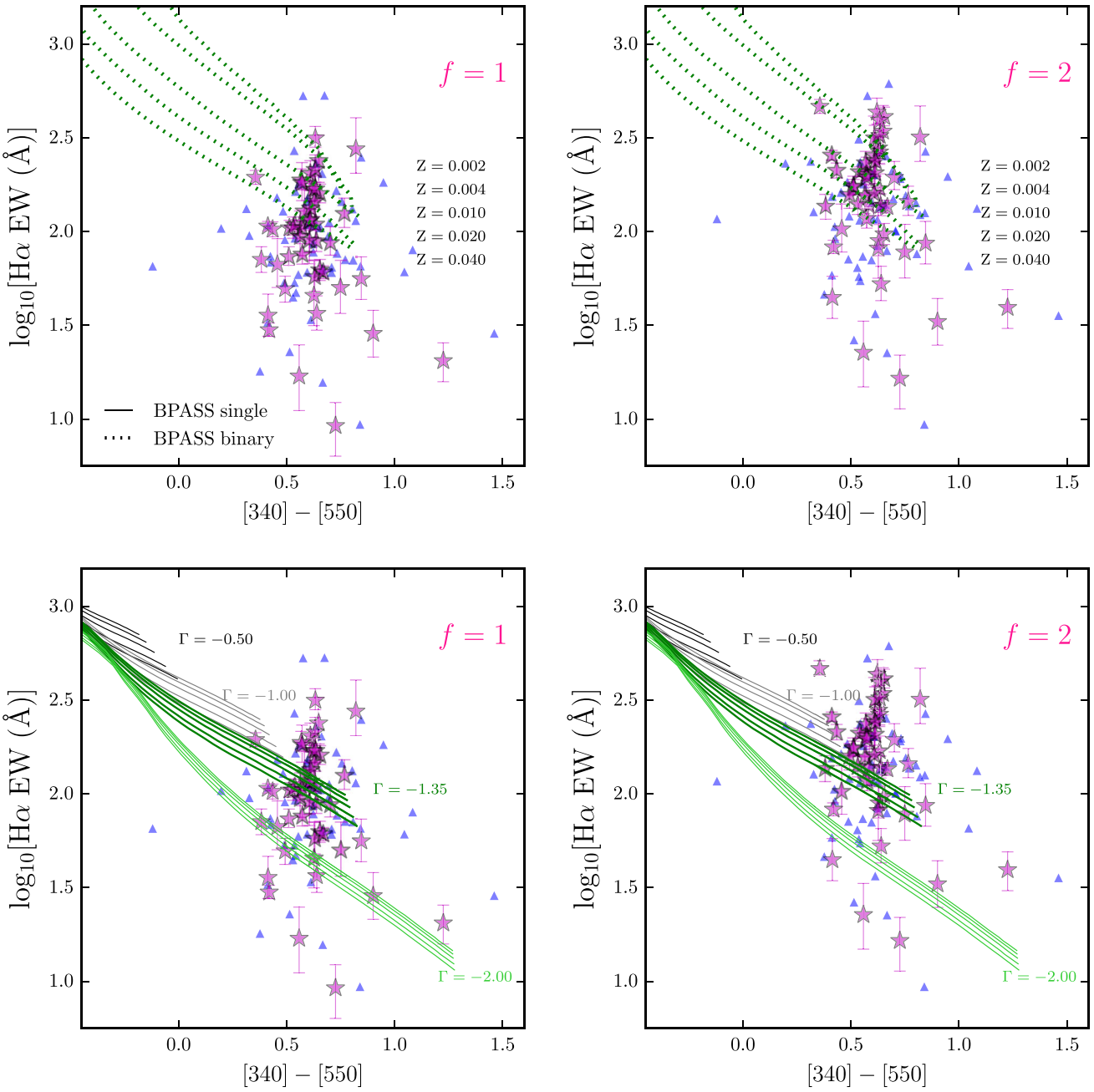

Figure 20. Effects of metallicity and high-mass cut-off of SSP models in the Hα EW versus [340]–[550] colour space. *Top left:* here, we show the evolution of Hα EW versus [340]–[550] colours of the BPASS binary model galaxies with different metallicities. All BPASS models shown here have an IMF of slope $\Gamma = -1.35$ and a constant SFH. From top to bottom, the metallicities of the tracks are $Z = 0.002$, 0.004, 0.010, 0.020 and 0.040. The overlaid ZFIRE-SP sample has been dust corrected with an $f = 1$. *Top right:* similar to the top left panel but with the ZFIRE-SP sample dust corrected with an $f = 2$. *Bottom left:* here, we show the evolution of Hα EW versus [340]–[550] colours of PEGASE constant SFH models with varying IMFs and high-mass cut-offs. From top to bottom, each set of tracks with similar colour has IMF slopes $\Gamma = -0.50$, -1.00, -1.35 and -2.00. Each set of IMFs is computed with varying high-mass cut-offs. From top to bottom for each IMFs, the high-mass cut-offs are, respectively, 120, 110, 100, 90 and 80 $M_\odot$. The overlaid ZFIRE-SP sample has been dust corrected with an $f = 1$. *Bottom right:* similar to the bottom left panel but with the ZFIRE-SP sample dust corrected with an $f = 2$.

statistically significant differences between high- and low-metallicity samples for Hα EWs and [340]–[550] colours.

We further use the Salpeter IMF tracks with constant SFHs to compare the EW excess of our continuum-detected sample with stellar mass and SFR. In Fig. 21 (left-hand panels), we show the EW excess as a function of stellar mass. We divide the sample into low-mass ($\log_{10}(M_*/M_\odot) < 10.0$) and high-mass ($\log_{10}(M_*/M_\odot) \geq 10.0$) bins and compute the scatter in EW excess to find that there is a greater tendency for low-mass galaxies to show larger scatter in EW offsets.

Looking for IMF change as a function of SFR is inherently problematic, especially if SFR is itself computed from the Hα flux assuming a universal IMF. Nevertheless, in order to compare with Gunawardhana et al. (2011), we show this in Fig. 21 (centre panels) and

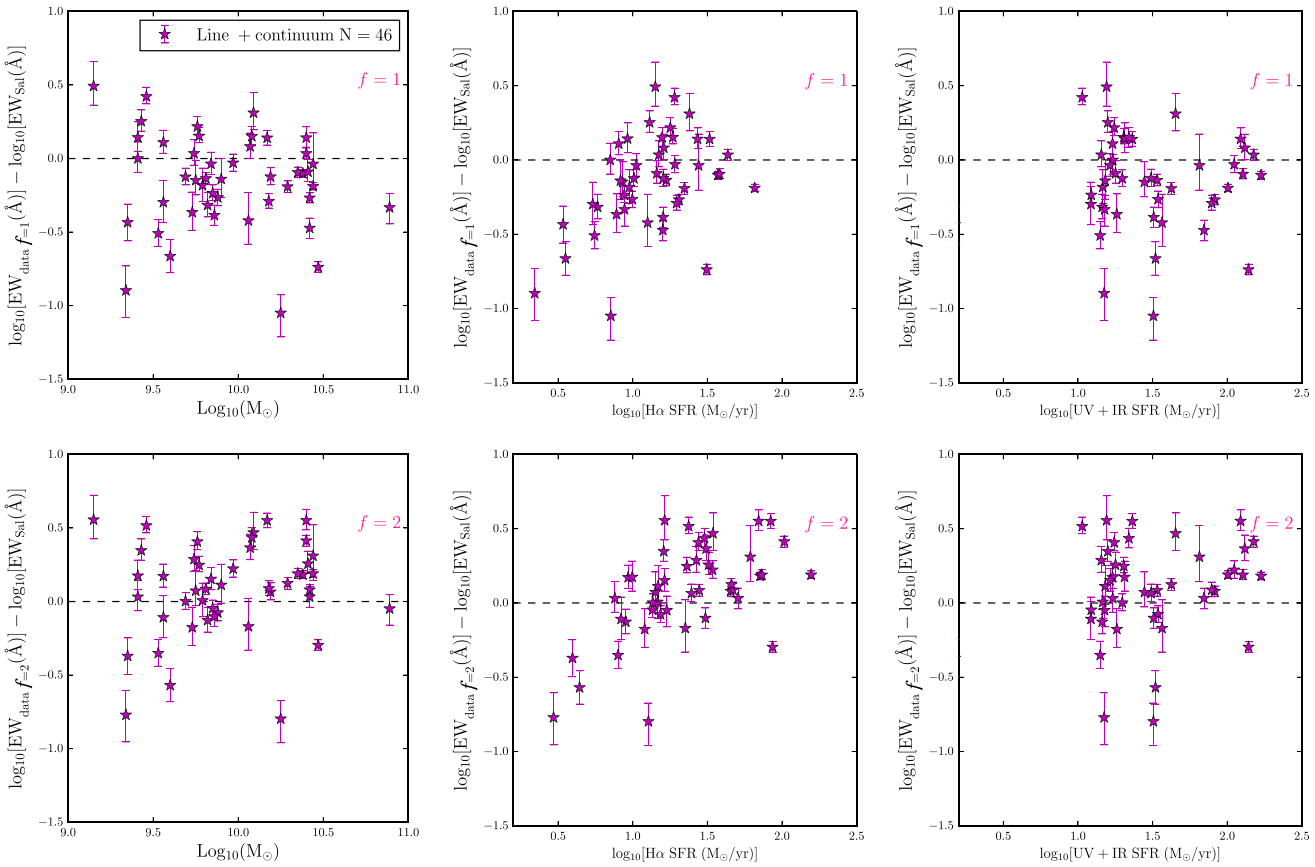

Figure 21. EW excess of the dust-corrected (top panels → $f = 1$, bottom panels → $f = 2$) continuum-detected sample from a PEGASE track of $\Gamma = -1.35$ IMF slope with constant SFH. *Left-hand panels:* EW excess as a function of stellar mass. *Centre panels:* EW excess as a function of Hα SFR. Hα SFR has been calculated using $f = 1$ and $f = 2$ in the top and bottom panels, respectively. *Right-hand panels:* EW excess as a function of UV+IR SFR. In all panels the dashed line denotes $y = 0$.

confirm the trend they found of EW offset for higher 'SFR' objects. However, we refrain from interpreting this as a systematic trend for IMF variation. By using best-fitting SEDs from ZFOURGE, we compute the UV+IR SFRs (Tomczak et al. 2014) and find that there is a greater tendency for low UV+IR SFR galaxies to show larger EW offsets, which is shown by Fig. 21 (right-hand panels). We note that the difference in SFR between Hα and UV+IR is driven by the different time-scales of SFRs probed by the two methods.

8 DISCUSSION

8.1 Comparison with local studies

Our study follows a method first outlined by Kennicutt (1983) and later implemented on large data sets by Hoversten & Glazebrook (2008) and Gunawardhana et al. (2011) to study the IMF of star-forming galaxies. We find that the distribution of Hα EWs and optical colours at $z \sim 2$ to be unlikely to be driven by a sample of galaxies with a universal Salpeter like IMF. Hoversten & Glazebrook (2008) found a trend with galaxy luminosity with low-luminosity galaxies in SDSS favouring a steeper IMF and the highest luminosity ones showing a Salpeter slope. Gunawardhana et al. (2011) found a systematic variance in IMF as a function of SFR in GAMA galaxies with the highest SFR galaxies lying above the Salpeter track. However, we note that the use of $(g - r)_{0.1}$ colour by the $z \sim 0$ studies may have given rise to additional complexities in the analysis, by introducing significant emission line contamination, and the use of SFR as a variable in IMF change is problematic as its calculation depends on IMF and Hα luminosity.

Comparing our results with the local galaxies of HG08 shows distinctive differences in the Hα EW versus $(g - r)_{0.1}$ colour distribution. Since galaxies at $z \sim 2$ had only ∼3.1 Gyr to evolve, we observe younger, bluer SPs giving rise to tighter $(g - r)_{0.1}$ colours (distributed around 0.082 mag with a standard deviation of 0.085 mag). However, HG08 galaxy sample comprises of much redder colours with a larger scatter in $(g - r)_{0.1}$. In a smooth star formation scenario, we interpret the large scatter of the HG08 sample in $(g - r)_{0.1}$ colour space to be driven by the large variety of ages of the galaxies.

Galaxies at $z \sim 2$ show a large range in Hα EWs compared to $z \sim 0$ results. In our analysis, we investigated several key factors that may contribute to the large scatter of Hα EW at $z \sim 2$. Compared to $z \sim 0$ galaxy populations, we expect galaxies at $z \sim 2$ to be young, actively star-forming in various environments and physical conditions that may be distinctively different from local conditions. Therefore, effects such as starbursts may be prominent and dust properties may have significant variation, which can influence the observed Hα EW values. Galaxy mergers and multiple starburst phases in the evolutionary history of $z \sim 0$ galaxies add additional layers of complexity. Furthermore, the presence of old SPs requires

Hα absorption to be corrected, which we expect to be negligible at $z \sim 2$. Due to the limited evolutionary time-scale at $z \sim 2$ (only 3 Gyr), we consider most of these effects to have no significant influence on our analysis. However, we cannot completely rule out effects of dust sight-lines to our analysis, which we discuss further in Section 8.3.

The development of much advanced stellar tracks and greater understanding of stellar properties allow us to explore uncertainties related to stellar modelling that may significantly influence the observed parameters of galaxies at $z \sim 2$.

8.2 What do we really find?

In Section 4.5, we showed that our ZFIRE-selected sample was not preferentially biased towards extremely high star-forming galaxies and that our mass-complete $z \sim 2$ sample is sensitive to quite low EWs. Since observational bias appears not to be the explanation we can investigate physical factors that drive the difference in the ZFIRE-SP sample from universal Salpeter like IMF scenarios in the Hα EW versus [340]–[550] colour plane.

Hα flux is a direct probe of the SFR on time-scales of ∼10 Myr, the continuum level at 6563 Å provides an estimate of the mass of the old SPs and therefore Hα EW is a proxy for the sSFR. Similarly, for monotonic SFHs, the optical colours change smoothly with time, so the [340]–[550] colour is a second proxy for the sSFR, but with different IMF sensitivity. Of the two sSFR measures, the Hα EW is the most sensitive to the highest mass stars, so one way to express our result is to state that there is an excess of ionizing photons (i.e. Hα) at a given SFR compared to a Salpeter-slope model.

In our sample, with $f = 2$ dust corrections, ∼50 per cent galaxies have an excess of high-mass stars for a given sSFR compared to the expectation by a Salpeter like IMF. By stacking galaxies in mass and [340]–[550] colour bins, we can average out stochastic variations in SFHs between galaxies. Our stacking results further confirmed that on average, for all masses and sSFR values, a universal IMF cannot produce the observed galaxy distribution in the Hα EW versus [340]–[550] colour space. We performed further analysis to understand other mechanisms that may drive this excess in Hα EW for a given sSFR.

8.3 Dust and starbursts

The dust extinction values in our analysis were derived using FAST, which uses underlying assumptions of IMF and SFH to produce best-fitting stellar parameters to galaxy observables. Our own analysis of dust showed SED-derived extinction values from the UV slope to have a strong dependence on the assumed IMF. However, for the purposes of testing consistency with a universal Salpeter-slope this suffices.

We further found that differential extinction in dust between the stellar continuum and nebular emission line regions can introduce significant scatter to galaxies in the Hα EW versus [340]–[550] colour parameter space. By analysing Balmer decrement values for a subset of galaxies in our sample, we found that there was significant scatter in the relation between extinction of nebular and stellar continuum regions (f), which can be attributed to differences in dust sight-lines between galaxies. Reddy et al. (2015) showed that this scatter in extinction to be a function of Hα SFR, where galaxies with higher star-forming activity shows larger nebular extinction compared to galaxies with low SFRs. We test this by allowing f values to vary as a free parameter in the Hα EW versus [340]–[550] colour space for each galaxy to force agreement with a universal Salpeter-like IMF. Our results showed extreme values for the distribution of f, including unphysical negative values, suggesting that it is extremely unlikely that the scatter in the Hα EW versus [340]–[550] colour space is driven solely by the variation of f values.

Starbursts in galaxies can introduce significant scatter in the Hα EW versus [340]–[550] colour space. We implemented a stacking procedure for the galaxies in mass and colour bins to remove stochastic SFHs of individual galaxies, treating them as an ensemble SP with a smooth SFH prior to $z \sim 2$. We found that our stacks on average (50 per cent of the $f = 1$ stacks and 100 per cent of the $f = 2$ stacks) favour shallower IMF slopes compared to the traditional $\Gamma = -1.35$ values from Salpeter. By performing Monte Carlo simulations of starbursts using PEGASE SSP models, we found that time-scales of bursts make it extremely unlikely for them to account for the galaxies which lie significantly above the $\Gamma = -1.35$ track.

8.4 Dependencies on SSP models and stellar libraries

We compared the evolution of model galaxies in the Hα EW versus [340]–[550] colours using PEGASE and S99 SSP codes to conclude that the evolution of these parameters are largely independent of the SSP models used for a given stellar library. The [340]–[550] colours were designed in order to avoid strong emission lines regions in the rest-frame optical spectra which averts the need of complicated photoionization codes to generate nebular emission lines. Hα flux is generated using a constant value to convert Lyman continuum photons to Hα photons, which is similar between PEGASE and S99.

We found that stellar libraries play a vital role in determining the evolutionary tracks of galaxies in the Hα EW versus [340]–[550] colour parameter space. Stellar libraries with rotation show higher amounts of ionizing flux which results in higher Hα EW values for a given [340]–[550] colour. Leitherer et al. (2014) further showed that rotation leads to larger convective cores in stars increasing the total bolometric luminosity, which can mimic a shallower IMF. At [340]–[550] = 0.61, introducing stellar rotation via Geneva stellar tracks with $Z = 0.014$ results in $\Delta\log_{10}[\mathrm{EW}(\log_{10}(\text{Å}))] \sim 0.09$. Therefore, we found that rotation cannot itself account for the scatter of our sample in Hα EW versus [340]–[550] colour parameter space at near solar metallicities.

Consideration of binary stellar systems is imperative to understand the stellar properties of $z \sim 2$ galaxies (Steidel et al. 2016). However, added complexity arises due to angular momentum transfer during binary star interactions. This may influence the rotation of the galaxies and therefore it is necessary to consider the evolution of binary stars with detailed prescriptions of stellar rotation. Metallicity of the stars become important in such scenarios, which is a strong factor that regulates the evolution of stellar rotation. However, adding additional degrees of freedom for SSP models makes it harder to constrain their values, thus resulting in extra uncertainties (Leitherer et al. 2014). At [340]–[550] = 0.61, introducing the effect of binaries via BPASS models resulted in $\Delta\log_{10}[\mathrm{EW}(\log_{10}(\text{Å}))] \sim 0.01$. Comparing results between S99 (single SP stellar tracks with and without rotation) and BPASS (single and binary stellar tracks with rotation), we found stellar rotation to have a larger contribution to the ΔEW compared to binaries. Direct comparisons require further work to investigate differences in the evolution of stellar systems between S99 and BPASS SSP codes.

We found low stellar metallicities ($Z \sim 0.002$) to have a strong influence in increasing the Hα EWs for a given [340]–[550] colour. At [340]–[550] = 0.61, reducing the metallicity of BPASS binary models from $Z = 0.02$ to $Z = 0.002$ resulted in

Table 2. Summary of scenarios investigated to explain the distribution of the ZFIRE-SP sample in the Hα EW versus [340]–[550] colour parameter space within a universal IMF framework.

Effect	Section	SSP model	SFH	Z	Median(ΔEW)		per cent[a]		Conclusion
					$f=1$	$f=2$	$f=1$	$f=2$	
Dust	4	PEGASE	Exp declining ($\tau = 1000$ Myr)	0.020	−0.13	0.10	20 per cent	46 per cent	Unlikely[b]
Observational bias	4.5	PEGASE	Exp declining ($\tau = 1000$ Myr)	0.020	–	–	–	–	Excluded
Starbursts	5	PEGASE	Constant	0.020	–	–	–	–	Excluded
Stellar rotation	6.1	S99	Constant	0.020	−0.39	−0.15	2 per cent	13 per cent	Probable[c]
Binaries	6.2	BPASS	Constant	0.020	−0.18	0.05	9 per cent	39 per cent	Future work[c]
Metallicity									
Single stellar systems	6.3	BPASS	Constant	0.020	−0.17	0.06	13 per cent	37 per cent	Unlikely
(with rotation)			Constant	0.010	−0.29	−0.06	4 per cent	22 per cent	Probable
			Constant	0.002	−0.47	−0.23	0 per cent	6 per cent	Probable
Binary stellar systems	6.3	BPASS	Constant	0.020	−0.18	0.05	9 per cent	39 per cent	Unlikely[c]
(with rotation)			Constant	0.010	−0.32	−0.08	4 per cent	22 per cent	Probable[c]
			Constant	0.002	−0.51	−0.28	2 per cent	9 per cent	Probable[c]
High-mass cut-off									
80 $M_\odot$	6.4	PEGASE	Constant	0.020	−0.01	0.22	28 per cent	54 per cent	Excluded[c]
120 $M_\odot$	6.4	PEGASE	Constant	0.020	−0.14	0.09	17 per cent	39 per cent	Excluded[c]

Notes. [a]The fraction of ZFIRE-SP sample galaxies with $>2\sigma$ positive deviations from the $\Gamma = -1.35$ tracks.
[b]Even though we cannot exclude effects from various dust sight-lines, we demonstrated that effects from dust cannot explain the excess of high Hα EW galaxies.
[c]Conclusions driven within the bounds of current SSP models, however, more sophisticated models are required on stellar rotation, binary evolution and high-mass evolution to fully constrain the effects.

$\Delta\log_{10}[\mathrm{EW}(\log_{10}(\text{Å}))] \sim 0.36$. This was largely driven by the increase in the number of ionization photons in the SPs due to lower opacities, lower mass-loss via stellar winds and sustained stellar rotation. Interactions between stars also contribute to an increase in ionizing flux. When considering the ionization energy generated by an SP, effects of stellar rotation is degenerated with the abundance of high-mass stars (see fig. 16 of Szécsi et al. 2015). Therefore, we cannot completely rule out effects of stars with extremely low metallicities to describe the distribution of our galaxies in the Hα EW versus [340]–[550] colour parameter space. In Section 6.3, we provided a thorough analysis of the gas-phase metallicities derived for the ZFIRE sample by Kacprzak et al. (2015) and Kacprzak et al. (2016) and inferred the metal abundances of stellar systems, which is a primary regulator of ionizing photons. However, uncertainties in deriving gas-phase abundances of elements via nebular emission line ratios (driven by our limited understanding of the ionization parameter at low metallicities), uncertainties in computing relative abundances of α elements in stellar systems and our limited understanding on linking gas-phase metallicities to stellar metallicities in $z \sim 2$ SPs constrains our ability to distinguish between effects of metallicity and IMF.

8.5 Case for the IMF

So far we have investigated various scenarios (summarized in Table 2) that could explain the distribution of the ZFIRE-SP sample galaxies in the Hα EW versus [340]–[550] colour without invoking changes in the IMF. However, none of the scenarios by itself could best describe the distribution of our galaxies.

The galaxies in our sample have stellar masses between $\log_{10}(M_*/\mathrm{M}_\odot) = 9$ and 10 and we expect these galaxies to grow in stellar mass during cosmic time to be galaxies with stellar masses of $\sim\log_{10}(M_*/\mathrm{M}_\odot) = 10$–11 at $z \sim 0$ (De Lucia & Blaizot 2007; van Dokkum et al. 2013; Genel et al. 2014; Papovich et al. 2015). Recent studies of ETGs with physically motivated models have shown the possibility for a two phase star formation (e.g. Ferreras et al. 2015). Furthermore, recent semi-analytic models have shown that a varying IMF best reproduces observed galaxy chemical abundances of ETGs (e.g. Lacey et al. 2016; Fontanot et al. 2017, and references therein). According to these models, ETGs during their starburst phases at high redshift are expected to produce higher fraction of high-mass stars (shallower IMFs). Gunawardhana et al. (2011) showed that $z \sim 0$ star-forming galaxies also show an IMF dependence, where highly star-forming galaxies prefer shallower IMFs.

If we consider a varying IMF hypothesis, our results are consistent with a scenario where star-forming galaxies form stars with a high fraction of high-mass stars compared to their local ETG counterparts. With lower metallicities and higher SFRs prominent at $z \sim 2$, we expect the fragmentation of molecular clouds to favour the formation of larger stars due to lower cooling efficiencies and higher heating efficiencies due to radiation from the young massive stars (Larson 2005). Krumholz et al. (2010) showed that radiation trapping prominent in high star-forming regions of dense gas surface density can also favour the formation of massive stars. If we allow the IMF to vary in our analysis, the distribution of the ZFIRE-SP sample in Hα EW versus [340]–[550] colour space can be explained, however values as shallow as $\Gamma = 0.5$ could be required. This could be problematic for chemical evolution models and have implications to how galaxies form and evolve (Romano et al. 2005). We note that invoking extremely shallow IMFs can have a significant influence on the inferred evolution of the universe. Therefore, it is imperative to fully understand these observations and test alternate explanations.

8.6 Effect of IMF variation on fundamental quantities

If the IMF does vary, we need to consider the potential effect on the basic parameters in our input ZFIRE survey, which were calculated using a Chabrier IMF (Chabrier 2003). First, we consider possible effects on the calculation of our rest frame [340]–[550] colours. This should not have a significant effect for several reasons: first

we are only using the spectral models as an interpolator, and by design we are interpolating only across a small redshift range. At $z = 2.1$, the interpolated and observed colours agree well as discussed in Appendix B. Secondly, we note that the main effect is an increased scatter in the EW axis (Fig. 8), once dust corrected the colours are quite tight. Finally, we note that at these young ages everything is quite blue, hence quite flat spectra are being interpolated.

Next is the effect on SFR and stellar mass, which have been used in many of the previous ZFIRE papers (Yuan et al. 2014; Kacprzak et al. 2015, 2016; Tran et al. 2015; Alcorn et al. 2016; Kewley et al. 2016; Nanayakkara et al. 2016), and here in our own mass selection. To quantify the change in mass, we run PEGASE for Γ and constant SFH models and estimate the change in R-band mass-to-light ratio ($\simeq K$ band at $z \simeq 2$) for ages 1–3 Gyr. We find for $-1.35 < \Gamma < -0.5$ the change in mass-to-light is <0.7 dex, with shallower IMFs resulting in a lower stellar mass. Thus, we conclude that our stellar mass selection is only slightly effected by the possible IMF variations we have identified.

The effect is much more severe for Hα-derived SFRs (Tran et al. 2015, 2017) as these directly count the number of the most massive stars, a sensitivity we have exploited in this paper to measure IMF. For $-1.35 < \Gamma < -0.5$, the change is $\sim$1.3 dex. UV- and far-IR-derived SFRs are more complicated. The rest-frame UV is more sensitive to intermediate-mass stars, at 1500 Å the change in flux is $\sim$0.4 dex for $-1.35 < \Gamma < -0.5$. The far-IR is from younger stars in deeper dust-enshrouded regions, at least in local galaxies (Kennicutt 1998b). It is common at high redshift to use an indicator, which combines UV and far-IR (e.g. Tomczak et al. 2014). These are often calibrated using SP models with idealized SFHs, and traditional IMFs and for a fixed dust mass the balance between UV and IR luminosities will depend on dust geometry, IMF and SFH (Kennicutt 1998b; Calzetti 2013). Therefore, IMF change could lead to difficulties in predicting the true underlying SFR of SPs.

9 SUMMARY AND FUTURE WORK

We have used data from the ZFIRE survey along with the multi-wavelength photometric data from ZFOURGE to study properties of a sample of star-forming galaxies in cluster and filed environments at $z \sim 2$. By using the Hα EW and rest-frame optical colours of the galaxies, we performed a thorough analysis to understand what physical properties could drive the distribution of galaxies in this parameter space. We have improved on earlier analysis by deriving synthetic rest-frame filters that remove emission line contamination. We analysed effects from dust, starbursts, metallicity, stellar rotation and binary stars in order to investigate whether the distribution of the ZFIRE-SP sample galaxies could be explained within a universal IMF framework.

We found the following:

(i) ZFIRE-SP sample galaxies show a large range of Hα EW values, with $\sim$1/3rd of the sample showing extremely high values compared to expectation from a $\Gamma = -1.35$ Salpeter-like IMF. Compared to the HG08 SDSS sample, galaxies at $z \sim 2$ show bluer colours with a larger scatter in Hα EW values.

(ii) The difference in extinction between nebular and stellar emission line regions (f) in galaxies can have a strong influence in determining the distribution of galaxies in the Hα EW versus [340]–[550] colour space. Our Balmer decrement studies for a subsample of galaxies showed a large scatter in f values. However, we showed that considering f value as a free parameter cannot describe the distribution of galaxies in the Hα EW versus [340]–[550] colour space.

(iii) Starbursts can increase the Hα EW to extreme values providing an alternative explanation to IMF for a subset of our galaxies with high Hα EW values. By implementing a stacking technique to remove stochastic SFHs of individual galaxies we concluded that on average our ZFIRE-SP sample still shows higher Hα EW values for a given [340]–[550] colour compared to a $\Gamma = -1.35$ Salpeter like IMF. We further used Monte Carlo simulations to study time-scales of starbursts to conclude that it was extremely unlikely that starbursts could explain the Hα EW versus [340]–[550] colour distribution of a large fraction of our galaxies.

(iv) Stellar rotation, binaries and the high-mass cut-off of SSP models could influence the distribution of galaxies in the Hα EW versus [340]–[550] colour parameter space. However, the individual effects of these were not sufficient to explain the distribution of the observed galaxies.

(v) Considering multiple effects together can describe the galaxies in our parameter space. We showed that the fraction of galaxies above the $\Gamma = -1.35$ tracks reduces to $\sim$5 per cent when considering stellar tracks with high initial rotations ($v_{\rm ini} = 0.4 v_{\rm crit}$) and equal dust extinction between nebular and stellar regions.

(vi) Including single or binary stars with stellar rotation in extreme low-metallicity scenarios can significantly increase the Hα EWs and is also one explanation to describe the distribution of our galaxies in Hα EW versus [340]–[550] colour parameter space. However, gas-phase metallicity analysis of the ZFIRE sample by Kacprzak et al. (2015) and Kacprzak et al. (2016) rules our such low metallicities for our sample. We note that calibration of emission line ratios and differences between stellar and ionized gas metallicities at $z \sim 2$ are uncertainties that may impact our inference about the stellar metallicity of our sample.

(vii) A non-universal high-mass IMF, varying between galaxies, could explain the distribution of galaxies in this parameter space. The Hα excess shows a broad trend with larger offsets for the less massive $z \sim 2$ galaxies. We also confirm the same systematic trend in IMF slope with Chabrier-derived SFR as shown by Gunawardhana et al. (2011) but we refrain from interpreting this.

(viii) Within the scope of our study, for $-1.35 < \Gamma < -0.5$ the variation in high-mass IMF slope can lead to changes in mass-to-light ratios of up to $\sim$0.7 dex. Furthermore, ignoring calibration offsets we compute that Hα SFRs can show deviations up to $\sim$1.3 dex.

IMF change is an important topic as the IMF determines basic parameters such as stellar mass and SFR, which are used to derive broad conclusions about galaxy evolution. What we observe is a population of galaxies with high Hα equivalent widths, i.e. an excess of ionizing photons for a given colour, and we have ruled out intermittent starbursts and alternate SP models as an explanation. Such high-EW objects appear to become more common at high redshift, for example similar observations have been reported at $z \sim 4$ by multiple studies (e.g. Malhotra & Rhoads 2002; Finkelstein et al. 2011; McLinden et al. 2011; Hashimoto et al. 2013; Stark et al. 2013) and have even been invoked at $z > 5$ as an explanation for cosmological re-ionization (Labbé et al. 2013; Labbé 2015; Schenker 2015; Stark et al. 2017). It seems reasonable to hypothesize that the abundance of high-EW objects is evolving towards high redshift and we are seeing this at $z \sim 2$.

Is IMF change responsible? This currently seems to be the only explanation that cannot be ruled out, but we do not yet understand what would drive it to vary between individual galaxies. Further study is required in order to fully comprehend the SP parameters of

the $z \sim 2$ galaxies to determine whether IMF is the main driver for the distribution of galaxies in the Hα EW versus rest-frame optical colour parameter space.

Future work should consider a more thorough statistical analysis using all the broad-band colour information and multiple spectral diagnostics including simultaneously modelling of possible effects of dust, starbursts, metallicity, stellar rotation, binary star evolution and high-mass cut-off of stellar systems together with systematic variances of the IMF. A new generation of stellar models is allowing many of these parameters to be varied and tested. The launch of the *James Webb Space Telescope* in 2018 will provide the opportunity to probe rest-frame optical stellar and NIR populations via high-signal/noise absorption lines and will revolutionize our understanding of the processes of star formation in the $z \sim 2$ universe.

ACKNOWLEDGEMENTS

The data presented herein were obtained at the W.M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W.M. Keck Foundation. The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain and without the generous hospitality of the indigenous Hawaiian community this study would have not been possible. TN, KG and GGK acknowledge Swinburne-Caltech collaborative Keck time, without which this study would have not been possible. We thank the anonymous referee for the constructive comments on our analysis. We thank Madusha Gunawardhana, Elisabete da Cunha, Richard McDermid, Luca Cortese, John Eldridge and Selma de Mink for insightful discussions. We thank Colin Jacobs for writing pypegase (https://github.com/coljac/pypegase), a PYTHON wrapper around PEGASE.2, which was instrumental to perform the SSP simulations. We thank Erik Hoversten for providing us with the SDSS data used in our analysis. We thank Naveen Reddy for providing us with the MOSDEF data used in his Reddy et al. (2015) analysis. We thank the Lorentz Centre and the scientific organizers of the 'The Universal Problem of the Non-Universal IMF' workshop held at the Loretnz Centre in 2016 December, which promoted useful discussions among the wider community on a timely concept. KG acknowledges the support of the Australian Research Council through Discovery Proposal awards DP1094370, DP130101460 and DP130101667. GGK acknowledges the support of the Australian Research Council through the award of a Future Fellowship (FT140100933).

Facilities: Keck:I (MOSFIRE).

REFERENCES

Alcorn L. Y. et al., 2016, ApJ, 825, L2
Astropy Collaboration et al., 2013, A&A, 558, A33
Baldry I. K., Glazebrook K., 2003, ApJ, 593, 258
Bastian N., Covey K. R., Meyer M. R., 2010, ARA&A, 48, 339
Bastian N. et al., 2017, MNRAS, 465, 4795
Beckwith S. V. W. et al., 2006, AJ, 132, 1729
Bessell M. S., 1990, PASP, 102, 1181
Boselli A., Boissier S., Cortese L., Buat V., Hughes T. M., Gavazzi G., 2009, ApJ, 706, 1527
Bouwens R. J. et al., 2016, ApJ, 833, 72
Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503
Brocklehurst M., 1971, MNRAS, 153, 471
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Bruzzese S. M., Meurer G. R., Lagos C. D. P., Elson E. C., Werk J. K., Blakeslee J. P., Ford H., 2015, MNRAS, 447, 618
Calzetti D., 2013, in Falcón-Barroso J., Knapen J. H., eds, Secular Evolution of Galaxies. Cambridge Univ. Press, Cambridge, p. 419
Calzetti D., Kinney A. L., Storchi-Bergmann T., 1994, ApJ, 429, 582
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Capak P. L. et al., 2015, Nature, 522, 455
Cappellari M. et al., 2012, Nature, 484, 485
Cappellari M. et al., 2013, MNRAS, 432, 1862
Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245
Chabrier G., 2003, PASP, 115, 763
Chattopadhyay T., De T., Warlu B., Chattopadhyay A. K., 2015, ApJ, 808, 24
Coil A. L. et al., 2015, ApJ, 801, 35
Conroy C., 2013, ARA&A, 51, 393
Conroy C., Dutton A. A., Graves G. J., Mendel J. T., van Dokkum P. G., 2013, ApJ, 776, L26
Cowley M. J. et al., 2016, MNRAS, 457, 629
Cullen F., McLure R. J., Khochfar S., Dunlop J. S., Dalla Vecchia C., 2017, preprint (arXiv:1701.07869)
de Mink S. E., Langer N., Izzard R. G., Sana H., de Koter A., 2013, ApJ, 764, 166
de Barros S., Reddy N., Shivaei I., 2016, ApJ, 820, 96
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
Driver S. P. et al., 2009, A&G, 50, e12
Ekström S. et al., 2012, A&A, 537, A146
Eldridge J. J., Stanway E. R., 2012, MNRAS, 419, 479
Eldridge J. J., Stanway E. R., 2016, MNRAS, 462, 3302
Erb D. K., Steidel C. C., Shapley A. E., Pettini M., Reddy N. A., Adelberger K. L., 2006, ApJ, 647, 128
Ferreras I., La Barbera F., de la Rosa I. G., Vazdekis A., de Carvalho R. R., Falcón-Barroso J., Ricciardelli E., 2013, MNRAS, 429, L15
Ferreras I., Weidner C., Vazdekis A., La Barbera F., 2015, MNRAS, 448, L82
Finkelstein S. L. et al., 2011, ApJ, 735, 5
Fioc M., Rocca-Volmerange B., 1997, A&A, 326, 950
Fontanot F., De Lucia G., Hirschmann M., Bruzual G., Charlot S., Zibetti S., 2017, MNRAS, 464, 3812
Fumagalli M., da Silva R. L., Krumholz M. R., 2011, ApJ, 741, L26
Fumagalli M. et al., 2012, ApJ, 757, L22
Genel S. et al., 2014, MNRAS, 445, 175
Georgy C., Ekström S., Meynet G., Massey P., Levesque E. M., Hirschi R., Eggenberger P., Maeder A., 2012, A&A, 542, A29
Georgy C. et al., 2013, A&A, 558, A103
Giacconi R. et al., 2001, ApJ, 551, 624
Gilmore G., 2001, in Tacconi L., Lutz D., eds, Starburst Galaxies: Near and Far. Springer-Verlag, Berlin, p. 34
Gräfener G., Vink J. S., 2015, A&A, 578, L2
Guiderdoni B., Rocca-Volmerange B., 1988, A&AS, 74, 185
Gunawardhana M. L. P. et al., 2011, MNRAS, 415, 1647
Guo Y. et al., 2016, ApJ, 833, 37
Hashimoto T., Ouchi M., Shimasaku K., Ono Y., Nakajima K., Rauch M., Lee J., Okamura S., 2013, ApJ, 765, 70
Hopkins A. M., Beacom J. F., 2006, ApJ, 651, 142
Hoversten E. A., 2007, PhD thesis, Johns Hopkins University, Baltimore, Maryland
Hoversten E. A., Glazebrook K., 2008, ApJ, 675, 163
Kacprzak G. G. et al., 2015, ApJ, 802, L26
Kacprzak G. G. et al., 2016, ApJ, 826, L11
Kashino D. et al., 2013, ApJ, 777, L8
Kauffmann G., 2014, MNRAS, 441, 2717
Kennicutt R. C., Jr, 1983, ApJ, 272, 54
Kennicutt R. C., Jr, 1998a, in Gilmore G., Howell D., eds, ASP Conf. Ser., Vol. 142, The Stellar Initial Mass Function (38th Herstmonceux Conference). Astron. Soc. Pac., San Francisco, p. 1

Kennicutt R. C., Jr, 1998b, ApJ, 498, 541
Kewley L. J., Ellison S. L., 2008, ApJ, 681, 1183
Kewley L. J., Dopita M. A., Leitherer C., Davé R., Yuan T., Allen M., Groves B., Sutherland R., 2013, ApJ, 774, 100
Kewley L. J. et al., 2016, ApJ, 819, 100
Klishin A., Chilingarian I., 2016, ApJ, 824, 17
Kobayashi C., Umeda H., Nomoto K., Tominaga N., Ohkubo T., 2006, ApJ, 653, 1145
Kriek M., van Dokkum P. G., Labbé I., Labbé I., Franx M., Illingworth G. D., Marchesini D., Quadri R. F., 2009, ApJ, 700, 221
Kroupa P., 2001, MNRAS, 322, 231
Krumholz M. R., Cunningham A. J., Klein R. I., McKee C. F., 2010, ApJ, 713, 1120
La Barbera F., Ferreras I., Vazdekis A., de la Rosa I. G., de Carvalho R. R., Trevisan M., Falcón-Barroso J., Ricciardelli E., 2013, MNRAS, 433, 3017
Labbé I., 2015, IAU Gen. Assem., 22, 2258492
Labbé I. et al., 2013, ApJ, 777, L19
Lacey C. G. et al., 2016, MNRAS, 462, 3854
Langer N., 2012, ARA&A, 50, 107
Larson R. B., 1973, MNRAS, 161, 133
Larson R. B., 1998, MNRAS, 301, 569
Larson R. B., 2005, MNRAS, 359, 211
Lee H.-c., Gibson B. K., Flynn C., Kawata D., Beasley M. A., 2004, MNRAS, 353, 113
Leitherer C., 1998, in Gilmore G., Howell D., eds, ASP Conf. Ser., Vol. 142, The Stellar Initial Mass Function (38th Herstmonceux Conference). Astron. Soc. Pac., San Francisco, p. 61
Leitherer C., Ekström S., 2012, in Tuffs R. J., Popescu C. C., eds, Proc. IAU Symp. 284, The Spectral Energy Distribution of Galaxies. Cambridge Univ. Press, Cambridge, p. 2
Leitherer C. et al., 1999, ApJS, 123, 3
Leitherer C., Ekström S., Meynet G., Schaerer D., Agienko K. B., Levesque E. M., 2014, ApJS, 212, 14
Levesque E. M., Leitherer C., Ekstrom S., Meynet G., Schaerer D., 2012, ApJ, 751, 67
Lin M.-K., Krumholz M. R., Kratter K. M., 2011, MNRAS, 416, 580
Lin C.-C., Hou J.-L., Chen L., Shao Z.-Y., Zhong J., Yu P.-C., 2015, Res. Astron. Astrophys., 15, 1325
McDermid R. M. et al., 2014, ApJ, 792, L37
McLean I. S. et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. SPIE, Bellingham, p. 84460J
McLinden E. M. et al., 2011, ApJ, 730, 136
Malhotra S., Rhoads J. E., 2002, ApJ, 565, L71
Mancini M., Schneider R., Graziani L., Valiante R., Dayal P., Maio U., Ciardi B., 2016, MNRAS, 462, 3130
Marmol-Queralto E., McLure R. J., Cullen F., Dunlop J. S., Fontana A., McLeod D. J., 2016, MNRAS, 460, 3587
Martín-Navarro I., Barbera F. L., Vazdekis A., Falcón-Barroso J., Ferreras I., 2015a, MNRAS, 447, 1033
Martín-Navarro I., La Barbera F., Vazdekis A., Ferré-Mateu A., Trujillo I., Beasley M. A., 2015b, MNRAS, 451, 1081
Martín-Navarro I. et al., 2015c, ApJ, 798, L4
Martín-Navarro I. et al., 2015d, ApJ, 806, L31
Martins F., Palacios A., 2013, A&A, 560, A16
Masters D. et al., 2014, ApJ, 785, 153
Meurer G. R., 2011, in Treyer M., Wyder T., Neill J., Seibert M., Lee J., eds, ASP Conf. Ser., Vol. 440, UP2010: Have Observations Revealed a Variable Upper End of the Initial Mass Function? Astron. Soc. Pac., San Francisco, p. 189
Meurer G. R. et al., 2009, ApJ, 695, 765
Meynet G., Maeder A., 2000, A&A, 361, 101
Miller G. E., Scalo J. M., 1979, ApJS, 41, 513
Morales-Luis A. B., Sánchez Almeida J., Pérez Montero E., Muñoz-Tuñon C., Aguerri J. A. L., Vilchez J. M., Terlevich E., Terlevich R., 2013, in Guirado J. C., Lara L. M., Quilis V., Gorgas J., eds, Proceedings of the 10th Scientific Meeting of the Spanish Astronomical Society (SEA'13), Highlights of Spanish Astrophysics VII. Valencia, Spain, p. 294
Murdin P. ed. 2000, Wolf, Charles J E (1827-1918)
Nanayakkara T. et al., 2016, ApJ, 828, 21
Nicholls D. C., Sutherland R. S., Dopita M. A., Kewley L. J., Groves B. A., 2016, MNRAS, preprint (arXiv:1612.03546)
Osterbrock D., 1989, Astrophyiscs of Gaseous Nebulae and Active Galactic Nuclei. University Science Books, Mill Valley, CA
Papovich C. et al., 2015, ApJ, 803, 26
Pauldrach A., Puls J., Kudritzki R. P., 1986, A&A, 164, 86
Pei Y. C., 1992, ApJ, 395, 130
Persson S. E. et al., 2013, PASP, 125, 654
Pettini M., Steidel C. C., Adelberger K. L., Dickinson M., Giavalisco M., 2000, ApJ, 528, 96
Pflamm-Altenburg J., Weidner C., Kroupa P., 2009, MNRAS, 395, 394
Potter A. T., Tout C. A., Eldridge J. J., 2012, MNRAS, 419, 748
Quider A. M., Pettini M., Shapley A. E., Steidel C. C., 2009, MNRAS, 398, 1263
Ramírez-Agudelo O. H. et al., 2013, A&A, 560, A29
Rasappu N., Smit R., Labbé I., Bouwens R. J., Stark D. P., Ellis R. S., Oesch P. A., 2016, MNRAS, 461, 3886
Reddy N. A. et al., 2015, ApJ, 806, 259
Relaño M., Kennicutt R. C., Jr., Eldridge J. J., Lee J. C., Verley S., 2012, MNRAS, 423, 2933
Rigby J. R., Rieke G. H., 2004, ApJ, 606, 237
Rivinius T., Carciofi A. C., Martayan C., 2013, A&AR, 21, 69
Romano D., Chiappini C., Matteucci F., Tosi M., 2005, in Corbelli E., Palla F., Zinnecker H., eds, Astrophysics and Space Science Library, Vol. 327, The Initial Mass Function 50 Years Later. Springer-Verlag, Dordrecht, p. 231
Salasnich B., Bressan A., Chiosi C., 1999, A&A, 342, 131
Salpeter E. E., 1955, ApJ, 121, 161
Sana H. et al., 2012, Science, 337, 444
Sana H. et al., 2013, A&A, 550, A107
Scalo J. M., 1986a, Fund. Cosmic Phys., 11, 1
Scalo J. M., 1986b, in De Loore C. W. H., Willis A. J., Laskarides P., eds, Proc. IAU Symp. 116, Luminous Stars and Associations in Galaxies. Reidel, Dordrecht, p. 451
Schenker M. A., 2015, PhD thesis, California Institute of Technology
Schombert J. M., Bothun G. D., Impey C. D., Mundy L. G., 1990, AJ, 100, 1523
Schwarzschild M., Spitzer L., 1953, The Observatory, 73, 77
Scoville N. et al., 2007, ApJS, 172, 1
Secchi A., 1866, Astron. Nachr., 68, 63
Shim H., Chary R.-R., Dickinson M., Lin L., Spinrad H., Stern D., Yan C.-H., 2011, ApJ, 738, 69
Shivaei I., Reddy N. A., Steidel C. C., Shapley A. E., 2015, ApJ, 804, 149
Skelton R. E. et al., 2014, ApJS, 214, 24
Smit R., Bouwens R. J., Labbé I., Franx M., Wilkins S. M., Oesch P. A., 2016, ApJ, 833, 254
Smith R. J., 2014, MNRAS, 443, L69
Sobral D., Best P. N., Smail I., Mobasher B., Stott J., Nisbet D., 2014, MNRAS, 437, 3516
Speagle J. S., Steinhardt C. L., Capak P. L., Silverman J. D., 2014, ApJS, 214, 15
Spitler L. R. et al., 2012, ApJ, 748, L21
Spitler L. R. et al., 2014, ApJ, 787, L36
Stanway E. R., Eldridge J. J., Becker G. D., 2016, MNRAS, 456, 485
Stark D. P., Schenker M. A., Ellis R., Robertson B., McLure R., Dunlop J., 2013, ApJ, 763, 129
Stark D. P. et al., 2017, MNRAS, 464, 469
Steidel C. C., Shapley A. E., Pettini M., Adelberger K. L., Erb D. K., Reddy N. A., Hunt M. P., 2004, ApJ, 604, 534
Steidel C. C. et al., 2014, ApJ, 795, 165
Steidel C. C. et al., 2016, ApJ, 826, 159
Stoll R., Prieto J. L., Stanek K. Z., Pogge R. W., 2013, ApJ, 773, 12
Straatman C. M. S. et al., 2016, ApJ, 830, 51

Szécsi D., Langer N., Yoon S.-C., Sanyal D., de Mink S., Evans C. J., Dermine T., 2015, A&A, 581, A15
Tadaki K.-i., Kodama T., Tanaka I., Hayashi M., Koyama Y., Shimakawa R., 2013, ApJ, 778, 114
Tomczak A. R. et al., 2014, ApJ, 783, 85
Tran K.-V. H. et al., 2015, ApJ, 811, 28
Tran K.-V. H. et al., 2017, ApJ, 834, 101
van Dokkum P. G., Conroy C., 2010, Nature, 468, 940
van Dokkum P. G., Conroy C., 2012, ApJ, 760, 70
van Dokkum P. G. et al., 2013, ApJ, 771, L35
Vazdekis A., Casuso E., Peletier R. F., Beckman J. E., 1996, ApJS, 106, 307
Vink J. S., de Koter A., 2005, A&A, 442, 587
Weidner C., Ferreras I., Vazdekis A., La Barbera F., 2013, MNRAS, 435, 2274
Weisz D. R. et al., 2012, ApJ, 744, 44
Whitaker K. E. et al., 2011, ApJ, 735, 86
Wolfe A. M., Gawiser E., Prochaska J. X., 2005, ARA&A, 43, 861
Woosley S. E., Heger A., 2006, ApJ, 637, 914
Woosley S. E., Weaver T. A., 1995, ApJS, 101, 181
Yoon S.-C., Langer N., Norman C., 2006, A&A, 460, 199
York D. G. et al., 2000, AJ, 120, 1579
Yu P.-C. et al., 2016, AJ, 151, 121
Yuan T.-T., Kewley L. J., Swinbank A. M., Richard J., 2012, ApJ, 759, 66
Yuan T. et al., 2014, ApJ, 795, L20
Yuan T., Kobayashi C., Kewley L. J., 2015, ApJ, 804, L14
Zanstra H., 1927, ApJ, 65, 50
Zeimann G. R. et al., 2014, ApJ, 790, 113

APPENDIX A: ROBUSTNESS OF CONTINUUM FIT AND Hα FLUX

The study presented in this paper relies on accurate computation of Hα flux and the underlying continuum level. In this section, we compare our computed continuum level and Hα flux values using two independent methods to investigate any systematic biases to our Hα EW values.

We examine the robustness of our continuum fit to the galaxies by using ZFOURGE imaging data to estimate a continuum level from photometry. By using a slit-box of size 0.7 arcsec × 2.8 arcsec overlaid on the 0.7 arcsec point spread function convolved FourStar K_s image, we calculate the photometric flux expected from the galaxy within the finite slit aperture. Justification of this slit-size comes from the spectrophotometric calibration of the ZFIRE data which is explained in detail in Nanayakkara et al. (2016). Since we remove slits that contain multiple galaxies within the selected aperture, only 38 continuum-detected galaxies and 39 galaxies with continuum limits are used in this comparison. We then convert the magnitude to f_λ as follows:

$$f_\lambda = 10^{-0.4(\mathrm{mag}+48.60)} \times \frac{3\times10^{-2}}{\lambda_c^2}\ \mathrm{erg\,s^{-1}\,cm^{-2}\,\text{Å}^{-1}}, \quad \text{(A1a)}$$

where λ_c is the central wavelength of the MOSFIRE K band which we set to 21 757.5 Å. Next, we compute the Hα flux contribution to f_λ by using the photometric bandwidth of FOURSTAR K_s band ($\Delta\lambda_{FS} = 3300$ Å).

$$F_{\mathrm{H}\alpha_{\mathrm{cont}}} = \frac{F_{\mathrm{H}\alpha}}{\Delta\lambda_{\mathrm{FS}}}. \quad \text{(A1b)}$$

We then remove the Hα flux contribution to the photometric flux to compute the inferred continuum level from photometry as shown below:

$$F_{\mathrm{cont}_{\mathrm{photo}}} = f_\lambda - F_{\mathrm{H}\alpha_{\mathrm{cont}}}. \quad \text{(A1c)}$$

Since the Hα flux is the dominant emission line for star-forming non-AGN galaxies, we ignore any contributions from other nebular emission lines to the photometric continuum level.

We compare this photometrically derived continuum level with the spectroscopic continuum level in Fig. A1 (left-hand panel). The median deviation of the detected continuum values are ∼0.026 in logarithmic flux values with a $1\sigma_{\mathrm{NMAD}}$ scatter of 0.12, which leads us to conclude that the photometrically derived continuum values agree well with the spectroscopic continuum detections, thus confirming the robustness of our continuum calculations. We further note that the large scatter of galaxies below the continuum detection level is driven by the increased fraction of sky noise, which is expected and further confirms that we have robustly established the continuum detection limit. There is no strong dependence of the Hα EW on the continuum detection levels, which suggests that the Hα EW values are not purely driven by weak continuum levels. Further analysis of such detection biases are shown in Section 4.5.

In order to test the robustness of the measured Hα flux, we compare the Hα fluxes between this study and the ZFIRE catalogue. Nebular line fluxes in the ZFIRE catalogue are measured by integrating a Gaussian fit to the emission lines (Nanayakkara et al. 2016). We follow a similar technique to calculate Hα fluxes for emission lines unless they show strong velocity structures.

It is vital to ascertain if Gaussian fits to emission lines would give drastically different Hα flux values compared to our visually integrated fluxes. In Fig. A1 (right-hand panel), we show this comparison for ZFIRE-SP sample galaxies that do not have strong sky line residuals. For the continuum detections in the above subset, the median deviation between the manual limits and Gaussian fits is 0.19×10^{-17} erg s^{-1} cm^{-2} Å^{-1} with $\sigma_{\mathrm{NMAD}} = 0.20\times10^{-17}$ erg s^{-1} cm^{-2} Å^{-1}. Therefore, Hα flux values agree with each other within error limits with minimal scatter. Single Gaussian fits would fail to describe Hα emission profiles of galaxies with strong rotations or galaxy that have undergone mergers. These features will require complicated multi-Gaussian fits to accurately provide the underlying Hα flux. All $3\sigma_{\mathrm{NMAD}}$ outliers of continuum detections contain profiles that cannot be described using single Gaussian fits. Therefore, we expect the direct integration to be the most accurate method to calculate the Hα flux for galaxies with velocity structures because it is independent of the shape of the Hα emission. We note that all galaxies with sky line contamination shows profiles that are well described by single Gaussian fits.

Due to above tests, we are confident that neither the Hα flux nor the continuum level calculations would give rise to systematic errors in our analysis. Therefore, we conclude that the Hα EW values derived for our continuum-detected ZFIRE-SP sample are robust.

A1 AGN contamination to Hα flux

As described in Section 2.1, we flag AGN of the ZFIRE sample following Coil et al. (2015) selection criteria. However, it is possible that weak AGN that are not flagged by our selection may still contaminate the ZFIRE-SP sample and contribute to higher Hα emission. In order to investigate effects from such AGN, we use Coil et al. (2015) selection and the measured [N II] fluxes to compute upper limits to Hα fluxes required for the galaxies to be flagged as AGN as follows:

$$f(\mathrm{H}\alpha)_{\mathrm{inf}} < \frac{f([\mathrm{N\,II}])}{0.316}, \quad \text{(A2)}$$

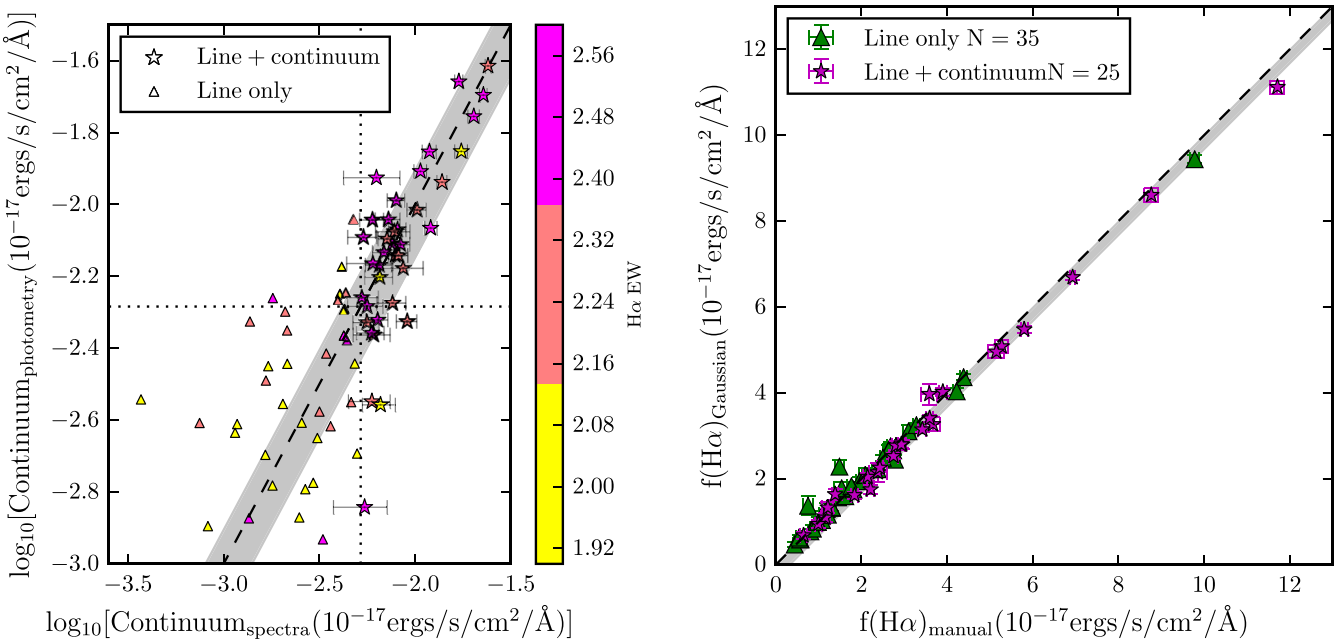

Figure A1. Left: we compare the continuum level derived from spectra with the expected continuum level from photometry. The stars represent objects with robust continuum detections. The remaining sample is shown as triangles. The error bars for the continuum detections come from bootstrap re-sampling. The one to one line is shown as a black dashed line. The median from one to one deviation of the continuum-detected objects is 0.06 logarithmic flux units with $\sigma_{\mathrm{NMAD}} = 0.14$ (shown by the grey shaded region). The horizontal and vertical dotted lines are the continuum detection levels. Right: we compare the Hα flux values computed using Gaussian fits to that of visually identified limits for galaxies which show no strong sky contamination. The magenta stars represent objects with robust continuum detections. The remaining sample is shown as green triangles. The error values are from the integration of the error spectrum. The one to one line is shown as a black dashed line. The median one to one deviation of the continuum-detected objects is 0.19 flux units with $\sigma_{\mathrm{NMAD}} = 0.20$ (shown by the grey shaded region).

where f([N II]) is the measured [N II] flux for our galaxies. We find that our measured Hα fluxes are $\sim\times2$ higher than the inferred Hα fluxes ($f(\mathrm{H}\,\alpha)_{\mathrm{inf}}$) computed using the above equation. Using the ratio of the measured and inferred Hα fluxes, we find the upper limit to the fraction of Hα photons produced by the strongest possible AGN that would not be flagged by the Coil et al. (2015) selection to be $\sim$0.4. Therefore, if our sample is contaminated by weak subdominant AGN, we expect the AGN to be responsible for $\sim$50 per cent of the observed Hα flux.

APPENDIX B: DERIVATION OF BOX-CAR FILTERS FOR IMF AND DUST ANALYSIS

B1 The choice of 340 and 550 filters

Due to the strong dependence of nebular emission line properties in the $(g-r)_{0.1}$ colour regime, we shift our analysis to synthetic box-car optical filters that avoid regions of strong emission lines. Fig. B1 shows the wavelength coverage of our purpose built [340] and [550] box-car filters along with the wavelength coverage of the FourStar filters in the rest frame of a galaxy at $z = 2.1$. It is evident from the figure that $J1_{z=2.1}$, $J3_{z=2.1}$ and $Hl_{z=2.1}$ filters avoid wavelengths with strong emission lines. We choose the median wavelength of $J1_{z=2.1}$ (3400 Å) and $Hl_{z=2.1}$ (5500 Å) filters to develop box-car filters with a wavelength coverage of 4500 Å. These box-car filters are used to compute optical colours for the ZFIRE-IMF sample.

B2 The choice of 150 and 260 filters

To be consistent with our IMF analysis, which employs the [340]–[550] colours, we compute UV filters employing a similar technique described in Appendix B1. Bessell (1990) *B* and *I* filters that samples the optical wavelength regime are chosen for this purpose. For a typical galaxy at $z = 2.1$, these filters sample the UV wavelength regime. Therefore, by dividing the wavelength coverage of the *B* and *I* filters by 3.1, we define a filter set that samples the UV region in the rest frame of a galaxy at $z = 2.1$

We then define two box-car filters that has similar wavelength coverage of $\sim$6700 Å to the blueshifted *B* and *I* filters. The bluer filter is centred at 1500 Å while the redder filter is centred at 2600 Å, both with a width of 673 Å. We name these filters [150] and [260], respectively, and are used in our analysis to investigate the IMF dependence of the dust parameters derived by FAST.

B3 Comparison between observed colours and EAZY-derived rest-frame colours

We use the observed FourStar *J*1 and *H*l fluxes and the best-fitting SED fits of our ZFIRE-SP sample to test the robustness of the observed colours with the EAZY-derived rest-frame colours. In Fig. B2 (left-hand panel), we show the differences between the observed ($J1-Hl$) colours and the rest-frame [340]–[550] colours (as described in Appendix B1) computed from the best-fitting EAZY SED templates. We compare $J1-Hl$ with [340]–[550] colours and expect them to approximately agree by construction at $z = 2.1$.

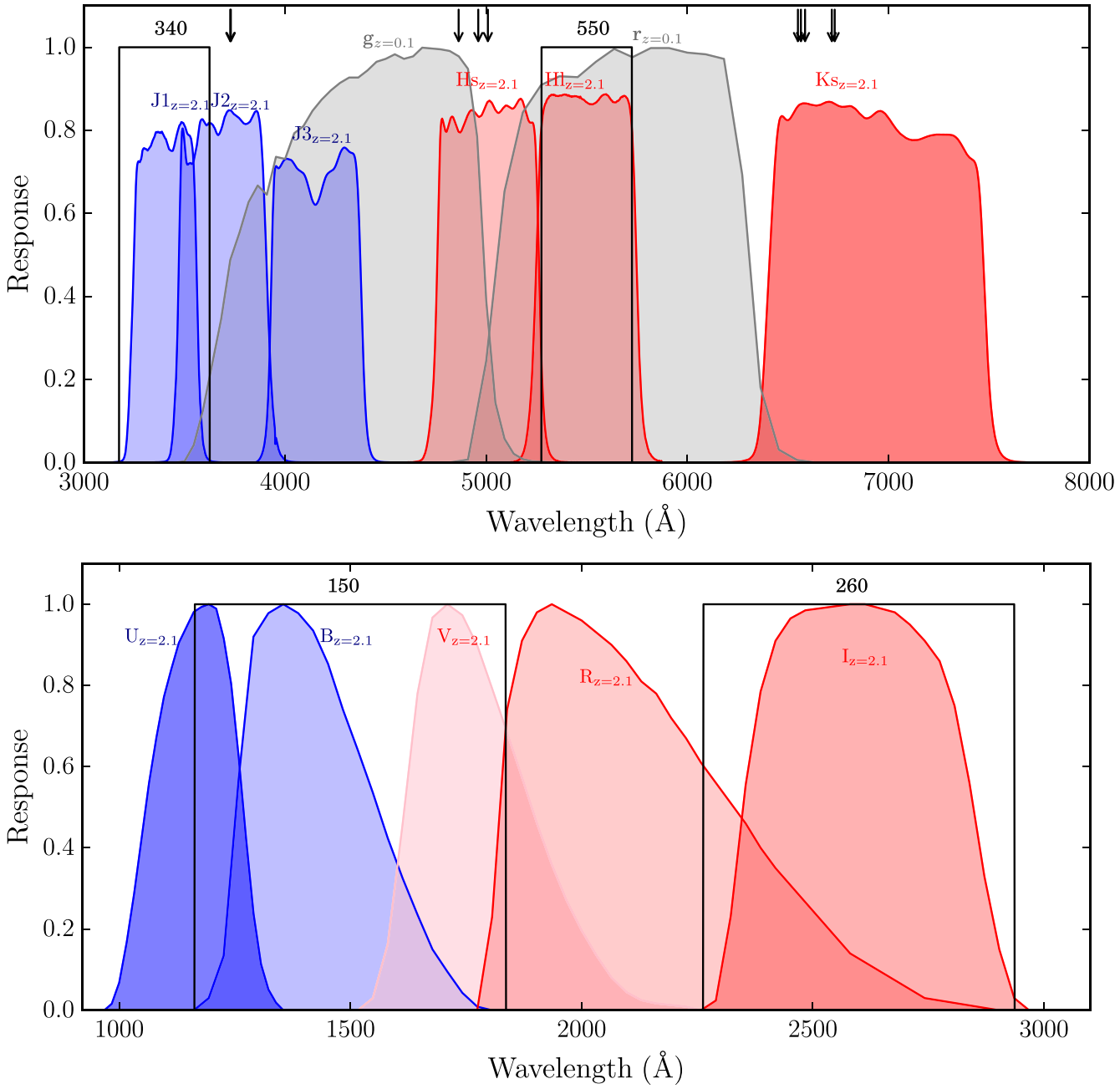

Figure B1. Top: the wavelength coverage of the [340] and [550] filters compared with the wavelength coverage of rest-frame FourStar NIR filters for a galaxy at $z = 2.1$. We also show the wavelength coverage of the $g_{z=0.1}$ and $r_{z=0.1}$ filters used by the HG08 analysis. The arrows denote locations of strong emission lines. Bottom: the wavelength coverage of the [150] and [340] filters along with Bessell (1990) filters de-redshifted from $z = 2.1$.

Using a PEGASE model spectrum, we compare the difference with z of $J1-Hl$ with [340]–[550] colours with what we expect from SED templates. Lines go through zero at $z = 2.1$ as we expect. The model spectrum is extracted at $t = 3100$ Myr from a galaxy with an exponentially declining SFH with a $p_1 = 1000$ Myr, $\gamma = -1.35$ IMF and no metallicity evolution. We use the model spectrum to compute the [340]–[550] colour. We then make a grid of redshifts between $z = 1.8$ and $z = 2.7$ with $\Delta z = 0.01$ and redshift the wavelength of the model spectra for redshifts in this grid by multiplying the wavelength by $(1 + z)$. For each redshift, we compute the $(J1-Hl)$ colours and since we only investigate the colour difference, there is no need to consider the redshift dimming or K-corrections etc.

Since the rest-frame filters assume that the galaxies are at $z = 2.1$, we expect the observed colours and rest-frame colours to agree at this redshift. Fig. B2 (left-hand panel) shows that for the model galaxy this expectation holds with a maximum deviation of $\sim\pm0.1$ mag in colour difference between $z = 1.8$ and $z = 2.7$. Galaxies in the ZFIRE-SP sample show a much larger deviation of $\sim\pm0.5$ mag, which we attribute to errors in photometry as evident from the large error bars. Furthermore, zero-point corrections in

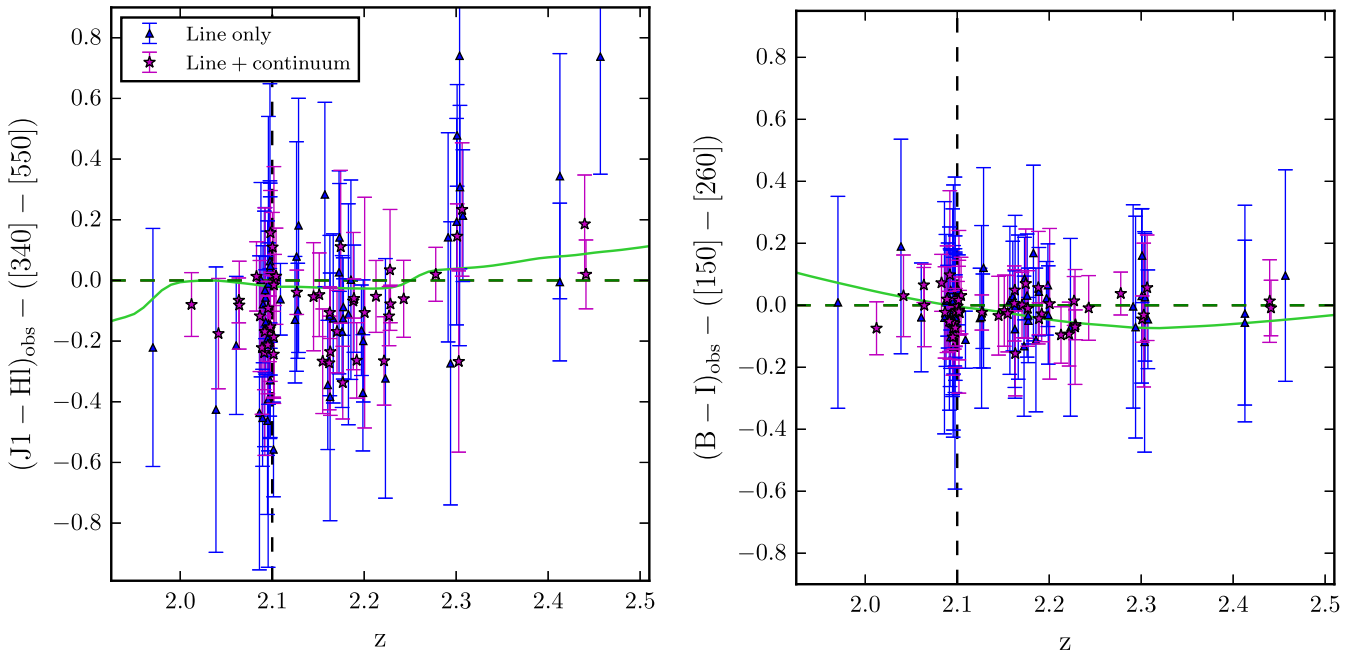

Figure B2. The colour difference between observed colours and rest-frame colours derived from best-fitting SED templates from EAZY. Rest-frame colours are computed in such a way that the wavelength coverage in the observed frame at $z = 2.1$ is approximately similar to the wavelength coverage of the rest-frame filters at $z = 0$. Left: the colour difference between the observed ($J1-Hl$) colours and the [340]–[550] colours of the galaxies in our IMF sample. Galaxies with continuum detections are shown as magenta stars while galaxies with only Hα emission are shown as blue triangles. The error bars are from the errors in $J1$ and Hl filters from the ZFOURGE survey photometry. The green line shows the evolution of the colour difference of a PEGASE model galaxy. The vertical dashed line denotes $z = 2.1$, which is the redshift used to de-redshift the NIR filters in order to compute rest-frame colours. The horizontal dashed line shown is the Δ(colour) $= 0$ line. Galaxies lying on this line shows no difference between the observed colours and the rest-frame colours derived via EAZY. Right: similar to left but for ($B - I$) and [150]–[260] colours.

the SED fitting techniques can give rise to additional systematic variations.

A similar analysis is performed on the ($B - I$) colours using the observed fluxes from the ZFOURGE survey and [150]–[260] colours (as described in Appendix B2) on the best-fitting EAZY SED templates. The same PEGASE model galaxy used for the ($J1-Hl$) comparison is used to derive the ($B - I$) colours by redshifting the spectra to redshifts between $z = 1.8$ and $z = 2.7$. Fig. B2 (right-hand panel) shows the comparison between observed and EAZY-derived colours along with the *ideal* expectation computed from PEGASE spectra. Due to the intrinsic shape of the SED, the redshift evolution of $\Delta(B - I)$ is opposite to that of $\Delta(J1-Hl)$.

APPENDIX C: DOES SSP MODELS GIVE IDENTICAL RESULTS?

In order to investigate whether there is a strong dependence of the Hα EW and/or $(g - r)_{0.1}$ colour evolution of model galaxies on the SSP models used, we compare the galaxy properties from PEGASE with that of S99 (Leitherer et al. 1999). S99 models support the use of multiple stellar libraries. For this analysis, we consider the Padova AGB stellar library that is an updated version of the Guiderdoni & Rocca-Volmerange (1988) stellar tracks that includes cold star parameters and thermally pulsating asymptotic giant branch (AGB) and post-AGB stars.

We compute PEGASE models using a constant SFR of 1×10^{-4} M$_\odot$ Myr^{-1} with various Γ values. PEGASE models are scale free and are generated using a baryonic mass reservoir normalized to 1 M$_\odot$. Having higher SFRs in PEGASE will result in the SFR exceeding the maximal amount of SFR possible for the amount of gas available in the galaxy reservoir before reaching $z \sim 2$. The other parameters of the PEGASE models are kept as mentioned previously.

S99 models employ a different approach to compute synthetic galaxy spectra where the SFR is not normalized and therefore the SFR should be kept at a reasonable level that allows the HR diagram to be populated with a sufficient number of stars during the time-steps the models are executed. For S99, we use an SFR of 1 M$_\odot$ using Padova AGB stellar libraries with a Z of 0.02, and similar IMFs to the PEGASE models. We do not change any other parameters in the S99 models from its default values. We state the S99 parameters below.

(i) Supernova cut-off mass is kept at 8 M$_\odot$.
(ii) Black hole cut-off mass is kept at 120 M$_\odot$.
(iii) Initial time is set to 0.01 Myr and time 1000 time steps are computed with logarithmic spacing up to 3100 Myr.
(iv) We consider the full isochrone for mass interpolation.
(v) We leave the indices of the evolutionary tracks at 0.
(vi) PAULDRACH/HILLIER option is used for the atmosphere for the low-resolution spectra.
(vii) Metallicity of the high-resolution models are kept at 0.02.
(viii) Solar library is used for the UV line spectrum.
(ix) In order to generate spectral features in the NIR, we use microturbulent velocities of 3 km s^{-1} and solar abundance ratios for alpha-element to Fe ratio.

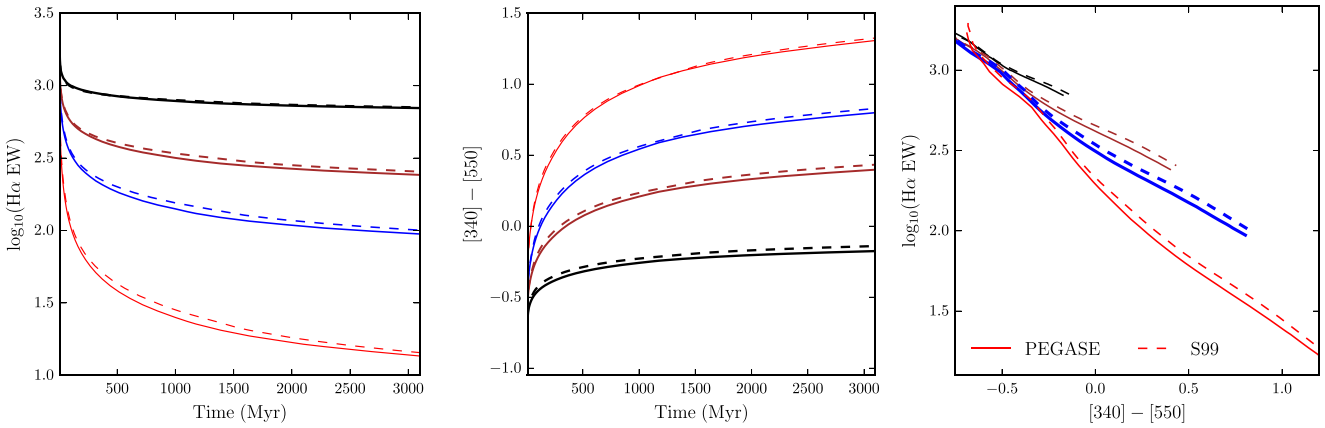

Figure C1. Comparison of model parameters between PEGASE and S99. We compare the evolution of Hα EW and [340]–[550] colours between PEGASE and S99 models. The S99 models are computed with similar to PEGASE using the updated Padova stellar tracks following prescriptions in Section 3. Each SSP model has been computed using varying Γ values of -0.5 (black), -1.0 (brown), -1.35 (blue) and -2.0 (red). The PEGASE models are shown as solid lines while S99 models are shown as dashed lines. Left: the evolution of Hα EW with time for PEGASE and S99 models. Centre: the evolution of PEGASE and S99 [340]–[550] colours with time. Right: the evolution of PEGASE and S99 Hα EW and [340]–[550] colours. The wavelength coverage used for [340]–[550] colours do not include any strong emission lines and therefore the colours are independent of photoionization properties of the galaxies.

To account for the difference in the SFRs between the SSP models, we scale the PEGASE Hα flux and the corresponding continuum level to the 1 $M_\odot$ yr^{-1} value by multiplying by 10^{10}. The scaling process assumes that the Hα luminosity $\propto$ SFR as shown by Kennicutt (1983). By interpolating the S99 models to the PEGASE time grid, we calculate the difference in the parameters between the two models for a given time.

All models are computed using a constant SFR. Since, the number of O and B stars that contribute strongly to the Hα flux is regenerated at a constant speed, the Hα flux reaches a constant value within a very short time-scale and maintains this value. The lifetime of these O and B stars are in the order of $\sim$10 Myr and therefore there is no effect from the accumulation of these stars to the Hα flux. Hα flux between the SSP models shows good agreement with shallower IMFs showing larger Hα flux values. This is driven by the increase in the fraction of larger O and B stars, which contributes to the increase in ionizing photons to boost the Hα flux. The Hα flux generated by the two SSP models agree within $\sim$0.03 dex. The discrepancy is slightly higher for steeper IMFs, perhaps driven by minor differences in the mass distribution of the stars in the SSP models.

The continuum level at 6563 Å also show good agreement between the SSP models and the differences are $\lesssim$0.1 dex. Unlike Hα flux, the continuum levels do not reach a constant value within 3 Gyr. This is driven by the larger lifetime of the A and G stars, which largely contributes to the galaxy continuum. The rate at which the continuum level increases is dependant on the IMF, where galaxies with steeper IMFs take longer times to reach a constant continuum level. However, having a higher fraction of smaller stars eventually leads to a higher continuum level compared to a scenario with a shallower IMF. Since the fraction of A and G stars are higher in a steeper IMF, the higher continuum level is expected. PEGASE and S99 models follow different time-scales for stellar evolution. For a given IMF, PEGASE models evolve the continuum level faster to reach a higher value compared to the S99 models. The discrepancy between the models increases up to $\sim$1500 Myr, after which it decreases to reach a constant value.

The change of the Hα EW between the two SSP models (shown by Fig. C1 left-hand panel) is driven by the differences in the Hα flux and the continuum level. Both the models behave similarly by decreasing Hα EW with time. Shallower IMFs show higher EW values driven by higher Hα flux and lower continuum values and the shape of the ΔEW function is driven by the differences in the continuum evolution.

Furthermore, in Fig. C1 (centre panel) we investigate the evolution of [340]–[550] colours derived between PEGASE and S99. Since the wavelength regime covered by [340] and [550] colours do not include emission lines, a direct [340]–[550] colour comparison between S99 models (with no emission lines) with PEGASE models (with emission lines) is possible. Models with different IMFs show distinctive differences between the derived [340]–[550] colours. Driven by the excess of higher mass blue O and B stars, galaxies with shallower IMFs show bluer colours compared to galaxies with steeper IMFs for a given time. Steeper IMFs show a better agreement between the two SSP models. Both, PEAGSE and S99 use the same stellar tracks from the Padova group and therefore we attribute the differences between the SSP models to differences in methods used by PEAGSE and S99 to produce the composite SPs.

In Fig. C1 (right-hand panel), we compare the evolution of Hα EW with [340]–[550] colours for PEGASE and S99 models. Following on close agreement between the evolution of Hα EW and [340]–[550] colours between the two SSP models, in the Hα EW versus [340]–[550] colour plane galaxies from both PEGASE and S99 show similar evolution. Therefore, our conclusions in this study are not affected by the choice of SSP model (PEGASE or S99) but we note that stellar libraries do play a more prominent role, which we discuss in detail in Section 6.

APPENDIX D: NEBULAR EXTINCTION PROPERTIES OF ZFIRE $z \sim 2$ SAMPLE

D1 Hβ detection properties

Fig. D1 (left-hand panel) shows the *UVJ* diagram (Spitler et al. 2014) of the ZFIRE Hβ targeted and detected sample. Rest-frame *UVJ* analysis shows that our Hβ- detected sample is a reasonably representative subset of our star-forming galaxies.

Fig. D1 (right-hand panel) shows Balmer decrement values as a function of stellar mass. We calculate the median Balmer decrement

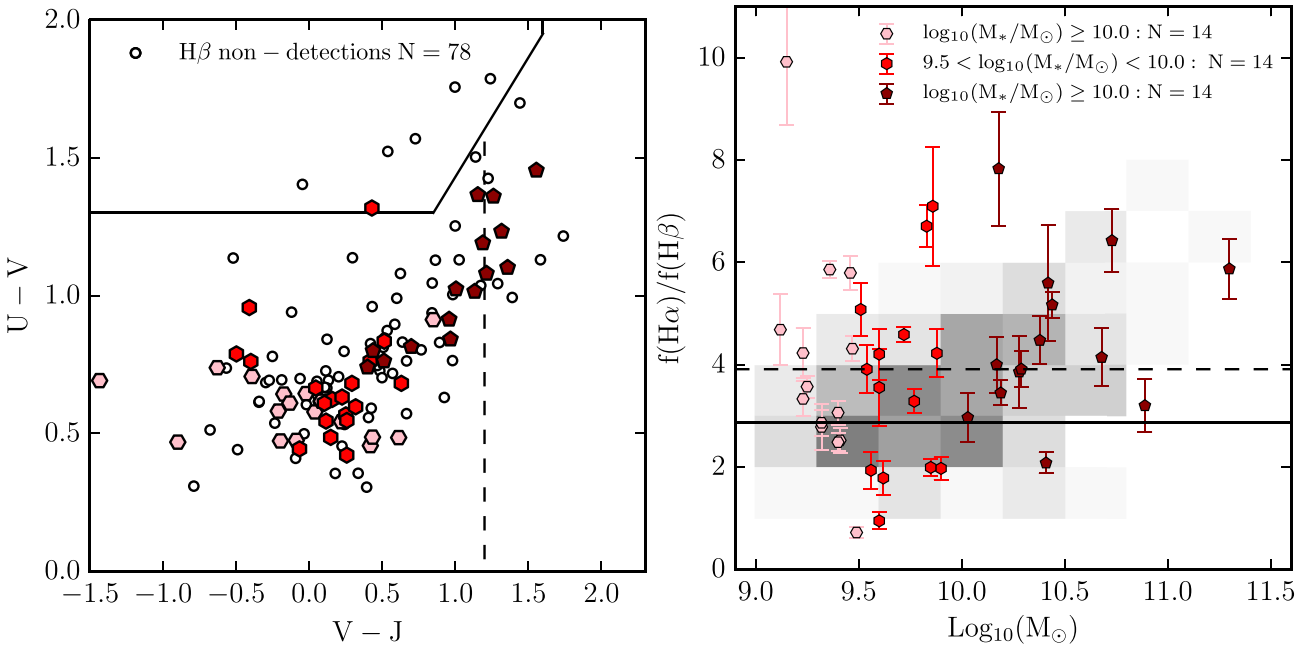

Figure D1. Left: rest-frame *UVJ* diagram of the ZFIRE Hβ targeted sample. Galaxies shown by green filled symbols have been detected in Hβ with SNR >5 while black open circles show Hβ non-detected galaxies. Note that not all galaxies with Hβ detections have been targeted in *K* band. The rest-frame colours have been derived using spectroscopic redshifts where available. Right: Balmer decrement versus mass of the ZFIRE sample. The black dashed horizontal line is the median Balmer decrement value (3.9) of the sample and the black solid line is the Balmer decrement = 2.86 limit from Case B recombination at $T = 10\,000$ K (Osterbrock 1989). All masses are derived from SED fitting by FAST. The 2D density histogram shows the distribution of values from Reddy et al. (2015). In both panels, the Hβ-detected sample is colour coded according to their stellar mass.

for our sample to be 3.9. Using a least-squares polynomial fit to the data, we find that galaxies with higher mass are biased towards high Balmer decrement values. There are nine galaxies with Balmer decrements below 2.86, which are below the theoretical minimum value for case B recombination at $T = 10\,000$ K (Osterbrock 1989).

D2 Derivation of the dust corrections to the ZFIRE-SP sample

In this section, we show how we used the Calzetti et al. (2000) and Cardelli et al. (1989) attenuation laws to derive extinction values for the ZFIRE-SP sample.

We first calculate the starburst reddening curve at 0.6563 μm using the following equation:

$$k'(\lambda) = 2.659\left(-1.857 + \frac{1.040}{\lambda}\right) + R'_v, \quad \text{(D1a)}$$

where λ is in μm. This equation is only valid for wavelengths between 0.63 μm < λ < 2.2 μm. Following Calzetti et al. (2000), the total attenuation (R'_v) is set to 4.05. We use the derived value for the reddening curve to calculate the attenuation of the continuum at 0.6563 μm ($A_c(0.6563)$).

$$A_c(0.6563) = k'(0.6563) \times \frac{A_v}{R'_v} = 0.82A_c(V). \quad \text{(D1b)}$$

Next, we use the Cardelli et al. (1989) prescription to calculate the attenuation of the nebular emission lines. This law is valid for both diffuse and dense regions of the ISM and therefore we expect it to provide a reasonable approximation to the ISM of galaxies at $z \sim 2$. We use the following equations to evaluate the extinction curve at 6563 Å.

$$x = 1/\lambda, \quad \text{(D2a)}$$

where λ is in μm and is between $1.1\,\mu\text{m}^{-1} \leq x \leq 3.3\,\mu\text{m}^{-1}$. Wavelength-dependent values $a(x)$ and $b(x)$ are defined as follows:

$$\begin{aligned} a(x) = {} & 1 + (0.176\,99 \times y) - (0.504\,47 \times (y^2)) \\ & - (0.024\,27 \times (y^3)) + (0.720\,85 \times (y^4)) \\ & + (0.019\,79 \times (y^5)) - (0.775\,30 \times (y^6)) \\ & + (0.329\,99 \times (y^7)) \end{aligned} \quad \text{(D2b)}$$

$$\begin{aligned} b(x) = {} & (1.413\,38 * y) + (2.283\,05 * (y^2)) + (1.072\,33 * (y^3)) \\ & - (5.384\,34 * (y^4)) - (0.622\,51 * (y^5)) \\ & + (5.302\,60 * (y^6)) - (2.090\,02 * (y^7)), \end{aligned} \quad \text{(D2c)}$$

where $y = x - 1.82$. Using $a(x)$ and $b(x)$ values, the attenuation of the nebular emission line at 0.6563 μm ($A_n(0.6563)$) can be expressed as follows:

$$A_n(0.6563) = A_n\left[a(0.6563^{-1}) + \frac{b(0.6563^{-1})}{R''_v}\right] = 0.82A_n(V). \quad \text{(D2d)}$$

R''_v is set to 3.1 following Cardelli et al. (1989). Colour excess is defined as

$$E(B - V) = A(V)/R_v. \quad \text{(D3)}$$

Calzetti et al. (1994) shows that at $z \sim 0$ newly formed hot-ionizing stars reside in dustier regions of a galaxy compared to old SPs. Ionizing stars mainly contribute to the nebular emission lines while the old SPs contribute the stellar continuum of a galaxy. Therefore,

they find that nebular emission lines of a galaxy to be ∼2 times more dust attenuated than the stellar continuum. Here, we denote this correction factor as f. Using n and c subscripts to denote the nebular and continuum parts, respectively,

$$E_n(B-V) = f \times E_c(B-V). \quad \text{(D4)}$$

Substituting equation (D3) into equation (D2d):

$$A_n(0.6563) = 0.82 \times R''_v \times E_n(B-V). \quad \text{(D5a)}$$

Using equation (D4):

$$A_n(0.6563) = 0.82 \times R''_v \times f \times E_c(B-V), \quad \text{(D5b)}$$

$E_c(B-V)$ is computed using the Calzetti et al. (2000) dust law,

$$A_n(0.6563) = 0.82 \times f \times A_c(V)\frac{R''_v}{R'_v}. \quad \text{(D5c)}$$

Therefore, we express the dust-corrected nebular line ($f_i(\mathrm{H}\alpha)$) and continuum flux ($f_i(\mathrm{cont})$) as follows:

$$f_i(\mathrm{H}\,\alpha) = f_{\mathrm{obs}}(\mathrm{H}\,\alpha) \times 10^{0.4(0.62\times f\times A_c(V))} \quad \text{(D6a)}$$

$$f_i(\mathrm{cont}) = f_{\mathrm{obs}}(\mathrm{cont}) \times 10^{0.4(0.82\times A_c(V))}, \quad \text{(D6b)}$$

where the subscript *obs* refers to the observed quantity while *i* refers to the intrinsic quantity. Since $\mathrm{EW}_i = f_i(\mathrm{H}\,\alpha)/f_i(\mathrm{cont})$, finally the dust-corrected Hα EW can be expressed as follows:

$$\log_{10}(\mathrm{EW}_i) = \log_{10}(\mathrm{EW}_{\mathrm{obs}}) + 0.4A_{\mathrm{c}}(V)(0.62f - 0.82). \quad \text{(D7)}$$

Next, we consider the dust correction for the $z = 0.1$ optical colours. Using Calzetti et al. (2000) attenuation law, we calculate the starburst reddening curve for these wavelengths using the following equation:

$$k'(\lambda) = 2.659\left(-2.156 + \frac{1.509}{\lambda} - \frac{0.198}{\lambda^2} + \frac{0.011}{\lambda^3}\right) + R'_v. \quad \text{(D8)}$$

This equation is different from equation (D1a), since this is valid for more bluer wavelengths between $0.12\,\mu\mathrm{m} < \lambda < 0.63\,\mu\mathrm{m}$. Similar to equation (D1b), we work out the attenuation for median wavelengths of the [340] and [550] filters (by definition the filter medians are, respectively, at 0.34 and 0.55 μm) as follows:.

$$f([340]) = f([340]_{\mathrm{obs}}) \times 10^{0.4\times 1.56A_c(V)} \quad \text{(D9a)}$$

$$f([550]) = f([550]_{\mathrm{obs}}) \times 10^{0.4\times 1.00A_c(V)}. \quad \text{(D9b)}$$

Dust-corrected fluxes are used to recalculate the [340]–[550] colours.

The median wavelengths of the $g_{0.1}$ and $r_{0.1}$ filters are, respectively, at 0.44 and 0.57 μm. Similar to [340] and [550] filters, we use equation (D8) to calculate the attenuation for $g_{0.1}$ and $r_{0.1}$ filters as follows:

$$f(g_i)_{0.1} = f(g_{\mathrm{obs}})_{0.1} \times 10^{0.4\times 1.25A_c(V)} \quad \text{(D10a)}$$

$$f(r_i)_{0.1} = f(r_{\mathrm{obs}})_{0.1} \times 10^{0.4\times 0.96A_c(V)}. \quad \text{(D10b)}$$

APPENDIX E: PEGASE SIMULATIONS OF STARBURST GALAXIES

Here, we describe the PEGASE simulations discussed in Section 5.3 to model the effects of starbursts in Hα EW versus [340]–[550] colour parameter space. We tune burst empirical parameters to maximize the number of high Hα EW objects. We consider four scenarios in our simulations as shown in Table E1. For each scenario, we model 100 galaxies and superimpose a single starburst on PEGASE model tracks with an IMF slope of $\Gamma = -1.35$ and a constant SFH.

PEGASE model galaxies are generated from an initial gas reservoir of 1 $\mathrm{M}_\odot$. Therefore, we normalize the total mass generated by the constant SFH and the starburst to 1 $\mathrm{M}_\odot$ in order to calculate the SFR for each time-step. These values are used to calculate SSPs with an IMF slope of $\Gamma = -1.35$ and upper and lower mass cut-offs set at 0.5 and 120 $\mathrm{M}_\odot$, respectively. We use a constant metallicity of 0.02 for all our simulations. The other parameters are kept similar at their default values as described in Section 3. Following this recipe, we generate the simulated galaxies with finer sampling of the time-steps around the time of burst to better resolve the effects of the bursts.

We use the simulated galaxies from Scenario 1 (Table E1) to further investigate the evolution of Hα EW and [340]–[550] colour of galaxies during a starburst phase. 40 galaxies are chosen at random from the simulations in Fig. E1 to show the deviation of the Hα EW values from the $\Gamma = -1.35$ IMF track. In a smooth SFH, the [340]–[550] colours are correlated with the age of the galaxies due to SPs moving away from the main sequence making the galaxy redder with time. However, galaxies undergo bursts at random times resulting them to deviate from the smooth SFH track at random times. Galaxies with higher burst fractions per unit time (f_{m}/τ_b), show larger deviations due to the extreme SFRs required to generate higher amount of mass within a shorter time period. The time-scale galaxies populate above the reference IMF track is small (in the order of $\lesssim$50 Myr) compared to the total observable time window of the galaxies at $z \sim 2$(∼3 Gyr). The Hα EW increases significantly soon after a burst within a short time-scale (in the order of few Myr) and decreases rapidly to be deficient in Hα EW compared to the reference constant SFH model. Once the burst has passed, the Hα EW increases at a slower phase until the tracks join the smooth SFH model.

In Fig. E1 (right-hand panel), we select all simulated galaxies to calculate the amount of time galaxies spend with higher Hα EW values ($0.05 < \Delta[\log_{10}(\mathrm{H}\,\alpha\ \mathrm{EW})]$) compared to models with smooth SFHs. We bin the galaxies according to the burst fraction per unit time to find that there is no strong dependence of it on ΔTime.

Table E1. Summary of scenarios investigated in our starburst simulations.

Scenario	Start of SFH (Myr)	Time of burst (Myr)	τ_b (Myr)	f_{m}
1	0	0–3250	100–300	0.10–0.30
2	0–2500	2585–3250	10–30	0.01–0.03
3	0–2500	2585–3250	10–30	0.10–0.30
4	0–2500	2585–3250	100–300	0.01–0.03

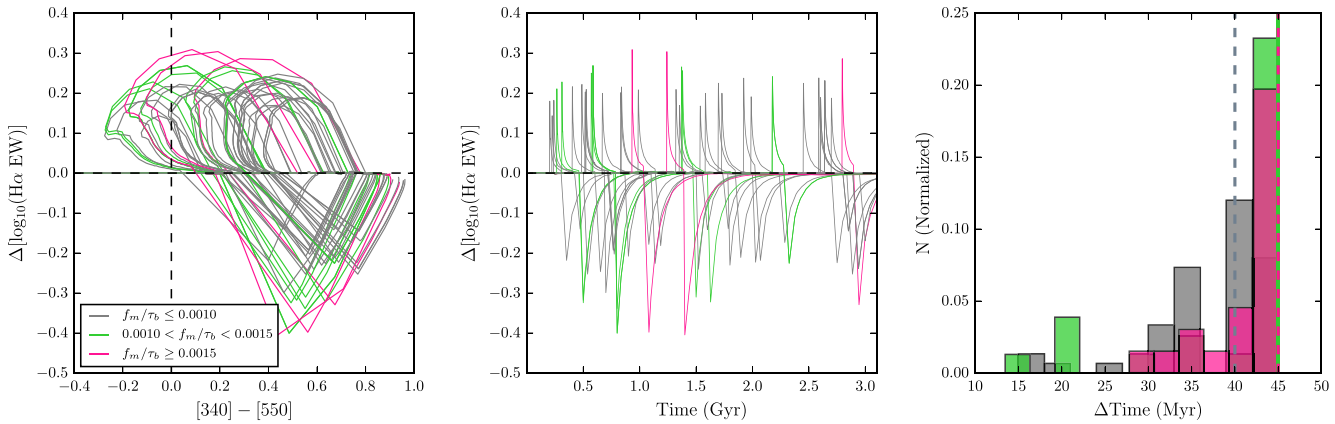

Figure E1. The Hα EW deviation of galaxies from our simulated sample (Table E1 Scenario 1). For each model galaxy, the deviations are calculated with respect to the $\Gamma = -1.35$ IMF constant SFH model at the same [340]–[550] colour. Left: Hα EW deviations as a function of [340]–[550] colour for 40 randomly selected galaxies. Galaxies are colour coded according to the fraction of stellar mass generated by the burst per unit time (measured in Myr^{-1}). The starbursts occur at random times, hence the galaxies deviate from the reference IMF track at random [340]–[550] colours. Centre: Hα EW deviations as a function of time. The galaxy sample is similar to the left-hand panel and are similarly colour coded according to the fraction of stellar mass generated by the burst per unit time. Right: the normalized histogram of the amount of time the starburst tracks stay at least 0.05 dex above the $\Gamma = -1.35$ smooth SFH model. All 100 galaxies in the simulation are shown here and have been colour coded according to the fraction of stellar mass generated by the burst per unit time. The vertical dashed lines show the median for each bin and are as follows: $\mu(f_\mathrm{m}/\tau_b \leq 0.0010) = 40$ Myr, $\mu(0.0010 < f_\mathrm{m}/\tau_b < 0.0015) = 45$ Myr and $\mu(f_\mathrm{m}/\tau_b \geq 0.0015) = 45$ Myr.

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