The Astrophysical Journal, 838:119 (15pp), 2017 April 1 https://doi.org/10.3847/1538-4357/aa61a6

Evolution of Dust-obscured Star Formation and Gas to $z = 2.2$ from HiZELS

A. P. Thomson[1], J. M. Simpson[2], Ian Smail[1], A. M. Swinbank[1], P. N. Best[2], D. Sobral[3], J. E. Geach[4], E. Ibar[5], and H. L. Johnson[1]
[1] Centre for Extragalactic Astronomy, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK; alasdair.thomson@durham.ac.uk
[2] Scottish Universities Physics Alliance, Institute for Astronomy, Royal Observatory Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK
[3] Department of Physics, Lancaster University, Lancaster LA1 4 YB, UK
[4] Centre for Astrophysics Research, Science & Technology Research Institute, University of Hertfordshire, Hatfield AL10 9AB, UK
[5] Instituto de Física y Astronomía, Universidad de Valparaíso, Avda. Gran Bretaña 1111, Valparaíso, Chile
Received 2016 July 14; revised 2017 February 8; accepted 2017 February 17; published 2017 March 31

Abstract

We investigate the far-infrared (far-IR) properties of galaxies selected via deep, narrow-band imaging of the Hα emission line in four redshift slices from $z = 0.40$–2.23 over $\sim$1 deg^2 as part of the High-redshift Emission Line Survey (HiZELS). We use a stacking approach in the *Herschel* PACS/SPIRE far-IR bands, along with 850 μm imaging from SCUBA-2 and Very Large Array 1.4 GHz imaging, to study the evolution of the dust properties of Hα-emitters selected above an evolving characteristic luminosity threshold, $0.2L^{\star}_{\rm H\alpha}(z)$. We investigate the relationship between the dust temperatures, $T_{\rm dust}$, and the far-infrared luminosities, $L_{\rm IR}$, of our stacked samples, finding that our Hα-selection identifies cold, low-$L_{\rm IR}$ galaxies ($T_{\rm dust} \sim 14$ K; $\log[L_{\rm IR}/L_{\odot}] \sim 9.9$) at $z = 0.40$, and more luminous, warmer systems ($T_{\rm dust} \sim 34$ K; $\log[L_{\rm IR}/L_{\odot}] \sim 11.5$) at $z = 2.23$. Using a modified graybody model, we estimate "characteristic sizes" for the dust-emitting regions of Hα-selected galaxies of $\sim$0.5 kpc, nearly an order of magnitude smaller than their stellar continuum sizes, which may provide indirect evidence of clumpy interstellar medium structure. Lastly, we use measurements of the dust masses from our far-IR stacking along with metallicity-dependent gas-to-dust ratios ($\delta_{\rm GDR}$) to measure typical molecular gas masses of $\sim 1 \times 10^{10}\,M_{\odot}$ for these bright Hα-emitters. The gas depletion timescales are shorter than the Hubble time at each redshift, suggesting probable replenishment of their gas reservoirs from the intergalactic medium. Based on the number density of Hα-selected galaxies, we find that typical star-forming galaxies brighter than $0.2L^{\star}_{\rm H\alpha}(z)$ comprise a significant fraction ($35 \pm 10\%$) of the total gas content of the universe, consistent with the predictions of the latest state-of-the-art cosmological simulations.

Key words: galaxies: evolution – galaxies: high-redshift – galaxies: ISM – galaxies: star formation

1. Introduction

It is now widely established that the star formation rate density of the universe increased from early times, reaching a peak around $z \sim 1$–2, and has been in steady decline ever since (e.g., Lilly et al. 1996; Hopkins & Beacom 2006; Madau & Dickinson 2014). Similar evolution has been claimed for the typical ratio of the star formation rate (SFR) and stellar mass ($M_{\star}$): the specific star formation rate (sSFR, e.g., Magdis et al. 2010). This has been linked, in part, to changes in the availability of molecular gas (the raw fuel for star formation) as a function of redshift (e.g., Genzel et al. 2010), and also potentially results in changes in the morphologies of star-forming galaxies, from smooth, disk-like structures in the local universe to more clumpy, irregular morphologies seen at $z \sim 2$ (e.g., Elmegreen et al. 2009; Swinbank et al. 2012).

However, existing studies of the evolving properties of star-forming galaxies are based on a range of techniques both for selecting star-forming galaxies, and for measuring their "instantaneous" SFRs (see, e.g., Kennicutt & Evans 2012). These include galaxies selected via their Lyα line emission (LAEs; Partridge & Peebles 1967), which are predominantly blue, less-massive galaxies (Oteo et al. 2015); via the "break" in their spectra near the Lyman limit at 912 Å (Lyman Break Galaxies; Steidel et al. 1996); through their rest-frame UV-optical colors (e.g., *BzK*; Daddi et al. 2004); and galaxies selected via their rest-frame far-infrared (IR) emission, which traces the dust heated by intense star formation episodes, such as the Ultra-Luminous InfraRed Galaxies (e.g., Lonsdale et al. 1990) and their high-redshift analogs, the submillimeter galaxies (SMGs; Smail et al. 1997; Barger et al. 1998; Casey et al. 2014). The varying utility of these different techniques, each with their own selection functions with redshift, SFR, and extinction has complicated the process of establishing a unified view of the cosmic star formation history.

Overcoming these limitations requires the use of homogeneous samples of star-forming galaxies selected from large-area, clean and deep multi-wavelength observations, which both minimize cosmic variance and circumvent the need to perform extrapolations down to faint luminosities. Narrow-band imaging techniques provide one means of making progress, as they can be used in sensitive, wide-field surveys to select large, representative samples of "typical" star-forming galaxies in a relatively clean way from a single emission line, which can act as an indicator for the SFR from low to high redshift. The Hα line ($\lambda_{\rm rest} = 656.3$ nm) is one of the most commonly employed SFR indicators, and has been used widely for star-forming galaxies at $z \lesssim 3$ (e.g., Kewley et al. 2002; Sobral et al. 2013). Hα emission arises predominantly from young, massive OB stars ($\lesssim$10 Myr old, and $\gtrsim 8\,M_{\odot}$), and can be used to measure the SFR, if corrected for the extinction ($A_{\rm H\alpha}$) due to scattering and absorption by dust (Kennicutt 1998).

Recent work in the optical/near-IR has revealed the existence of an apparent correlation between the SFR and stellar mass ($M_{\star}$), the so-called "main sequence," whose logarithmic slope is thought to be close to linear (Daddi et al. 2007; Elbaz et al. 2007), with evidence of a possible deviation to sub-linear slopes ($\sim$0.6–0.8; Whitaker et al. 2012; Magnelli

et al. 2014) at high $M_\star$. Recently, Schreiber et al. (2016) presented an analysis of the far-IR properties of a mass-selected sample of main sequence galaxies ($M_\star \gtrsim 10^{10}\,M_\odot$), using a stacking analysis to peer below the confusion limit of deep 250, 350, and 500 μm images, taken as part of the GOODS-*Herschel*[6] and CANDELS-*Herschel* key programs. They explain the flattening of the main sequence at high-$M_\star$ not as being due to high-$M_\star$ galaxies lacking fuel (i.e., low $M_{\rm gas}$), but rather due to their having a lower star formation efficiency (SFE $\equiv$ SFR/$M_{\rm gas}$). Possible explanations for this low SFE are that the SFR is artificially suppressed, either (i) by radio-mode active galactic nucleus (AGN) feedback (e.g., Bower et al. 2006), which drives gas out of a galaxy, preventing it from cooling to form stars, or (ii) by morphological quenching (e.g., Martig et al. 2009), wherein the internal kinematics of a galaxy's stellar disk set up differential torques, which act to prevent the cold gas clouds from fragmenting. An alternative explanation for flattening of the main sequence in high-$M_\star$ galaxies is simply that the most massive galaxies contain old stellar bulges, which are almost completely decoupled from the ongoing star formation (and the molecular gas reservoir that fuels it), and that by instead considering the stellar mass only in the disk, a constant main sequence slope emerges (Abramson et al. 2014).

To fully understand the evolution of star-forming galaxies, we must also investigate the link between star formation and its fuel supply, by tracking the typical molecular gas mass, $M_{\rm H_2}$, at different epochs. However, this is observationally challenging; H_2 lacks a strong dipole and thus does not radiate strongly (Carilli & Walter 2013). Tracer molecules such as ^{12}CO, the second most abundant molecule in the interstellar medium (ISM), have been observed out to $z \sim 6$ (e.g., Riechers et al. 2013), but are observationally expensive to detect, requiring long integrations with interferometers, hence it is difficult to construct statistical samples (e.g., Thomson et al. 2012; Bothwell et al. 2013). However, some progress can be made by using the relationship between dust and gas in local galaxies, which allows estimates of the gas mass to be obtained very quickly via observations of the Rayleigh–Jeans continuum (see Scoville et al. 2014 for a thorough discussion of this method).

Here we combine results from a narrow-band Hα survey with measurements of the dust content of Hα-selected galaxies in order to address the issue of the evolution of normal star-forming galaxies.

The High-redshift(Z) Emission Line Survey (HiZELS: Geach et al. 2008) conducted observations through specially designed, narrow-band filters on the Wide Field CAMera (WFCAM) on the United Kingdom Infra-Red Telescope (UKIRT[7]). HiZELS used narrow-band filters in the J, H, and K bands (NB_J, NB_H, and NB_K), corresponding to redshifted Hα emission at $z = 0.84$, 1.47, and 2.23, along with complementary observations with the Subaru telescope in the NB921 filter[8] (Sobral et al. 2012), which are sensitive to Hα emission at $z = 0.40$.

Prior to the launch of *Herschel*, far-IR studies of (SFR-selected) HiZELS galaxies by Geach et al. (2008) and Garn et al. (2010) relied on *Spitzer* 24, 70, and 160 μm imaging—in many cases, offering only upper limits—to constrain the dust properties in two redshift slices at $z = 0.84$ and $z = 2.23$. Later, Ibar et al. (2013) used *Herschel* PACS/SPIRE data covering the peak of the dust spectral energy distribution (SED) to investigate the far-IR properties of the HiZELS sample at $z = 1.47$, finding that Hα-selected galaxies are highly efficient star-forming systems, which lie somewhat above the main sequence.

In this paper, we build upon this earlier work by investigating the relationships between SFR, $M_\star$, Hα extinction ($A_{\rm H\alpha}$), dust mass ($M_{\rm dust}$), and temperature ($T_{\rm dust}$) across all four redshifts surveyed by HiZELS, in a self-consistent manner. We use the deepest available wide area images of the COSMOS and UDS fields in five far-IR bands covering the dust peak (100–500 μm), and supplement this with new photometry at 850 μm from the Submillimeter Common-User Bolometer Array 2 (SCUBA-2) to trace the dust mass, as well as 1.4 GHz radio imaging from the Karl G. Jansky Very Large Array (VLA). We employ stacking techniques to circumvent the limitations created by the poor angular resolution of these images, allowing us to study the dust properties of moderately star-forming galaxies (SFR $\sim 20\,M_\odot\,{\rm yr}^{-1}$) selected in a uniform manner across the full redshift range of the HiZELS survey.

This paper is laid out as follows: in Section 2 we present our analysis, including an outline of the sample selection, a description of the method used to correct the Hα luminosities of our sample for extinction, and details of our stacking analysis. In Section 3 we discuss our main results, beginning with the luminosities, dust masses, and temperatures derived from our SED fits, and a comparison of the IR and Hα-derived SFRs, which offers additional insight into the extinction of Hα emission by dust. In Section 4, we develop a framework in which to use the "fundamental metallicity relation (FMR)" to constrain the gas-to-dust ratios of Hα-selected galaxies, and so estimate the contribution made by Hα-selected star-forming galaxies to the total H_2 content of the universe. We give our main conclusions in Section 5. Throughout, we use *Planck* cosmology, with $H_0 = 70\,{\rm km\,s^{-1}\,Mpc^{-1}}$, $k = 0$, $\Omega_m = 0.3$, and $\Omega_\lambda = 0.7$, and assume a Chabrier (2003) initial mass function (IMF).

2. Analysis

2.1. Sample Selection and Observations

The starting point of our analysis is the catalog of 3004 Hα-selected sources from HiZELS presented in Sobral et al. (2013), comprising 1771 and 1233 Hα-selected star-forming galaxies in the COSMOS and UDS fields, respectively. These total 1108 galaxies at $z = 0.40$, 635 at $z = 0.84$, 511 at $z = 1.47$, and 750 at $z = 2.23$, down to typical SFR limits of $\sim$0.1, 1.0, 2.0, and 4 $M_\odot\,{\rm yr}^{-1}$, respectively. We summarize the properties of the sample in Table 1.

We exploit the multitude of mid/far-IR data available in the UDS and COSMOS fields, comprising: (i) 250, 350, and 500 μm *Herschel* SPIRE observations (see Section 2.3); and (ii) cold dust-sensitive 850 μm observations taken with SCUBA-2 on the James Clerk Maxwell Telescope (JCMT) as part of the SCUBA-2 Cosmology Legacy Survey (S2CLS: Geach et al. 2013, 2017). In addition, in the COSMOS field, we include 100 and 160 μm *Herschel* PACS observations from the

[6] *Herschel* was an ESA space observatory with science instruments provided by European-led Principal Investigator consortia and with important participation from NASA.
[7] UKIRT programs U/CMP/3 and U/10B/07.
[8] Subaru program S10B-144S.

Table 1
Properties of the HiZELS Sample

		$N_{\rm galaxies}$[a]	Volume (10^4 Mpc^3 deg^{-2})	SFR limit[b] ($M_\odot$ yr^{-1})	$N_{\rm stack}$[c]	$\log_{10}(L^\star({\rm H}\alpha))$ (erg s^{-1})	$\langle {\rm SFR}_{\rm H\alpha}\rangle$[d] ($M_\odot$ yr^{-1})	$\langle M_\star^{\rm H\alpha}\rangle$[e] ($\times 10^{10} M_\odot$)
$z = 0.40$		1108	5.13	…	52	$42.15^{+0.47}_{-0.12}$	2.1	3.2 ± 0.9
	COSMOS	445	…	0.1	36	…	…	…
	UDS	663	…	0.2	16	…	…	…
$z = 0.84$		635	14.65	…	397	$42.37^{+0.07}_{-0.05}$	6.9	0.9 ± 0.2
	COSMOS	425	…	0.9	240	…	…	…
	UDS	210	…	0.9	152	…	…	…
$z = 1.47$		511	33.96	…	449	$42.75^{+0.06}_{-0.05}$	26.0	1.2 ± 0.3
	COSMOS	323	…	1.9	274	…	…	…
	UDS	188	…	4.0	175	…	…	…
$z = 2.23$		750	38.31	…	535	$43.17^{+0.08}_{-0.06}$	34.4	2.3 ± 0.4
	COSMOS	578	…	3.5	388	…	…	…
	UDS	172	…	7.7	146	…	…	…

Notes.
[a] Total number of Hα-selected star-forming galaxies in each field, at each redshift.
[b] SFR limit determined by converting the faintest Hα luminosity in each subsample, using the Kennicutt (1998) conversion factor.
[c] Number of galaxies in each stack, selected with $L_{\rm H\alpha} \geqslant 0.2 L^\star_{\rm H\alpha}(z)$.
[d] Median dust-corrected Hα-derived SFR of galaxies contributing to stacked subsamples.
[e] Representative Hα-weighted stellar mass of each subsample (See Section 2.7).

PACS Evolutionary Probe (PEP; Lutz et al. 2011) survey, and 1.4 GHz radio continuum observations from the Very Large Array, taken as part of the VLA-COSMOS survey (Schinnerer et al. 2004).

To study evolution in the properties of "normal" star-forming galaxies, we select galaxies close to the knee of the Hα luminosity function, with dust-corrected luminosities $L_{\rm H\alpha} \geqslant 0.2 L^\star_{\rm H\alpha}(z)$ (where $L^\star_{\rm H\alpha}(z)$ denotes the characteristic "break" in the Hα luminosity function at each HiZELS redshift, measured by Sobral et al. 2014).

Dividing the HiZELS population on an evolving luminosity cut in this manner gives us four matched, SFR-selected subsamples, comprising typical star-forming galaxies spanning the characteristic Hα luminosity at each redshift.

2.2. Dust Corrections

In order to account for obscuration of Hα photons by dust within the ISM of each HiZELS galaxy, we perform an empirical dust-correction to the measured line fluxes. We begin with the observed Hα luminosities ($L_{\rm H\alpha,obs}$) and stellar continuum reddening values, $E(B-V)$, for each galaxy reported by Sobral et al. (2014). Next, we use the reddening law of Calzetti et al. (2000) to estimate the Hα extinction arising from diffuse dust, $A_{\rm H\alpha,cont} = \kappa(\lambda) A_V / R'_V$, where $A_V = 3 \times E(B-V)$ is the stellar extinction. For $\lambda_{\rm H\alpha,rest} = 656$ nm, $\kappa(\lambda)$ is expressed as

$$\kappa(656\ {\rm nm}) = 2.659\left(-1.857 + \frac{1.040}{\lambda/\mu{\rm m}}\right) + R'_V. \quad (1)$$

$R'_V = 4.08 \pm 0.88$ is the typical effective obscuration at V-band (Calzetti et al. 2000). In addition to the extinction from the diffuse ISM, observations of nearby starburst galaxies by Calzetti et al. (1994, 2000) indicate the need for an "extra" extinction component, $A_{\rm H\alpha,ext}$, to account for attenuation occurring locally in the birth clouds around young OB stars. Using deep, multi-wavelength *Hubble Space Telescope* (*HST*) data from the CANDELS project, Wuyts et al. (2013) propose an empirical method to estimate $A_{\rm H\alpha,extra}$, which is based in part on the dependency of the average molecular cloud mass on the galaxy-integrated gas fraction:

$$A_{\rm H\alpha,ext} = A_{\rm H\alpha,cont}\,(0.90 - 0.15 A_{\rm H\alpha,cont}). \quad (2)$$

We hence determine the intrinsic (i.e., dust-corrected) Hα luminosities of our Hα-selected galaxies, $L_{\rm H\alpha,int}$ (hereafter, $L_{\rm H\alpha}$) from their uncorrected luminosities ($L_{\rm H\alpha,obs}$) using:

$$\log_{10}(L_{\rm H\alpha,int}) = \log_{10}(L_{\rm H\alpha,obs}) + 0.4 A_{\rm H\alpha,cont} + 0.4 A_{\rm H\alpha,ext}. \quad (3)$$

2.3. Herschel SPIRE Deblending

The *Herschel* SPIRE mosaics of the UDS and COSMOS fields were taken as part of the HerMES survey (Oliver et al. 2012), and were retrieved from the DR2 HerMES data release. Due to the coarse resolution of the SPIRE maps (15″, 22″, and 32″ at 250, 350, and 500 μm, respectively), the images need to be deblended before reliable flux densities can be derived. This deblending allows us to overcome the confusion limit, which smears together nearby sources, and can bias flux density measurements—even in stacks (e.g., Oliver et al. 2012; Magnelli et al. 2013).

To create the prior catalog, we used the *Spitzer*/MIPS 24 μm photometric catalogs from Magnelli et al. (2013) which are derived by simultaneous point-spread function fitting to the prior positions at IRAC/3.6 μm. The 24 μm catalogs were limited to a 3σ detection limit of 50 μJy. To deblend the images, we followed the same procedure described in Swinbank et al. (2014) (see also Stanley et al. 2015). For each of the SPIRE bands, we created a model image in which flux was added at the positions of the galaxies in the prior catalog and convolved with the SPIRE beam, before a residual map (*data* − *model*) was created. The fluxes of the galaxies in the model image were then randomly perturbed, and the process repeated until the residual map converged on a minimum. Since

~10% of the Hα-selected galaxies in HiZELS are also 24 μm sources, their flux has been removed from the image when creating the residual image during the deblending process and so this flux needs to be added back into the map before stacking. We therefore reinsert the flux of each 24 μm detected HiZELS galaxy back into each of the SPIRE residual images using the appropriate PSF. This finally leaves us with a deblended, residual image (that includes the HiZELS galaxies that were also 24 μm sources) that can be stacked to investigate the far-infrared properties of our sample. The decision to use 24 μm/radio-detected galaxies as priors for the deblending is motivated by the assumption that those galaxies that dominate the mid-IR/radio bands will account for the majority of the total flux in the SPIRE images. However, the inherent clustering of field galaxies which are *not* in the prior catalog (but do still contribute flux to the SPIRE images) could bias the results of our stacks, if not carefully accounted for. Due to the increasing beam size at longer wavelengths, we may expect any such bias to be more severe at 500 μm than at shorter wavelengths. We performed a series of tests (Appendix A.1) to quantify this effect, but found the bias in our stacked flux densities due to clustering to be comparable to (or smaller than) the statistical uncertainties on the photometry. Moreover the "bias" was not found to be dependent on wavelength; we therefore opted not to apply a systematic correction to our flux densities after stacking in the residual images.

The smaller beam sizes in the PACS 100/160 μm, SCUBA-2 870 μm, and VLA 1.4 GHz images, coupled with the low surface density of 870 μm sources (Chen et al. 2016) mean that the images at these wavelengths do not require deblending. Hence, we stack at the positions of HiZELS galaxies directly in the calibrated PACS, SCUBA-2, and VLA maps, and in the residual, deblended SPIRE images.

2.4. AGN Contribution

While the Hα line is commonly used as a SFR indicator, the intense UV radiation fields that are responsible for its production can also be found in the vicinity of AGNs; therefore it is prudent to consider the contribution made by non star formation-dominated Hα-selected galaxies to the stacked far-IR flux densities. We identify 41 Hα-emitting AGNs by searching for X-ray sources within 1″ of each HiZELS galaxy in the *Chandra* COSMOS Legacy Survey (Civano et al. 2016) and Subaru-*XMM* Deep Survey (SXDS Ueda et al. 2008) catalogs. Of these 41 AGNs, none fall within the $L_{\rm H\alpha} \geqslant 0.2L^{\star}_{\rm H\alpha}$ sample at $z = 0.40$, however there is 1 X-ray AGN in the sample at $z = 0.84$, 11 (2.4% of the sample) at $z = 1.47$, and 7 (1.2%) at $z = 2.23$. We note that, of the 41 X-ray AGNs, only 16 are individually detected in any far-IR/radio band (see also Calhau et al. 2017). In addition to the 41 X-ray AGNs, we identify a further 17 candidate AGNs via their mid-IR colors (0, 5, 5, and 7 at the respective HiZELS redshifts), following the technique of Donley et al. (2012).

We performed our stacking analysis both with these sources present and with them excluded, finding the fluxes in both cases to be consistent within the errors. In an analysis which involved stacking all the candidate AGNs in HiZELS at $z = 1.47$, Ibar et al. (2013) found the far-IR SEDs of AGNs to be potentially more luminous ($\log[L_{\rm IR}/L_{\odot}] \sim 11.6$) and warmer ($\Delta T_{\rm dust} \sim 7$ K) on average than purely star-forming galaxies, but with significant uncertainties due to the small sample size. Given the small number of X-ray or IRAC-identified AGNs in our $>0.2L^{\star}_{\rm H\alpha}(z)$ subsamples, the lack of any measurable effect on the SEDs regardless of whether or not we exclude them from our stacks is therefore unsurprising. Nevertheless, we exclude all 58 candidate AGNs from our subsequent analysis.

2.5. Stacking Analysis

We use the deepest available extragalactic far-IR/sub-mm observations from *Herschel* and SCUBA-2. Focusing our study only on those Hα-selected galaxies that are individually detected in the far-IR/sub-mm data would introduce a strong bias toward galaxies with the most extreme SFRs, characteristic of "starburst" systems; in order to study *typical* star-forming galaxies, and understand their dust properties, it is necessary to peer below the confusion limit in the far-IR wave-bands, and use a stacking approach.

Given the similar depths of the original Hα observations from which the COSMOS and UDS samples were drawn ($\log_{10}[L_{\rm H\alpha}/{\rm erg\ s^{-1}}] \geqslant 39.95$), we opt to simultaneously stack in both fields.

We begin by separating the COSMOS and UDS galaxies, extracting thumbnails around each galaxy above our $0.2L^{\star}_{\rm H\alpha}(z)$ cuts, and running a median stacking algorithm at each wavelength on these thumbnails. In each of the far-IR/sub-mm maps, we account for the background emission by subtracting the median flux of 1000 random positions from the map. In the PACS 100 and 160 μm bands, we measure fluxes by summing the pixel values of the stacks within a 7.″2 or 12″ aperture, respectively, and multiplying by the empirically derived aperture correction ($a(\lambda)$) and high-pass filtering correction ($g(\lambda)$) specified in the PACS PEP data release notes.[9]

In the deblended SPIRE 250, 350, and 500 μm and SCUBA-2 850 μm bands, the angular resolution is low enough (⩾15″) that we do not resolve the emission; hence we measure fluxes from the peak pixel value within 1″ of the centroid of each stack (to allow for any small systematic misalignments in the astrometry of the images).

At the native 1.″5 resolution of the VLA COSMOS image, we anticipate that some of our sources may indeed be resolved; because we are only interested in measuring the total fluxes of the stacks, we convolve the VLA image with a 5″ Gaussian kernel prior to stacking, using the Astronomical Image Processing System (*AIPS*) task CONVL (setting the FACTOR parameter to scale the map by the ratio of beam areas, in order to preserve the absolute flux scale). We then measure the 1.4 GHz flux densities of each of the stacked, smoothed radio images by measuring the peak pixel value within 1″ of the centroid of the stack.

In addition to creating the median stacks at each wavelength, we also create error images by bootstrap re-sampling each of the thumbnails that are used in the stacking procedure. We determine the flux uncertainties in our stacks from these error images in the same manner as we measured the fluxes from the corresponding stacks, i.e., aperture photometry in the PACS images, and taking peak pixel values in the SPIRE, SCUBA-2 and VLA images. The derived flux densities are reported in Table 2.

[9] http://www.mpe.mpg.de/resources/PEP/DR1_tarballs/readme_PEP_global.pdf

Table 2
Stacked Photometry

	$S_{100\,\mu m}$ (mJy)	$S_{160\,\mu m}$ (mJy)	$S_{250\,\mu m}$ (mJy)	$S_{350\,\mu m}$ (mJy)	$S_{500\,\mu m}$ (mJy)	$S_{850\,\mu m}$ (mJy)	$S_{1.4\,GHz}$ (μJy)
$z = 0.40$	<4.4	<7.6	4.3 ± 0.7	3.5 ± 0.7	4.1 ± 0.6	1.0 ± 0.3	<13.3
$z = 0.84$	2.5 ± 0.6	3.1 ± 1.0	3.2 ± 0.5	3.1 ± 0.5	2.8 ± 0.4	<0.4	8.9 ± 1.8
$z = 1.47$	<1.7	<2.7	3.3 ± 0.4	3.9 ± 0.6	2.8 ± 0.5	<0.5	<7.9
$z = 2.23$	<1.4	<2.4	2.1 ± 0.4	2.2 ± 0.5	1.8 ± 0.4	0.4 ± 0.1	<6.1

Note. Non-detections are represented as 3σ limits.

2.6. SED Fitting

We create a set of composite photometry at each redshift, comprising the combined flux densities at 250, 350, 500, and 850 μm from COSMOS and UDS (weighted by the number of sources in each field), and the 100 and 160 μm, plus 1.4 GHz fluxes from the COSMOS-only stacks (since we lack coverage in these bands in the UDS field). Combining the photometry in this manner has the effect of reducing the flux uncertainties near the peak of the SED.

We measure the properties of our galaxy stacks from this set of photometry by fitting isothermal modified blackbody (graybody) templates[10] of the form

$$S(\nu) \propto \frac{\nu^{\beta+3}}{\exp(h\nu/kT_{\rm d}) - 1} \quad (4)$$

between 100 and 850 μm, where β is the dust emissivity index. We account for bands in which the galaxy is not detected by extending the formalism of Sawicki (2012) to the far-IR and sub-mm fluxes. A full description of this methodology can be found in Sawicki (2012), but briefly, we measure the goodness-of-fit of graybody templates to the photometry (comprising detections in bands i and non-detections in bands j, respectively) by adopting the modified χ^2 statistic:

$$\chi^2 = \sum_i \left(\frac{f_{d,i} - sf_{m,i}}{\sigma_i}\right)^2 - 2\sum_j \ln \int_{-\infty}^{f_{{\rm lim},j}} \exp\left[-\frac{1}{2}\left(\frac{f - sf_{m,j}}{\sigma_j}\right)^2\right] df \quad (5)$$

where $f_{m,i}$ is the model flux density in the ith band, $f_{d,i}$ is the measured flux density in the same band, σ_i is the uncertainty on the observed flux density, s is the flux scaling between the model and the data (a parameter which is fit numerically[11]), and $f_{{\rm lim},j}$ is the 1σ flux uncertainty in the jth band. In the case where the stack is detected in every band (i.e., there are no upper limits), the second sum is set to zero, and the fitting reduces to a simple χ^2 determination.

The functional form of the modified blackbody curve is such that its partial derivatives with respect to β and $T_{\rm dust}$ are correlated, and thus exact values for either parameter cannot be analytically determined without first fixing (i.e., assuming) the other. While methods exist in the literature to numerically disentangle this correlation (e.g., Kirkpatrick et al. 2013; Hunt et al. 2015), data of high signal-to-noise ratio (S/N $\geqslant$ 10) are typically required. In the present case, the S/N of our far-IR/sub-mm stacks is insufficient to break this degeneracy; to facilitate comparison with the literature, we therefore fix $\beta = 1.5$, but note that allowing for a range of $\beta = 1$–2.5 (as reported in recent works, e.g., Casey et al. 2011; Chapin et al. 2011) would introduce an additional uncertainty on all quantities derived from the SED fits (e.g., Ibar et al. 2015).

We estimate the uncertainties in our isothermal graybody fits using a Markov Chain Monte Carlo (MCMC) approach; we model the stacked photometry for each HiZELS subsample using the isothermal graybody described in Equation (4) and use the affine-invariant, MCMC sampler, EMCEE (Foreman-Mackey et al. 2013). For each set of photometry, we employ 50 "walkers" for a combined total of 10^6 steps. The "burn-in" phase is considered over after 5000 steps of each walker. This is a conservative approach, and an investigation of the resultant data indicates that the "burn-in" phase is indeed complete and that the chain is well-mixed. The best-fit parameters are taken as those corresponding to the maximum likelihood sample of the chain, and the uncertainties are the 16–84th percentile of each parameter distribution.

In addition to fitting this isothermal graybody component (from which $T_{\rm dust}$, $M_{\rm dust}$, and $L_{\rm IR}$ are measured), we also fit to the data (including the radio) models from a suite of 184 template SEDs from Chary & Elbaz (2001), Draine & Li (2007), and Rieke et al. (2009), plus the SEDs of the starburst galaxies M 82, Arp 220, and SMM J21352–0102 (Swinbank et al. 2010). In each case, the template SEDs are set to the redshift of the stack, and each template is re-scaled in flux to fit the stacked photometry. The range of best-fitting templates for each stack that satisfies $|\chi^2_{\rm reduced} - \chi^2_{\rm reduced,min}| \leqslant 1$ is shown in Figure 1. While the templates implicitly account for both cold and warm dust—and hence provide the most secure estimate of the total 8–1000 μm luminosity, $L_{8-1000\,\mu m}$—in order to facilitate comparison with the literature, we hereafter focus primarily on the cold dust

[10] We also try fitting dual-temperature graybody SEDs, and check for improvements in the fit by measuring the Bayesian Information Criterion parameter, BIC $\equiv \chi^2 + k \ln N$, where k is the number of degrees of freedom of the model and N is the number of data points constraining the fit (Wit et al. 2012). BIC favors models that improve χ^2, but penalizes models that require several extra degrees of freedom to deliver marginal improvements in χ^2. We find in three out of four cases that the BIC formally favors an isothermal fit, however at $z = 0.84$, the two-component fit lowers BIC by ~50%. The total luminosity and dust mass of the $z = 0.84$ two-component fit are $\log_{10}(L_{\rm IR}/L_\odot) = (10.95 \pm 0.15)$ and $M_{\rm dust} = (1.0 \pm 0.8) \times 10^8\,M_\odot$, respectively, and the luminosity-weighted temperature is $T_{\rm dust} = 35 \pm 7$ K. In the interest of consistency across all four redshifts, we hereafter use properties measured from the single-temperature dust SEDs, but incorporate the differences between properties measured from the isothermal and dual-temperature fits into the errors on all derived quantities in Table 3 as a model-dependent systematic uncertainty.

[11] In the case where the stack is detected in each band, scaling the best-fit template to the photometry is trivial; however in cases with non-detections, we need to numerically find the value of s for which $\partial\chi^2/\partial s = 0$.

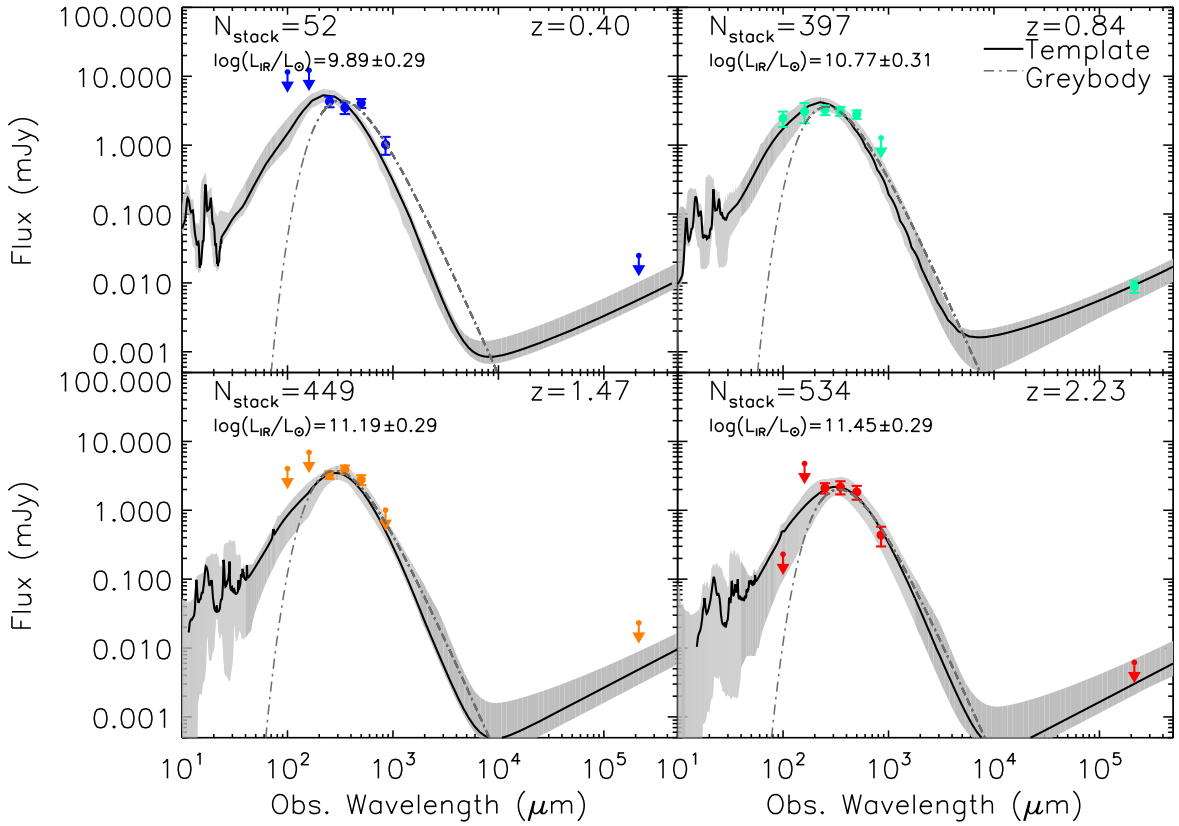

Figure 1. Far-infrared/sub-mm spectral energy distributions of stacks comprising all HiZELS star-forming galaxies with $L_{\mathrm{H}\alpha} \geqslant 0.2L^{\star}_{\mathrm{H}\alpha}(z)$, in each of the four redshift slices. Stacked photometry is shown with colored points, non-detections are represented as 3σ upper limits with arrows. We show the best-fitting template SED for each stack, along with the range of templates within $\chi^2_{\mathrm{red}} = 1$ as a gray band. The best-fitting isothermal graybody is shown with a dotted–dashed line (see Section 2.6 for details). The templates implicitly account for the effects of warm dust (which dominates the SED at $\lambda_{\mathrm{rest}} \lesssim 100\ \mu$m wavelengths); however, in order to facilitate comparison with the literature, we take L_{IR} to be the the integral under the isothermal graybody, representing the dominant cold dust component only.

Table 3
Properties of Stacked Hα-selected Samples

	$\log_{10}(L_{\mathrm{IR}})$ ($L_{\odot}$)	$T_{\mathrm{dust,BB}}$ (K)	$T_{\mathrm{dust,Tem}}$ (K)	$\log_{10}(L_{\mathrm{IR}}/L_{\mathrm{H}\alpha})$	M_{dust} ($\times 10^8\ M_{\odot}$)	$A_{\mathrm{H}\alpha,\mathrm{cont}}$	$A_{\mathrm{H}\alpha,\mathrm{ext}}$
$z = 0.40$	9.89 ± 0.29	14 ± 3	22 ± 1	-1.80 ± 0.05	4.5 ± 2.4	0.92	0.70
$z = 0.84$	10.77 ± 0.31	25 ± 6	32 ± 6	-2.16 ± 0.06	1.3 ± 0.8	1.18	0.85
$z = 1.47$	11.19 ± 0.29	27 ± 7	30 ± 7	-2.01 ± 0.05	2.2 ± 1.1	1.05	0.78
$z = 2.23$	11.45 ± 0.29	34 ± 8	34 ± 7	-2.15 ± 0.06	1.2 ± 0.6	0.92	0.70

Note. Uncertainties on all far-IR derived properties are measured by adding (in quadrature) the statistical uncertainties from the isothermal graybody fits to the model-dependent systematic, measured as the offset between the isothermal and dual-temperature fits (Section 2.6).

component, and take L_{IR} to be the integral under the isothermal graybody curves (unless noted otherwise). The average ratio between the two is $\log_{10}(L_{8-1000\ \mu\mathrm{m}}) = 1.03 \times \log_{10}(L_{\mathrm{IR}})$.

We measure dust masses, M_{dust}, from our stacked far-IR/sub-mm SEDs via the monochromatic flux density S_{ν} (at frequency ν). M_{dust} and S_{ν} are related through the relationship

$$S_{\nu} \propto \kappa_{\nu} B_{\nu}(T)(1+z) M_{\mathrm{dust}} / D_{\mathrm{L}}^2 \quad (6)$$

where κ_{ν} is the frequency-dependent dust absorption coefficient, $B_{\nu}(T)$ is the Planck function at temperature T, and D_{L} is the luminosity distance to the source (e.g., Casey et al. 2012). In the interest of consistency, we choose to measure M_{dust} from the flux density of the best-fitting modified blackbody at the same rest-frame wavelength (850 μm) in each case. We measure $S_{850\ \mu\mathrm{m}}$ in each case by interpolating the best-fit graybody (i.e., with $\beta = 1.5$), and fix $\kappa_{850\ \mu\mathrm{m}} = 0.07 \pm 0.02\ \mathrm{m^2\ kg^{-1}}$, which James et al. (2002) find to provide a suitable fit to both dwarf and massive star-forming galaxies in the SCUBA Local universe Galaxy Survey. We report the implied dust masses in Table 3;[12] uncertainties on the dust masses are calculated by propagating the uncertainties on T_{d} from the SED fits through Equation (6).

2.7. The Stellar Masses of Stacked Samples of Hα-emitters

In order to relate the far-IR properties of our stacked Hα-selected samples to their stellar content, it is necessary to determine the "representative" stellar mass of each stacked sample. Sobral et al. (2014) measured the stellar masses of

[12] Recently, Clark et al. (2016) measured the equivalent quantity at 500 μm in 22 massive, dusty galaxies from the *Herschel* Reference Survey, finding $\kappa_{500} = 0.051\ \mathrm{m^2\ kg^{-1}}$ (equivalent to $\kappa_{850} \sim 0.02$). Adopting this value instead would increase our dust masses by a factor ~2–3×.

individual HiZELS galaxies via multi-band UV–mid-IR SED fits. While one approach to calculating the "representative" stellar mass for our stacks would be to simply calculate the median mass of the galaxies in each subsample, this does not take into account the fact that, in our SFR-selected subsamples, the contribution each galaxy makes to the far-IR flux densities is a function of SFR. A more appropriate approach, therefore, is to weight the stellar masses of our galaxies by $L_{\rm H\alpha}$, such that the characteristic stellar mass in each stacked subsample, $\langle M_\star^{\rm H\alpha}\rangle$, is:

$$\langle M_\star^{\rm H\alpha}\rangle = \frac{\sum_i L_{{\rm H\alpha},i} M_\star^i}{\sum_i L_{{\rm H\alpha},i}}. \quad (7)$$

These Hα luminosity-weighted stellar masses are $\sim$1–3 $\times$ 10^{10} $M_\odot$, and are around $\sim$7$\times$ higher than the simple medians of the stellar masses of the galaxies in each stack. We estimate the typical uncertainties on these weighted masses by bootstrap re-sampling the errors on the numerator and the denominator, and propagating them.

We test the reliability of this weighting scheme by splitting each $>0.2L^\star(z)$ sample into two smaller subsamples, comprising the brightest and faintest 50% of Hα-luminosity galaxies above $0.2L^\star(z)$ at each redshift. We then stack the far-IR photometry for these and derive $L_{\rm IR}$ for each subsample from the stacked photometry as before. We estimate the IR-weighted representative stellar masses of our $>0.2L^\star(z)$ subsamples from these fits as

$$\langle M_\star^{\rm IR}\rangle = \frac{\langle M_\star^{\rm H\alpha\ bright}\rangle \langle L_{\rm IR}^{\rm H\alpha\ bright}\rangle + \langle M_\star^{\rm H\alpha\ faint}\rangle \langle L_{\rm IR}^{\rm H\alpha\ faint}\rangle}{\langle L_{\rm IR}^{\rm H\alpha\ bright}\rangle + \langle L_{\rm IR}^{\rm H\alpha\ faint}\rangle}. \quad (8)$$

We measure $\langle M_\star^{\rm IR}\rangle = (7.5 \pm 1.9) \times 10^9\ M_\odot$, $(7.3 \pm 2.0) \times 10^9\ M_\odot$, $(10.3 \pm 1.7) \times 10^9\ M_\odot$, and $(11.6 \pm 4.8) \times 10^9\ M_\odot$ in each of our four samples, with ascending redshift, in excellent agreement with the Hα-weighted stellar masses at $z = 0.84$ and $z = 1.47$ (within 25%), and within a factor $\sim$2$\times$ of the Hα-weighted mass at $z = 2.23$. At $z = 0.40$, the discrepancy between the Hα-weighted and IR-weighted stellar masses is a factor $\sim$4$\times$; however, much of this may be explained by the small number of $>0.2L^\star_{\rm H\alpha}(z)$ HiZELS galaxies at this redshift (52), which results in the stacked SEDs for the two halves of this sample being characterized by upper limits at 100, 160, and 350 μm. This hence leads to large uncertainties in $L_{\rm IR}$. We hereafter use our Hα-weighted stellar masses to characterize our samples, and report these in Table 1.

3. Results and Discussion

3.1. SED Fits

We show our stacked photometry in Figure 1, along with the best-fitting template and graybody SEDs. Each of our Hα-selected subsamples is detected in all three SPIRE bands, while the PACS data (which cover only the COSMOS field) typically present only upper limits (with $>3\sigma$ detections in both bands only at $z = 0.84$). Properties measured from these SED fits are presented in Table 3. The measured infrared luminosities of our Hα-selected galaxies increase from $\log_{10}[L_{\rm IR}/L_\odot] = 9.9$–$11.5$ with redshift, while the graybody dust temperatures increase from $T_{\rm dust} = 14$–34 K.

In our three higher redshift bins, we see that the best-fitting isothermal graybody is consistent with the $\pm1\sigma$ range of template SED fits. At $z = 0.40$, the isothermal graybody is $\sim$8 K colder than the best-fitting template ($T_{\rm graybody} = 14 \pm 3$ K versus $T_{\rm template} = 22 \pm 1$ K). The coldest available template in the template library is 20 K, which indicates that much of this discrepancy can be accounted for by the template library's inadequate sampling of sufficiently cold dust temperatures. The $z = 0.40$ graybody fit is also significantly colder than any of the 500 individually far-IR detected $z = 0$–3 star-forming galaxies observed by Genzel et al. (2015). However, our $z = 0.40$ Hα-selected galaxies are typically a factor 10$\times$ less massive than the galaxies detected by Genzel et al. (2015), and have a lower SFR $\sim 2\ M_\odot\ {\rm yr}^{-1}$.

We see no evidence for strong evolution of $M_{\rm dust}$ with redshift, with each of our stacked SEDs for Hα-selected star-forming galaxies being consistent with the mean dust mass of $(1.6 \pm 0.1) \times 10^8\ M_\odot$ measured by Rowlands et al. (2014) for a sample of $z < 0.5$ dusty star-forming galaxies.

3.2. Extinction Properties of Hα Emitters

In applying individual dust corrections to each of our Hα-selected star-forming galaxies based on the Calzetti redenning law, we have potentially altered the compositions of the samples in our $>0.2L^\star(z)$ stacks relative to samples selected with no extinction correction applied; any non-zero dust correction will tend to scatter galaxies with lower observed Hα fluxes above our luminosity cut-offs, which would otherwise not have surpassed this threshold.

To investigate whether this effect has in any way biased our conclusions, we now perform a series of tests comparing the properties of our dust-corrected subsamples with analogous subsamples to which *no* dust correction is applied. The median $E(B-V)$ of galaxies derived via multi-wavelength SED fits in the *corrected* subsamples are 0.35 at $z = 0.40$, 0.45 at $z = 0.84$, 0.40 at $z = 1.47$, and 0.35 at $z = 2.23$, corresponding to $A_V = 1.12$, 1.44, 1.28, and 1.12, respectively. The Hα extinctions ($A_{\rm H\alpha}$) measured from this analysis (via the process outlined in Section 2.2), along with the number of HiZELS galaxies above our $0.2L^\star(z)$ thresholds are listed in Table 3.

We also select (and stack) comparison subsamples consisting of the same number of galaxies at each redshift, representing those galaxies with the highest *uncorrected* $L_{\rm H\alpha}$, and fit far-IR SEDs to the photometry in the manner set out in Section 3.1.

In Figure 2 we plot the ratio of the median Hα luminosity of galaxies in each of these subsamples to $L_{\rm IR}$, as measured from the best-fit isothermal graybodies. If no $A_{\rm H\alpha}$ correction is applied, the ratio $\log_{10}[L_{\rm H\alpha}/L_{\rm IR}]$ provides a proxy for the amount of reddening due to dust. We see that this ratio remains approximately constant from $z = 0.40$–2.23 for both the corrected and uncorrected subsamples. This argues for no significant evolution in the reddening of "typical" Hα-selected galaxies as a function of redshift (see also Garn et al. 2010; Sobral et al. 2012; Ibar et al. 2013), with a scalar offset between the corrected and uncorrected subsamples of $\Delta \log_{10}[L_{\rm H\alpha}/L_{\rm IR}] \sim 0.2$.

Next, we compare $\log_{10}[L_{\rm H\alpha}/L_{\rm IR}]$ with the observed median A_V of each subsample. The lack of any strong correlation between the ratio $\log_{10}[L_{\rm H\alpha}/L_{\rm IR}]$ and A_V seen in Figure 2 for our uncorrected subsamples is consistent with the lack of correlation seen in corrected subsamples, and again suggests that stacking on sub-populations defined relative to $L^\star_{\rm H\alpha}$ selects similar galaxies at all redshifts (albeit with a tendency to select intrinsically more luminous galaxies at higher redshift as $L^\star(z)$ evolves). However, the subsamples selected on the basis of

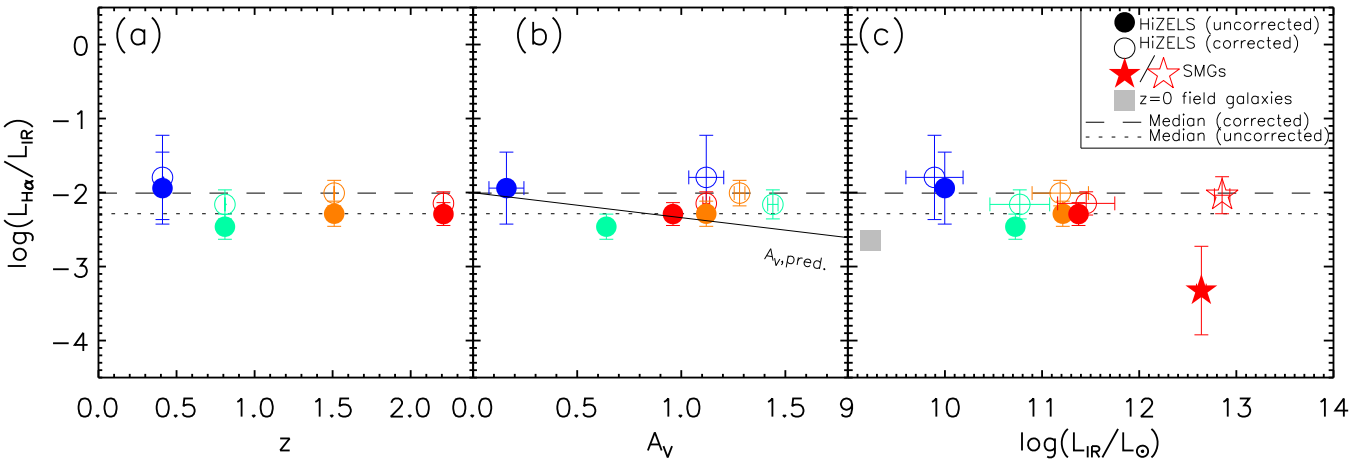

Figure 2. (a) We show the relationship between the extinction proxy $\log_{10}[L_{H\alpha}/L_{IR}]$ and redshift. We see no evolution in the ratio $\log_{10}[L_{H\alpha}/L_{IR}]$—using samples selected via either their extinction corrected (see Section 3.2) or uncorrected $L_{H\alpha}$—with redshift, suggesting that our selection of galaxies above an evolving Hα luminosity cut yields subsamples with similar ISM properties at different redshifts. (b) We compare $\log_{10}[L_{H\alpha}/L_{IR}]$ as a reddening proxy with the measured stellar extinctions, A_V, finding our extinction-corrected Hα-selected stacks to have similar A_V. We plot the empirically derived relationship between the two quantities, based on the $L_{H\alpha}$- and L_{IR}-to-SFR conversion factors of Kennicutt & Evans (2012) and Kennicutt (1998), plus the Calzetti et al. (2000) reddening law, and find that this exactly intersects our sample. (c) If $\log_{10}[L_{H\alpha}/L_{IR}]$ is a suitable proxy for the dust extinction, then we may expect to see a negative correlation between this ratio and L_{IR}, given the latter's connection with the cool dust component of the ISM. We see no strong trend, which may indicate that very heavily extinguished (low $L_{H\alpha}/L_{IR}$), dusty star-forming galaxies do not dominate our stacks. We show the median results from the dust-corrected samples of local field galaxies (Kewley et al. 2002) and SMGs (Takata et al. 2006, open star), as well as the non-corrected SMG sample of Swinbank et al. (2004, filled star). In each panel, the median corrected and uncorrected $\log_{10}[L_{H\alpha}/L_{IR}]$ of our Hα-selected subsamples are also shown. The color coding is the same as in Figure 1.

their *uncorrected* Hα luminosities have a much lower typical A_V than those in the corrected subsamples. This suggests that applying individual dust-corrections to the Hα luminosities (measured from their stellar continuum fits) prior to stacking is an important step, as it allows high-SFR (but dusty) galaxies to satisfy our $0.2L_{H\alpha}(z)$ criterion, which otherwise would be excluded on the basis of their low (observed) Hα fluxes.

In Figure 2, as a sanity check, we also show the implied relationship between $\log_{10}[L_{H\alpha}/L_{IR}]$ and $A_{V\,pred}$, where $A_{V\,pred}$ is the predicted extinction derived (for the uncorrected sample) by equating the L_{IR}-to-SFR and $L_{H\alpha}$-to-SFR indicators of Kennicutt & Evans (2012) and Kennicutt (1998),[13] respectively, and solving for the "intrinsic" luminosity ratio that would be seen if there were no dust extinction:

$$\begin{aligned}\log(\mathrm{SFR}/M_\odot\ \mathrm{yr}^{-1}) &= \log(L_{H\alpha}/\mathrm{erg\ s}^{-1}) - 41.36 \\ &= \log(L_{IR}/\mathrm{erg\ s}^{-1}) - 43.41 \end{aligned} \quad (9)$$

such that

$$\log(L_{H\alpha}/L_{IR})|_{int} = -2.05. \quad (10)$$

Next, we compute $A_{H\alpha}$ (again, for the uncorrected subsamples) by comparing the observed and intrinsic $L_{H\alpha}$-to-L_{IR} ratios

$$A_{H\alpha} \equiv \frac{\log(L_{H\alpha}/L_{IR})|_{obs} - \log(L_{H\alpha}/L_{IR})|_{int}}{0.4} \quad (11)$$

before finally deriving $A_{V,\,pred}$ via the Calzetti et al. (2000) law, $A_{V,\,pred} = A_{H\alpha}R'_V/\kappa(\lambda)$. The predicted trend is for low-extinction galaxies to have higher Hα-to-far-IR ratios, and conversely for higher-extinction galaxies ($A_V \gtrsim 1$) to have lower Hα-to-far-IR ratios, due to the increased absorption of Hα photons in the dustier regions which give rise to far-IR emission.

[13] The latter corrected by a factor 1.6× to account for the change from a Salpeter (1955) to a Chabrier (2003) IMF.

We see from Figure 2 that our four (extinction-corrected) stacked HiZELS samples occupy a region of moderate $A_V \sim 1$–1.5. The uncorrected Hα-selected stacks, as previously noted, have lower A_V than the corrected stacks. The likely explanation for this is simply that only those galaxies with low A_V are likely to exceed $0.2L^\star_{H\alpha}(z)$ if no correction is applied, and that it is the inclusion of intrinsically luminous (but heavily extinguished) galaxies—after applying a dust correction—that raises the median A_V in the corrected subsamples. We also note that the difference in A_V between the corrected/uncorrected subsamples at $z = 0.40$ is greater than the difference between any other two subsamples at a given redshift. This is most likely because at $z = 0.40$ the volume probed by our Hα observations is smaller than at any other redshift, meaning fewer galaxies exceed $0.2L^\star_{H\alpha}$ than at any other redshift—hence the up-scattering of a small number of additional, heavily extinguished galaxies above $0.2L^\star_{H\alpha}$ by performing a full dust correction produces larger random shifts in the derived properties of the $z = 0.40$ subsample than at other redshifts, where the larger survey volumes mitigate the effect of up-scattering a small number of heavily extinguished galaxies.

In Figure 2, we also show $\log_{10}[L_{H\alpha}/L_{IR}]$ as a function of L_{IR}, along with comparison samples from the literature. We see no evidence of strong trends in $L_{H\alpha}/L_{IR}$ as a function of L_{IR}, in either the corrected or uncorrected HiZELS subsamples, with $\log_{10}[L_{H\alpha}/L_{IR}]$ lying within $\pm 1\sigma$ of the median at each redshift, regardless of whether we apply individual $A_{H\alpha}$ corrections or not. Hence, although we cannot claim to observe *strong* evolution in $L_{H\alpha}/L_{IR}$ with L_{IR}—which we would expect, if higher-L_{IR} galaxies were found to be dustier on average—the relationship between $L_{H\alpha}/L_{IR}$ and L_{IR} is at least *weakly* consistent with our expectations. These results suggest that, while our extinction-corrected Hα-selected stacks do contain some IR-luminous, dusty galaxies, they are not dominated by extreme sources.

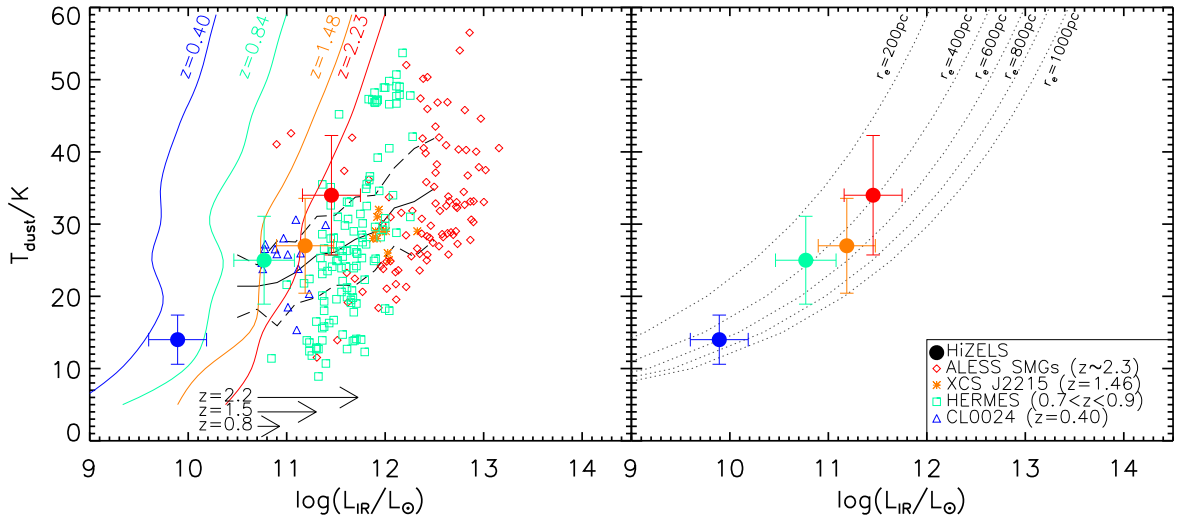

Figure 3. Left: dust temperature, $T_{\rm dust}$, vs. the modified blackbody luminosity, $L_{\rm IR}$, for the four extinction-corrected, Hα-selected HiZELS samples, using properties measured from the best-fitting SEDs (Section 2.5). HiZELS stacks are plotted with large filled symbols, and use the same color coding as in Figure 1. We show the relation for $0.1 < z < 2$ IR-luminous star-forming galaxies (black line with $\pm 2\sigma$ scatter) from Symeonidis et al. (2013). Typical selection boundaries are shown with color coded lines for each redshift (see text for details). The arrows at the bottom of the plot indicate the expected increase in $L^{\star}_{\rm IR}(z)$ relative to $z = 0.40$, given the evolution in $L^{\star}_{\rm H\alpha}(z)$, and the common scaling between $L_{\rm H\alpha}$ and $L_{\rm IR}$ proposed in Kennicutt & Evans (2012). This evolution roughly matches that seen in our samples, in the sense that straightforwardly converting $L^{\star}_{\rm H\alpha}(z)$ to SFR$^{\star}(z)$ to $L^{\star}_{\rm IR}(z)$ using the scaling relations of Kennicutt & Evans (2012) places the arrowheads at infrared luminosities that are close to those we measure from the stacks. We also show comparison samples observed in the far-IR at the redshifts of our HiZELS stacks, including a population of ALMA-detected $z > 2$ submillimeter galaxies (Swinbank et al. 2014), submillimeter-detected starburst galaxies in the $z = 1.46$ cluster, XCS J2215 (Ma et al. 2015), a population of $\langle z \rangle \sim 0.8$ far-infrared selected starbursts from Casey et al. (2012), and starburst galaxies in the core of the $z = 0.4$ cluster CL0024 + 17 (Johnson et al. 2016). Right: as the left plot, but with the comparison samples and the local relation removed. If the dust is well described by a graybody SED, then it ought to obey a modified Stefan–Boltzmann (SB) law, $L_{\rm IR} \propto R^2 T^4_{\rm dust}$. We show curves of constant size, using the prescription outlined in Section 3.4, finding that the dust emission from our Hα stacks is well described as having a constant characteristic size $\sim$500 pc, with increasing temperature and luminosity at higher redshifts. We note that this size is $\sim$10$\times$ smaller than the median stellar sizes of HiZELS galaxies, which may suggest the dust is bound in clumpier knots and filaments, and does not directly trace the stellar emission.

3.3. Luminosity–Temperature Relation

In Figure 3, we show the measured infrared luminosities and dust temperatures for each of the four HiZELS stacks. We see that both $L_{\rm IR}$ and $T_{\rm dust}$ appear to increase with redshift in our sample from $z = 0.40$ to 2.23. Due to the evolving Hα luminosity threshold—and the fact that both $L_{\rm H\alpha}$ and $L_{\rm IR}$ correlate with the SFR—we expect $L_{\rm IR}$ of our stacked Hα-selected galaxies to increase with redshift,. This behavior is indeed seen, with $L_{\rm IR}$ increasing from $10^{9.9\pm0.3}\,L_{\odot}$ to $10^{11.5\pm0.3}\,L_{\odot}$ between $z = 0.40$ and 2.23—an increase of a factor $\sim$40$\times$, compared to the increase of a factor $\sim$10$\times$ in $L^{\star}_{\rm H\alpha}(z)$.

We determine the effective selection boundaries for our HiZELS samples on the plot of $T_{\rm dust}$ against $L_{\rm IR}$ by generating 1000 isothermal graybody SEDs at each redshift (with fixed $\beta = 1.5$), in increments of $T_{\rm dust} = 5$ K between 5 and 60 K. At each redshift, and at each temperature, we find the lowest-$L_{\rm IR}$ graybody SED that is above the measured 3σ detection limits of at least two of the SPIRE band stacked images. Our HiZELS stacked subsamples are typically 3$\times$ more luminous than the selection limits at the corresponding temperature and redshift.

For comparison, we plot the $L_{\rm IR}$–$T_{\rm dust}$ relation from Symeonidis et al. (2013), defined for $0.1 < z < 2$ LIRG/ULIRG galaxies in the *Herschel* Multi-Tiered Extragalactic Survey (HerMES; Oliver et al. 2010) and PACS Evolutionary Probe (PEP; Lutz et al. 2011) surveys. In addition, we also populate Figure 3 with data from comparison samples at similar redshifts to the HiZELS slices taken from the literature. These include the $z = 0.40$ galaxy cluster Cl 0024 + 16 (Johnson et al. 2016), $\langle z \rangle = 0.8$ Herschel SPIRE detected SFGs from Casey et al. (2012), Ultra-Luminous InfraRed Galaxy (ULIRG) starbursts in the $z = 1.46$ cluster XCS J2215 (Ma et al. 2015), and also $z = 2$–3 ALMA-detected SMGs in the Extended *Chandra* Deep Field South (ALESS; Swinbank et al. 2014).

We see that our Hα-selected stacks are typically a little warmer (at fixed $L_{\rm IR}$) than the Symeonidis et al. (2013) local relation, but similar in temperature to the comparison samples (which comprise individually far-IR detected galaxies) at each redshift. The ability of our stacking approach to extend the $L_{\rm IR}/T_{\rm dust}$ relationship to lower luminosities compared with the individual detections in the comparison samples is also apparent; at $z = 0.40$ and 0.84, our Hα-selected stacks are both lower in luminosity and (a little) cooler than the individually detected galaxies in their respective comparison samples. At $z = 1.47$ and 2.23, our stacked Hα-selected samples are roughly an order of magnitude fainter than the individually detected galaxies in their respective comparison samples, but have similar dust temperatures.

3.4. Dust Sizes

To interpret the locations of the HiZELS stacks on the $L_{\rm IR}$–$T_{\rm dust}$ plot, we employ a simple model that relates the luminosity, temperature, and expected size of the far-IR emitting regions in our sample. For a spherical blackbody source, the Stefan–Boltzmann law ($L = 4\pi R^2 \sigma T^4$, where σ is the Stefan–Boltzmann constant and R is the radius of the emitting source) provides the natural framework within which to interpret the $L_{\rm IR}$–$T_{\rm dust}$ relation. In the case of graybody emission, a form of the Stefan–Boltzmann law will apply, in which the Stefan–Boltzmann constant, σ, is replaced with $\tilde{\sigma}(T, \beta, \lambda_0)$ for a given set of dust properties. Choosing $\beta = 1.5$ (as we did during the SED fitting), and fixing

$\lambda_0 \sim 100\ \mu$m as the reference wavelength at which the dust opacity is unity, Yan & Ma (2016) investigated this relationship in a sample of high-redshift, dusty star-forming ULIRGs, finding that, for dust temperatures between 10 and 100 K, $\tilde{\sigma}$ can be approximated as:

$$\frac{\tilde{\sigma}(T_{\rm d})}{\sigma} = 10^{-3}(-3.03T_{\rm d}^{1.5} + 45.55T_{\rm d} - 127.53). \quad (12)$$

We can use this to estimate the "effective radii" of the dust-emitting regions of our Hα-selected stacks via their locations on the $L_{\rm IR}$–$T_{\rm dust}$ plane as:

$$R_{\rm eff} = \sqrt{\frac{L_{\rm IR}}{4\pi\tilde{\sigma}T_{\rm d}^4}}. \quad (13)$$

The effective radii for the dust-emitting regions of the Hα-selected galaxies in our stacks are between 0.4 and 0.6 kpc. We can compare these to the stellar sizes from Stott et al. (2013), who measured the typical sizes of Hα-selected galaxies by fitting Sérsic profiles to their rest-frame optical continuum emission. The dust sizes we derive from our analysis are, at all redshifts, nearly an order of magnitude smaller than the stellar continuum sizes, $\langle r_e \rangle = 3.6 \pm 0.3$ kpc. Such a discrepancy may indicate that the bulk of the dust in Hα-selected star-forming galaxies is not distributed smoothly throughout the ISM, but is instead bound up in (one or more) dense clumps. Direct observational evidence for the existence of dense, star-forming clumps within the ISMs of high-redshift galaxies has traditionally been limited to extreme starbursting systems (e.g., Swinbank et al. 2010; Hodge et al. 2012); however, recent work by Zanella et al. (2015), based on spatially resolved maps of [O II], [O III] and Hβ line emission in a galaxy cluster at $z = 1.99$, suggests that similar structures may exist in more modestly star-forming systems as well. We note that a putative dust size of ∼500 pc corresponds to an angular scale ∼0.″06 at $z = 2.23$, which is far below the angular resolution limit of *Herschel*, and for a galaxy at $\log_{10}[L_{\rm IR}/L_\odot] \sim 11$, would require a significant investment of long-baseline ALMA time to measure directly.

4. The Evolving Gas Content of the Universe

In order to understand galaxy evolution, it is necessary to link populations of galaxies at high and low redshift. One means by which high-redshift galaxies can be linked to local populations is by determining the typical timescale on which a galaxy at a given redshift forms stars, and exhausts its gas supply. The length of time a star-forming galaxy can support its present rate of star formation is determined by its gas depletion timescale ($\tau_{\rm dep}$), defined—in the absence of inflows/outflows of material—as the ratio of the available molecular gas supply ($M_{\rm H_2}$) to the rate at which that gas is being converted into stars (i.e., the SFR).

As noted previously, direct observations of the atomic and molecular gas content of distant galaxies are expensive in terms of telescope time, making them unfeasible for large samples. However, we can obtain indirect constraints on the total gas masses of our Hα-selected galaxies by invoking the common assumption that the total gas mass ($M_{\rm gas} \equiv M_{\rm H\,I} + M_{\rm H_2}$) is proportionally linked to $M_{\rm dust}$ (which we have measured from our far-IR stacks) via a gas-to-dust ratio, $\delta_{\rm GDR}$.

In the local universe, star-forming galaxies of approximately solar metallicity, $Z \equiv 12 + \log({\rm O/H}) = 8.7$ (Asplund et al. 2009), have $\delta_{\rm GDR} \sim 140$ (Draine et al. 2007). More recent work by Leroy et al. (2011), across a more diverse sample of local galaxies, has found that this ratio is sensitive to (and scales inversely with) the metallicity, as

$$\log_{10}\delta_{\rm GDR} = (9.4 \pm 1.1) - (0.85 \pm 0.13)[12 + \log_{10}({\rm O/H})]. \quad (14)$$

In order to constrain the gas-to-dust ratio for our Hα-selected samples, it is therefore necessary to first estimate their metallicities. Metallicities of extragalactic sources are commonly estimated via emission line diagnostics, e.g., the [O III]/[N II] ratio (Pettini & Pagel 2004), however obtaining such line diagnostics for a statistically large sample is relatively expensive in terms of telescope time. Stott et al. (2013) carried out multi-object spectrometer observations of the [O III] and [N II] lines for a subset of 381 HiZELS galaxies, in order to calibrate the "FMR," a 3D surface relating a galaxy's metallicity with its (much more easily observable) stellar mass and SFR.

Schreiber et al. (2016) perform a similar analysis on a sample of $z = 0.7$–1.3 star-forming galaxies, selected from the CANDELS field using the FMR calibration of Kewley & Ellison (2008), in which they measure $\delta_{\rm GDR} = 150$–380. The implied $\log_{10}[M_{\rm gas}/M_\odot] = 9$–$10$ for their stellar mass-selected sample galaxies is in good (within ∼30%) agreement with the gas masses they measured directly via ^{12}CO +H I spectroscopy. While a direct comparison of the results of the their mass-selected sample with those of our SFR-selected sample is difficult, the broad conclusion from Schreiber et al. (2016)—that in spite of the potential uncertainties in inferring $M_{\rm gas}$ from $M_{\rm dust}$ via a metallicity-dependent $\delta_{\rm GDR}$, the derived gas masses correlate well with those obtained from spectroscopy—is likely to hold for our Hα-selected sample as well.

4.1. The Metallicities of Hα-selected Galaxies

We measure the metallicities of our stacked Hα-selected samples by exploiting the FMR—a 3D plane defined by Mannucci et al. (2010) as

$$12 + \log({\rm O/H}) = a_0 + a_1 m + a_2 s + a_3 m^2 + a_4 ms + a_5 s^2 \quad (15)$$

where $m = \log_{10}[M_\star/M_\odot] - 10$, and $s = \log_{10}[{\rm SFR}/M_\odot\ {\rm yr}^{-1}]$. Recently, Stott et al. (2013), used Subaru FMOS observations of the [N II]-to-Hα line ratio to investigate this metallicity relationship in a subsample of the HiZELS galaxies, comprising 381 bright Hα-selected galaxies at $z = 0.84$ and 1.47. They found that the bright Hα-selected galaxies occupy an FMR that is relatively flat across the mass range of their sample, and depends primarily on SFR, with best-fit coefficients $a_0 = 8.77$, $a_1 = 0.00$, $a_2 = -0.055$, $a_3 = 0.00$, $a_4 = 0.019$, and $a_5 = -0.101$.

We measure $\langle {\rm SFR_{H\alpha}} \rangle$ for our stacked subsamples from the median dust-corrected Hα luminosities (Table 1), and use the Hα-weighted stellar masses, $\langle M_\star \rangle$, of each stacked subsample (measured in Section 2.7) to derive metallicities of $12 + \log({\rm O/H}) = 8.75 \pm 1.60$ at $z = 0.40$, 8.65 ± 0.71 at $z = 0.84$, 8.49 ± 0.71 at $z = 1.47$, and 8.46 ± 0.65 at $z = 2.23$.

Given these metallicities, the Leroy et al. (2011) relation (Equation (14)) implies gas-to-dust ratios $\delta_{\rm GDR} = 90 \pm 30$, 110 ± 20, 150 ± 20, and 160 ± 20 for each of the four

Table 4
Gas Properties of HiZELS Galaxies

	$12 + \log(\mathrm{O/H})$[a]	δ_{GDR}	$M_{\mathrm{H_2}}$ ($\times 10^9\,M_\odot$)	τ_{dep} (Gyr)	τ_{H} (Gyr)	$\rho_{\mathrm{H_2}}/\rho_C$ (%)	$\rho_{\mathrm{H_2}}/\rho_{\mathrm{H_2,\,TOT}}$ (%)	f_{gas}[b]
$z = 0.40$	8.7 ± 1.6	90 ± 30	10.1 ± 5.3	5 ± 3	9.4	0.017 ± 0.016	50 ± 20	0.5 ± 0.3
$z = 0.84$	8.7 ± 0.7	110 ± 20	3.5 ± 2.0	0.5 ± 0.3	6.6	0.012 ± 0.007	20 ± 10	0.7 ± 0.4
$z = 1.47$	8.5 ± 0.7	150 ± 20	8.3 ± 2.8	0.3 ± 0.2	4.4	0.015 ± 0.008	20 ± 10	0.7 ± 0.3
$z = 2.23$	8.5 ± 0.6	160 ± 20	4.6 ± 1.9	0.1 ± 0.1	3.0	0.019 ± 0.007	30 ± 10	0.5 ± 0.3

Notes.
[a] We calculate $12 + \log(\mathrm{O/H})$ from $\langle \mathrm{SFR_{H}}\alpha \rangle$ and $\langle M_\star \rangle$ using the fundamental metallicity relation (FMR) of Stott et al. (2013), and use these metallicities to calclate δ_{GDR} using the relation of Leroy et al. (2011). We note that the Leroy et al. (2011) relation is sensitive to the metallicity, which in turn renders the results in this table sensitive to the choice of FMR (see Appendix A.2 for details).
[b] f_{gas} includes both atomic and molecular components.

HiZELS stacks, the last three of which are consistent with the typical Milky Way value $\delta_{\mathrm{GDR}} = 140$ (Draine et al. 2007). In Appendix A.2 we discuss the sensitivity of the gas masses estimated in this manner to the coefficients used to fit the FMR, comparing the results we derive using the Stott et al. (2013) fit for HiZELS galaxies to similar fits performed in [O II] and *r*-band selected samples in Sloan Digital Sky Survey (SDSS) by Mannucci et al. (2010) and Lara-López et al. (2010).

4.2. The Gas Masses of HiZELS Star-forming Galaxies

Using the Stott et al. (2013) FMR in conjunction with the metallicity-dependent model of Leroy et al. (2011), we estimate gas-to-dust ratios $\delta_{\mathrm{GDR}} \sim 90$–$160$ for our Hα-selected subsamples. Based on the atomic-to-molecular gas mass relation $\log_{10}(M_{\mathrm{H_2}}) = 0.99 \times \log_{10}(M_{\mathrm{H\,I}}) - 0.42$, measured in a sample of local massive galaxies by Saintonge et al. (2011), we infer a typical molecular-to-total gas mass ratio of $\sim$0.24. The molecular gas masses implied by this analysis are summarized in Table 4, and lie in the range $(3.5$–$10.1) \times 10^9\,M_\odot$; hence the gas depletion timescales (assuming these galaxies maintain their current SFR) are $\tau_{\mathrm{dep}} = 5 \pm 3$, 0.5 ± 0.4, 0.3 ± 0.2, and 0.1 ± 0.1 Gyr for the galaxy samples at $z = 0.40$, 0.84, 1.47, and 2.23, respectively.

We see from Table 4 that these molecular gas depletion timescales are significantly shorter than the Hubble time (τ_{H}) at each redshift (especially at $z \geqslant 0.84$). We perform a least-squares fit to the relation $\tau_{\mathrm{dep}} \sim (1 + z)^N$, finding the data to be fit best by the exponent $N = -3.9 \pm 1.0$. This exponent is steeper than expected in a simple model where the gas depletion timescale is proportional to the dynamical timescale ($N = -1.5$; Davé et al. 2011).

We measure total gas fractions ($f_{\mathrm{gas}} \equiv (M_\star + M_{\mathrm{gas}})/M_{\mathrm{gas}}$) of our H$\alpha$-selected stacks (including both atomic and molecular gas), finding $f_{\mathrm{gas}} = 50 \pm 30\%$ at $z = 0.40$, $70 \pm 40\%$ at $z = 0.84$, $70 \pm 30\%$ at 1.47, and $50 \pm 30\%$ at $z = 2.23$. These gas fractions are high—and carry significant uncertainties—but are consistent with the results of Tacconi et al. (2013), who measured the gas masses of star-forming galaxies at $z = 1.2$ and 2.2 via their ^{12}CO $J = 3$–2 emission in the PHIBSS survey. If we assume an exponentially declining star formation history (SFH), then this (along with the short depletion timescales at $z > 0.8$) suggests a scenario in which the typical star-forming galaxies near the peak of cosmic star formation activity—which we select with our $>0.2L^\star_{\mathrm{H}\alpha}(z)$ cut—must undergo replenishment of their gas reservoirs (e.g., Tacconi et al. 2013). We note that while Tacconi et al. (2013) find evidence of decreasing gas fractions with increasing stellar mass, the large uncertainties in our gas mass estimates prevent us from identifying such a correlation within our sample.

It is illustrative to compare the contribution that these Hα-selected galaxies make to the gas content of the universe as a function of redshift, with predictions from models. For each of our four subsamples of SFR-selected galaxies, we estimate their contribution to the cosmological gas density (relative to the critical density at $z = 0$, $\rho_{C\,z=0} \equiv 3H_0^2/(8\pi G)$) as

$$\frac{\rho_{\mathrm{H_2}}}{\rho_{C\,z=0}} = \frac{N_{\mathrm{stack}}\,\delta_{\mathrm{GDR}}\,M_{\mathrm{dust}}}{\rho_{C\,z=0}\,V_{\mathrm{H},z}}. \quad (16)$$

We find that the gas reservoirs of these Hα-selected star-forming galaxies account for $\sim$1%–2% of the critical density at each redshift.

In Figure 4, we use our dust-based estimates of the molecular gas mass to measure the contribution made by Hα-selected star-forming galaxies to the total $\mathrm{H_2}$ density of the universe, as a function of redshift. We compare our gas mass densities with samples from the literature, including Kereš et al. (2005), who measured the gas content at $z = 0$ as $\rho_{\mathrm{gas}}/\rho_{C\,z=0} \sim 1\%$, along with intermediate redshift observations of $z = 0.2$–0.8 ULIRGs from Combes et al. (2011, 2013), and $z \sim 1$–2.5 BX/BM and *BzK* star-forming galaxies from Tacconi et al. (2013) and Daddi et al. (2010). We also include the $\mathrm{H_2}$ mass densities of $S_{870\,\mu\mathrm{m}} > 1$ mJy SMGs observed with ALMA from Swinbank et al. (2014).

We compare each of these populations with the latest cosmological hydrodynamical estimate of the evolving total $\mathrm{H_2}$ gas budget of the universe from the Evolution and Assembly of GaLaxies and their Environments (EAGLE; Schaye et al. 2015) simulation. Using the evolving total $\mathrm{H_2}$ density from EAGLE (Lagos et al. 2015), we find that Hα-selected galaxies brighter than $0.2L^\star_{\mathrm{H}\alpha}(z)$ account for $\sim 50 \pm 20\%$, $20 \pm 10\%$, $20 \pm 10\%$, and $30 \pm 10\%$ of the total cosmic gas supply at $z = 0.40$, 0.84, 1.47, and 2.23, respectively (see Table 4). In comparison, the *BzK* star-forming galaxies account for $\sim$10% of the total $\mathrm{H_2}$ density from the EAGLE simulation at their redshifts, and the $\sim$1 mJy SMGs typically less than 5%.

In their work with the EAGLE simulation, Lagos et al. (2015) propose that the molecular hydrogen density, $\rho_{\mathrm{H_2}}$, is dominated at $z \lesssim 2$ by galaxies more massive than $\log_{10}[M_\star/M_\odot] = 9.7$, at $z \gtrsim 2$ by galaxies less massive than $\log_{10}[M_\star/M_\odot] = 9.7$, and at all redshifts, by galaxies with SFR $\gtrsim 10\,M_\odot\,\mathrm{yr}^{-1}$. Given the median $\langle \mathrm{SFR_{H}}\alpha \rangle$ and $\langle M_\star \rangle$ of the galaxies in our subsamples and the uncertainties in our gas mass estimates (which result in

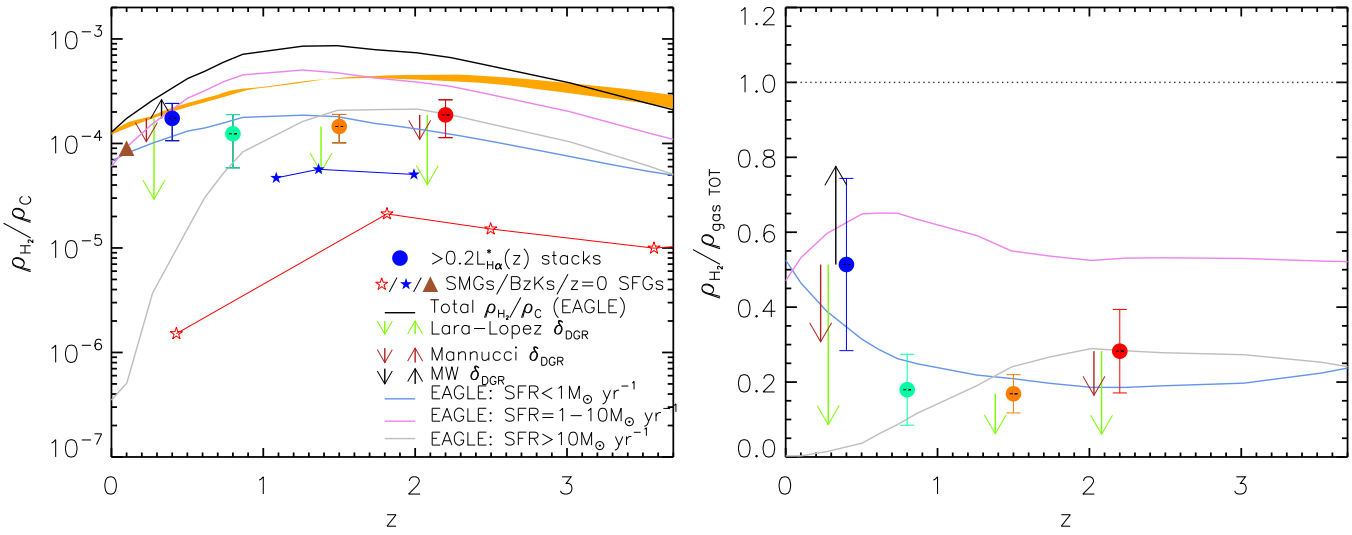

Figure 4. Left: the contribution of far-IR stacked $>0.2\,L^{\star}_{\rm H\alpha}$, star-forming galaxies to the total molecular gas (H_2) density of the universe from the EAGLE simulation (Lagos et al. 2015, thick black line). Stacked Hα-selected galaxies are shown at $z = 0.40$, 0.84, 1.47, and 2.23 with filled, colored points, using the same color scale as Figure 1. Gas masses are measured from the dust continuum fits using the Stott et al. (2013) fundamental metallicity relation (FMR) to measure the metallicity from the SFR and $\langle M_{\star}\rangle$ of each stacked sample, and then using the relation between metallicity and the gas-to-dust ratio ($\delta_{\rm GDR}$) described by Leroy et al. (2011). Colored arrows to the left of each point show the effect of using the FMR fits of Lara-López et al. (2010) and Mannucci et al. (2010) to derive $\delta_{\rm GDR}$, or of using a fixed Milky Way $\delta_{\rm GDR} = 140$ (Draine et al. 2007); we discuss the sensitivity of our results to the choice of FMR relation in detail in Appendix A.2. We do not plot arrows where the change is less than the 1σ error on the HiZELS data point. Also shown are the $z = 0$ data from Kereš et al. (2005), along with higher-redshift *BzKs* and SMGs from Daddi et al. (2010), Tacconi et al. (2013), and Swinbank et al. (2014), respectively, plus model predictions from Lagos et al. (2015) for the contribution of galaxies with given SFRs. We see that our stacked Hα-selected galaxies account for a significant fraction of the gas located in 1–10 $M_{\odot}$ yr^{-1} star-forming galaxies at $z = 0.40$ and 2.23. At $z = 0.84$ and 1.47, a significant fraction of the gas remains unaccounted for; however our analysis suggests that stacking Hα-selected SFGs recovers more of the gas in the universe than is measured by summing individually detected *BzKs*/SMGs over the same redshift range. For comparison with EAGLE, we also show the range of H_2 densities from the models of (Popping et al. 2014, orange band). Right: as in the left panel, but normalized by the total gas density. We see that our stacks all lie below the predicted contribution to $\rho_{\rm H_2}$ from galaxies with SFR = 1–10 $M_{\odot}$ yr^{-1} in EAGLE, indicating that the contribution made by HiZELS Hα-selected galaxies to the total gas density is significant, but that our Hα cuts have also missed a significant contribution to the gas budget from low-SFR galaxies.

the large error bars in Figure 4), our results are broadly consistent with this picture; for $z > 0.84$, as $\langle {\rm SFR_{H\alpha}}\rangle$ and $\langle M_{\star}\rangle$ for our Hα-selected samples increase, so we find our galaxies contribute a (slightly) greater fraction of the total gas supply. This trend is exactly reversed with the $z = 0.40$ measurement; however, this may be the result of cosmic variance, given that the volume probed at this redshift is the smallest of any of the HiZELS redshift slices.

In order to further investigate the contribution made by Hα-selected galaxies to the evolving gas density of the universe, we use estimates of $\rho_{\rm H_2}/\rho_{\rm C\,z=0}$ from the EAGLE simulation, broken down as a function of SFR (C. Lagos 2017, private communication), which allow us to obtain a prediction for the fraction of $\rho_{\rm H_2}/\rho_{\rm C\,z=0}$ held in *all* galaxies whose SFR is higher than the lower limit implied by our $>0.2L^{\star}_{\rm H\alpha}(z)$ cut, regardless of whether those galaxies would be selected as narrow-band excess sources in the HiZELS survey or not. We find that Hα-selected galaxies brighter than $0.2L^{\star}_{\rm H\alpha}$ account for $100 \pm 30\%$ of the gas in galaxies at $z = 0.40$ above our evolving SFR cut, falling to $40 \pm 10\%$ at $z = 0.84$, $20 \pm 10\%$ at $z = 1.47$, and $40 \pm 10\%$ at $z = 2.23$. Hence, we find that flux and equivalent width-limited Hα-selection provides an efficient means of selecting galaxies that host a significant fraction of the molecular gas in the universe.

5. Conclusions

In this paper, we have investigated the evolution of the far-IR properties of Hα-emitting star-forming galaxies across the peak of the star formation activity in the universe. By bringing together 100–500 μm imaging from *Herschel* PACS/SPIRE with 850 μm observations from SCUBA-2 and deep 1.4 GHz VLA imaging, we have employed stacking in order to measure the far-IR properties of "typical" Hα-selected star-forming galaxies below the confusion limit, spanning the knee of the (evolving) Hα luminosity function ($>0.2L^{\star}_{\rm H\alpha}(z)$).

1. We have measured the infrared luminosities, $L_{\rm IR}$, dust masses, $M_{\rm dust}$, and temperatures, $T_{\rm dust}$, of our stacked samples, finding them to evolve from cold ($\sim$14 K) systems, with Milky Way-like luminosities ($L_{\rm IR} \sim 10^{10}\,L_{\odot}$) at $z = 0.40$, to warmer, LIRG-like systems ($T_{\rm dust} \sim 34$ K, $L_{\rm IR} \sim 10^{11}\,L_{\odot}$) systems at $z = 2.23$. The evolution in their far-IR properties is comparable to the evolution in $L^{\star}_{\rm H\alpha}(z)$ used to construct our samples.
2. Comparing the infrared and Hα-derived SFRs of those galaxies contributing to the stacks allows us to obtain new constraints on the levels of Hα extinction due to dust. We see no evidence of strong trends in the ratio $\log_{10}[L_{\rm H\alpha}/L_{\rm IR}]$ with either redshift, $L_{\rm IR}$, or stellar continuum extinction, A_V, suggesting that our selection based on $L_{\rm H\alpha}$ relative to an evolving luminosity cut has selected galaxies with "similar" ISM properties at all redshifts.
3. Our stacking method allows us to study the relationship between the far-IR luminosities and dust temperatures of "normal" star-forming galaxies that are otherwise below the confusion limit in the *Herschel* SPIRE bands. Comparing the results of our stacking analysis with samples from the literature, we find our Hα-selected galaxies follow a similar trend in terms of $L_{\rm IR}$–$T_{\rm dust}$ to that seen in local galaxies, but are typically warmer (at a given $L_{\rm IR}$) than comparison samples at the same redshift.

Using a modified Stefan–Boltzmann law, we estimate the characteristic sizes of the dust-emitting regions of HiZELS galaxies, finding them to be ∼0.5 kpc, nearly an order of magnitude smaller than their stellar sizes and providing tentative evidence of their having a clumpy ISM.

4. Using a two-step approach that entails estimating the dust metallicity from a "FMR," and then calculating a metallicity-dependent gas-to-dust ratio, we use the far-IR photometry for our stacked Hα-selected samples to estimate their typical gas masses. By comparing these with the SFRs, we estimate gas depletion timescales $\tau_{\rm dep} \sim 0.1$–5 Gyr across $z = 2.2$–0.40. These short gas depletion timescales (relative to the Hubble time at each redshift) suggest that refueling of these gas reservoirs—likely either by steady accretion or by an accumulation of minor mergers—may have taken place at some point in the past. The total gas mass fractions are high —$f_{\rm g} \backsimeq 0.6 \pm 0.1$—at all redshifts. However, the choice of FMR used to measure $\delta_{\rm GDR}$ is a potentially significant systematic, and accounts for an additional factor ∼2× uncertainty on the gas masses at all redshifts (Appendix A.2).
5. Combined with the number densities of Hα-selected galaxies at $z = 0.40$, 0.84, 1.47, and 2.23, this allows us to estimate the contribution of $>0.2L_{\rm H\alpha}(z)$, Hα-selected galaxies to the evolving H_2 content of the universe. We find that galaxies satisfying our selection criteria comprise a significant fraction of the total H_2 in the universe ($35 \pm 10\%$ on average), as predicted by the state-of-the-art cosmological hydrodynamical simulation EAGLE, modulo the aforementioned uncertainties due to the FMR.

Our results are important in order to understand the nature and evolution of luminous Hα emitters within the context of the evolving properties of the star-forming galaxy population. As the typical SFRs of galaxies increase from the local universe toward the peak of star formation at $z = 1$–2, so too does the relative prevalence of dusty systems, observable in the far-IR, leading to increasing biases in surveys that rely exclusively on UV/optical tracers at higher redshifts.

A.P.T., I.R.S., and A.M.S. acknowledge support from STFC (ST/L00075/X). A.P.T. and I.R.S. also acknowledge support from the ERC Advanced Investigator Programme DUSTYGAL (#321334); I.R.S. acknowledges a Royal Society/Wolfson merit award. D.S. acknowledges financial support from the Netherlands Organisation for Scientific Research (NWO) through a Veni Fellowship. D.S. also acknowledges funding from FCT through a FCT Investigator Starting Grant and Start-up Grant (IF/01154/2012/CP0189/CT0010) and from FCT grant PEst-OE/FIS/UI2751/2014. A.P.T. thanks John Stott and Claudia Lagos for sending insightful comments on the fundamental metallicity relation, and sharing tables of data from EAGLE, respectively. The SCUBA-2 850 μm data presented in this paper were taken as part of Program ID MJLSC02. It is a pleasure to thank the entire staff of the JCMT for their superb support throughout the S2CLS campaign. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. This research has made use of data from HerMES project (http://hermes.sussex.ac.uk/). HerMES is a *Herschel* Key Programme utilizing Guaranteed Time from the SPIRE instrument team, ESAC scientists and a mission scientist. *Herschel*-ATLAS is a project with *Herschel*, which is an ESA Space Observatory with science instruments provided by European-led Principal Investigator consortia and with important participation from NASA. The H-ATLAS Web site is www.h-atlas.org. The authors wish to thank all staff at the (former) JAC and their successors at EAO for their help in conducting the observations at the UKIRT telescope and their ongoing support.

Appendix

A.1. The Effect of Clustering on Stacked Flux Densities

We investigate possible biases in our stacking analysis due to the inherent clustering of galaxies within the large-beam SPIRE images by performing a set of simulations, using model images of the COSMOS field at 250, 350, and 500 μm.

We begin with a blank grid, using the same astrometry and pixel scale as for the real SPIRE images, and add delta functions at the positions of each of the ∼47,000 24 μm/radio-detected galaxies in the prior catalog, whose flux densities in each of the model images are set to the flux densities measured when the raw SPIRE images were originally deblended. Next, we seed the maps with delta functions at the positions of the 1771 HiZELS galaxies in COSMOS (applying a 0.″5 exclusion radius around each galaxy to exclude those HiZELS galaxies that were also priors, in order to avoid injecting the same galaxies twice). The SPIRE flux densities of these model HiZELS galaxies in the model SPIRE images are set to those of a 30 K graybody SED (the weighted average temperature of the real HiZELS SEDs) at the appropriate redshifts, scaled by each galaxy's measured $L_{\rm H\alpha}$ under the condition that $\rm SFR_{IR} = SFR_{H\alpha}$. We convolve each grid of delta functions with the appropriate SPIRE PSF, before addding Gaussian random noise from a distribution matching the rms of the original residual images (i.e., 1.6 mJy, 2.1 mJy, and 3.1 mJy at 250, 350, and 500 μm, respectively). We then deblend these model SPIRE images using the same technique as was used for the real maps, requiring the code to fit sources in each of the bands at the positions of every galaxy injected into the image (i.e., 24 μm/radio priors *and* model HiZELS galaxies). This generates a new list of deblended photometry, along with a new residual image at each wavelength. We perform the same stacking analysis as outlined in Section 2.3 by reinjecting the model HiZELS sources into the model residual images and stacking those galaxies that lie above the evolving cut, $L_{\rm H\alpha} \geqslant 0.2L^{\star}_{\rm H\alpha}$ (these simulations are called "clustered HiZELS stacks").

We then repeat this entire process, generating new model maps by injecting scaled delta functions at the positions of the 24 μm/radio-detected galaxies; however, this time, we inject the model HiZELS graybodies at random positions in the field, and run the deblending code again. By randomizing the positions of the model HiZELS galaxies, we erase their clustering signature. We again stack the $L_{\rm H\alpha} \geqslant 0.2L^{\star}_{\rm H\alpha}$ model HiZELS galaxies by reinjecting sources into the residual images at these randomized positions, and measure their 250, 350, and 500 μm flux densities ("unclustered HiZELS stacks").

By comparing the flux densities measured from these two sets of simulated SPIRE stacks, it is possible to directly

quantify the effect that HiZELS galaxies clustering with each other has on their measured stacked flux densities. We find that the differences between the flux densities of our "clustered" and "unclustered" model HiZELS galaxies are (0.05 ± 0.03) mJy at 250 μm, (-0.02 ± 0.04) mJy at 350 μm, and (0.07 ± 0.09) mJy at 500 μm. We therefore find no evidence that the clustering of HiZELS galaxies (with each other) has significantly boosted the flux densities of our stacks (presented in Table 2).

A final source of potential bias is the clustering of (non-Hα detected) field galaxies around the HiZELS galaxies, i.e., galaxies that contribute flux in the real SPIRE images but are not included in either the deblending prior list or our target list. Since the deblending code is required to minimize the residuals (i.e., |data − model|), and since the typical source density of optically selected galaxies per SPIRE beam is high, the effect of having a galaxy population in the SPIRE images that is not in the prior list would be that their flux would be erroneously "pulled in" and assigned to the positions of the priors. If those prior galaxies were then reinjected (with their best-fit fluxes) into the residual images (following the process used for the real data) and stacked, their stacked flux densities would be higher than their "true" flux densities. We attempt to quantify this effect by repeating the deblending process yet again, generating a new set of model SPIRE images comprising: (i) the 47,000 24 μm/radio priors; (ii) the 1771 model HiZELS galaxies; and (iii) 295,000 field galaxies at the positions of *I*-band selected galaxies in COSMOS (Ilbert et al. 2008, again with a 0.″5 exclusion radius around each prior/HiZELS galaxy to avoid injecting the same galaxies multiple times. This step excludes ∼45,000 duplicates from the *I*-band catalog). We inject these field galaxies into the SPIRE images with fluxes set to those of graybody SEDs with $T_{\rm dust}$ drawn randomly in the range 25–50 K, and shifted to the photometric redshifts estimated by Ilbert et al. (2008). Each of these field galaxy SEDs is scaled in luminosity to an SFR drawn randomly between SFR = 0.1–10 $M_{\odot}\,{\rm yr}^{-1}$. The median flux densities of these field galaxies are $S_{250} = (0.756 \pm 0.012)$ mJy, $S_{350} = (0.405 \pm 0.005)$ mJy, and $S_{500} = (0.175 \pm 0.002)$ mJy. We deblend these new model images *using the same prior list as before*, and once again stack at the positions of the $>0.2L^{\star}_{\rm H\alpha}(z)$ galaxies (these are termed "stacks with field galaxies"). By comparing the flux densities measured in the "with field galaxies" and "clustered HiZELS" stacks, we are able to quantify the degree to which the fluxes of our HiZELS stacks may have been boosted due to the presence of a population of clustered field galaxies, which is not accounted for in the deblending.

We find offsets in the flux densities of our "with field galaxies" and "clustered HiZELS" stacks to be $\Delta S_{250} = -0.5 \pm 0.2$ mJy, $\Delta S_{350} = 0.2 \pm 0.3$ mJy, $\Delta S_{500} = 0.2 \pm 0.5$ mJy. These offsets are comparable to the statistical uncertainties in the stacked flux densities from the real images at each wavelength and are not systematically dependent on wavelength/beam size, therefore we do not apply a correction to the flux densities measured from stacking in the real SPIRE data. Instead, we incorporate this effect into our results by combining the size of the offsets in flux densities between our simulations and the statistical uncertainties in our stacks in quadrature (Table 2). From this series of tests, we conclude that the contribution to the flux densities in the SPIRE images from field galaxies *not* in our prior catalog is low (at least in the regions around the SFR-selected HiZELS galaxies). We note that the magnitude of this effect is likely to be a function of the flux distribution of the field galaxy population, which is the very unknown that stacking analyses intend to address. In our simulations, galaxies are injected into the model SPIRE images with flux densities that are appropriate given the *reasonable*, physically motivated assumptions we have made for their observed-frame colors and SFR. In a forthcoming paper, R. Cochrane (2017, private communication) will show that HiZELS galaxies are typically star-forming centrals located in relatively low mass haloes ($M_{\rm halo} \sim 10^{12}$–$10^{13}\,M_{\odot}$), which may provide a physical explanation for our finding that contamination in the far-IR from galaxies *other* than those we have already accounted for in the deblending process is low.

A.2. Sensitivity of Metallicities to the SFR and Stellar Mass

In Section 4, we measured the gas masses of our HiZELS galaxies using a two-step process in which we used dust masses measured from our far-IR SED fits, and converted these to total (H I+H_2) gas masses using a metallicity-dependent gas-to-dust ratio, $\delta_{\rm GDR}$. Because we do not have emission line diagnostics to constrain the metallicty for each of our HiZELS galaxies, we exploited the mass–metallicity–SFR relation (the so-called "FMR"; Equation (15)) to infer the typical metallicities of our stacked subsamples from their broadband photometric properties. Our analysis utilized the parameterization of the FMR carried out by Stott et al. (2013), who measured its shape in a subset of the Hα-selected galaxies present in this sample, finding the metallicities of HiZELS galaxies to be a strong function of the SFR, and with only a weak dependence on the product of the stellar mass and SFR entering via the a_4 coefficient.

A number of works have sought to characterize the FMR, using photometry and spectral line data drawn from different samples of galaxies. Notably, Mannucci et al. (2010) studied the FMR in a sample of $z \sim 0.1$, [O II]-selected star-forming galaxies from the SDSS, finding it to be well-characterized by the coefficients $a_0 = 8.90$, $a_1 = 0.37$, $a_2 = -0.14$, $a_3 = -0.19$, $a_4 = 0.12$, and $a_5 = -0.054$, while in another work, Lara-López et al. (2010) found the surface of the FMR for $z \sim 0.1$, *r*-band selected galaxies in SDSS-DR7 to be fit by $\log(M_{\star}) = 1.122 \times [12 + \log({\rm O/H})] + 0.474 \times \log({\rm SFR}) - 0.097$.

If we use the Mannucci et al. (2010) and Lara-López et al. (2010) measurements of the FMR in lieu of the Stott et al. (2013) FMR, then the metallicities of our samples are respectively raised to $[12 + \log({\rm O/H}] = 9.01 \pm 0.75$ (9.66 ± 1.71) at $z = 0.40$, 8.72 ± 0.48 (8.99 ± 1.91) at $z = 0.84$, 8.63 ± 0.54 (9.01 ± 3.22) at $z = 1.47$, and 8.73 ± 0.53 (9.24 ± 3.32) at $z = 2.23$. Following Equation (14), these metallicities imply gas-to-dust ratios $\delta_{\rm GDR} = 55 \pm 9$ (15 ± 5), 96 ± 10 (57 ± 23), 114 ± 14 $(55\pm)$, and 94 ± 11 (34 ± 24), at the four HiZELS redshifts, respectively.

We see that the higher metallicities implied by the Lara-López et al. (2010) FMR fit act to lower $\delta_{\rm GDR}$ considerably (∼3–6×) compared to the Stott et al. (2013) fit, whereas the Mannucci et al. (2010) FMR and Stott et al. (2013) fits typically agree with each other to within a factor ∼2×. While a quantitative analysis of the FMR is beyond the scope of this paper, it is important to bear in mind the implications these three measurements have for our gas mass measurements in Section 4. To begin to understand the reasons why these three measurements of the FMR differ, it is important to understand the compositions of the samples in which they were measured.

The galaxies we stack in our subsamples are Hα-selected, with $>0.2L^{\star}_{\mathrm{H}\alpha}(z)$, and typical $\langle \mathrm{SFR} \rangle = 20\,M_{\odot}\,\mathrm{yr}^{-1}$ and $\langle \log M_{\star} \rangle = 10.5$ (see Table 1). The analysis of Stott et al. (2013) is based on a (different) subset of galaxies drawn from the same (HiZELS) parent sample with similar mass ($\langle \log M_{\star} \rangle = 9.5$), albeit with slightly lower typical SFR ($\langle \mathrm{SFR} \rangle = 11\,M_{\odot}\,\mathrm{yr}^{-1}$). The works of Mannucci et al. (2010) and Lara-López et al. (2010) are both based on significantly less active galaxies ($\langle \mathrm{SFR} \rangle = 1.5\,M_{\odot}\,\mathrm{yr}^{-1}$) at lower redshift, albeit at a similar stellar mass ($\langle \log M_{\star} \rangle = 10.1$).

Recent work on the nature of the FMR by Telford et al. (2016), using updated line ratio diagnostics, finds a generally weaker anti-correlation between metallicity and SFR (at a given $M_{\star}$) than that reported in Lara-López et al. (2010), which may reconcile some of the $\sim$3–6$\times$ discrepancy between these δ_{GDR} estimates. However, this anti-correlation is thought to be stronger in galaxies whose current or recent SFRs are higher than their past average SFRs. Without knowing the detailed star formation histories of our Hα-selected galaxies, or having the line ratio diagnostics necessary to directly measure the FMR in our Hα-selected subsamples, we have little choice but to adopt an FMR; we chose to adopt the Stott et al. (2013) FMR, as it is calculated in the region of SFR–$M_{\star}$ parameter space that most closely matches that of our $>0.2L^{\star}_{\mathrm{H}\alpha}$ subsamples, but we reiterate that the gas-to-dust ratios (and hence gas masses) we subsequently derive are sensitive to this choice of FMR.

References

Abramson, L. E., Kelson, D. D., Dressler, A., et al. 2014, ApJL, 785, L36
Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481
Barger, A. J., Cowie, L. L., Sanders, D. B., et al. 1998, Natur, 394, 248
Bothwell, M. S., Smail, I., Chapman, S. C., et al. 2013, MNRAS, 429, 3047
Bower, R. G., Benson, A. J., Malbon, R., et al. 2006, MNRAS, 370, 645
Calhau, J., Sobral, D., Stroe, A., et al. 2017, MNRAS, 464, 303
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Calzetti, D., Kinney, A. L., & Storchi-Bergmann, T. 1994, ApJ, 429, 582
Carilli, C. L., & Walter, F. 2013, ARA&A, 51, 105
Casey, C. M., Berta, S., Béthermin, M., et al. 2012, ApJ, 761, 140
Casey, C. M., Chapman, S. C., Neri, R., et al. 2011, MNRAS, 415, 2723
Casey, C. M., Narayanan, D., & Cooray, A. 2014, PhR, 541, 45
Chabrier, G. 2003, PASP, 115, 763
Chapin, E. L., Chapman, S. C., Coppin, K. E., et al. 2011, MNRAS, 411, 505
Chary, R., & Elbaz, D. 2001, ApJ, 556, 562
Chen, C.-C., Smail, I., Ivison, R. J., et al. 2016, ApJ, 820, 82
Civano, F., Marchesi, S., Comastri, A., et al. 2016, ApJ, 819, 62
Clark, C. J. R., Schofield, S. P., Gomez, H. L., & Davies, J. I. 2016, MNRAS, 459, 1646
Combes, F., García-Burillo, S., Braine, J., et al. 2011, A&A, 528, A124
Combes, F., García-Burillo, S., Braine, J., et al. 2013, A&A, 550, A41
Daddi, E., Bournaud, F., Walter, F., et al. 2010, ApJ, 713, 686
Daddi, E., Cimatti, A., Renzini, A., et al. 2004, ApJ, 617, 746
Daddi, E., Dickinson, M., Morrison, G., et al. 2007, ApJ, 670, 156
Davé, R., Finlator, K., & Oppenheimer, B. D. 2011, MNRAS, 416, 1354
Donley, J. L., Koekemoer, A. M., Brusa, M., et al. 2012, ApJ, 748, 142
Draine, B. T., Dale, D. A., Bendo, G., et al. 2007, ApJ, 663, 866
Draine, B. T., & Li, A. 2007, ApJ, 657, 810
Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, A&A, 468, 33
Elmegreen, B. G., Elmegreen, D. M., Fernandez, M. X., & Lemonias, J. J. 2009, ApJ, 692, 12
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Garn, T., Sobral, D., Best, P. N., et al. 2010, MNRAS, 402, 2017
Geach, J. E., Chapin, E. L., Coppin, K. E. K., et al. 2013, MNRAS, 432, 53
Geach, J. E., Dunlop, J. S., Halpern, M., et al. 2017, MNRAS, 465, 1789
Geach, J. E., Smail, I., Best, P. N., et al. 2008, MNRAS, 388, 1473
Genzel, R., Tacconi, L. J., Gracia-Carpio, J., et al. 2010, MNRAS, 407, 2091
Genzel, R., Tacconi, L. J., Lutz, D., et al. 2015, ApJ, 800, 20
Hodge, J. A., Carilli, C. L., Walter, F., et al. 2012, ApJ, 760, 11
Hopkins, A. M., & Beacom, J. F. 2006, ApJ, 651, 142
Hunt, L. K., Draine, B. T., Bianchi, S., et al. 2015, A&A, 576, A33
Ibar, E., Lara-López, M. A., Herrera-Camus, R., et al. 2015, MNRAS, 449, 2498
Ibar, E., Sobral, D., Best, P. N., et al. 2013, MNRAS, 434, 3218
Ilbert, O., Salvato, M., Capak, P., et al. 2008, in ASP Conf. Ser. 399, Panoramic Views of Galaxy Formation and Evolution, ed. T. Kodama, T. Yamada, & K. Aoki (San Francisco, CA: ASP), 169
James, A., Dunne, L., Eales, S., & Edmunds, M. G. 2002, MNRAS, 335, 753
Johnson, H. L., Harrison, C. M., Swinbank, A. M., et al. 2016, MNRAS, 460, 1059
Kennicutt, R. C., & Evans, N. J. 2012, ARA&A, 50, 531
Kennicutt, R. C., Jr. 1998, ApJ, 498, 541
Kereš, D., Katz, N., Weinberg D., H., & Dave, R. 2005, MNRAS, 363, 2
Kewley, L. J., & Ellison, S. L. 2008, ApJ, 681, 1183
Kewley, L. J., Geller, M. J., Jansen, R. A., & Dopita, M. A. 2002, AJ, 124, 3135
Kirkpatrick, A., Calzetti, D., Galametz, M., et al. 2013, ApJ, 778, 51
Lagos, C. d. P., Crain, R. A., Schaye, J., et al. 2015, MNRAS, 452, 3815
Lara-López, M. A., Cepa, J., Bongiovanni, A., et al. 2010, A&A, 521, L53
Leroy, A. K., Bolatto, A., Gordon, K., et al. 2011, ApJ, 737, 12
Lilly, S. J., Le Fevre, O., Hammer, F., & Crampton, D. 1996, ApJL, 460, L1
Lonsdale, C. J., Hacking, P. B., Conrow, T. P., & Rowan-Robinson, M. 1990, ApJ, 358, 60
Lutz, D., Poglitsch, A., Altieri, B., et al. 2011, A&A, 532, A90
Ma, C.-J., Smail, I., Swinbank, A. M., et al. 2015, ApJ, 806, 257
Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415
Magdis, G. E., Rigopoulou, D., Huang, J.-S., & Fazio, G. G. 2010, MNRAS, 401, 1521
Magnelli, B., Lutz, D., Saintonge, A., et al. 2014, A&A, 561, A86
Magnelli, B., Popesso, P., Berta, S., et al. 2013, A&A, 553, A132
Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, MNRAS, 408, 2115
Martig, M., Bournaud, F., Teyssier, R., & Dekel, A. 2009, ApJ, 707, 250
Oliver, S. J., Bock, J., Altieri, B., et al. 2012, MNRAS, 424, 1614
Oliver, S. J., Wang, L., Smith, A. J., et al. 2010, A&A, 518, L21
Oteo, I., Sobral, D., Ivison, R. J., et al. 2015, MNRAS, 452, 2018
Partridge, R. B., & Peebles, P. J. E. 1967, ApJ, 147, 868
Pettini, M., & Pagel, B. E. J. 2004, MNRAS, 348, L59
Popping, G., Somerville, R. S., & Trager, S. C. 2014, MNRAS, 442, 2398
Riechers, D. A., Bradford, C. M., Clements, D. L., et al. 2013, Natur, 496, 329
Rieke, G. H., Alonso-Herrero, A., Weiner, B. J., et al. 2009, ApJ, 692, 556
Rowlands, K., Dunne, L., Dye, S., et al. 2014, MNRAS, 441, 1017
Saintonge, A., Kauffmann, G., Kramer, C., et al. 2011, MNRAS, 415, 32
Salpeter, E. E. 1955, ApJ, 121, 161
Sawicki, M. 2012, PASP, 124, 1208
Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521
Schinnerer, E., Carilli, C. L., Scoville, N. Z., et al. 2004, AJ, 128, 1974
Schreiber, C., Elbaz, D., Pannella, M., et al. 2016, A&A, 589, A35
Scoville, N., Aussel, H., Sheth, K., et al. 2014, ApJ, 783, 84
Smail, I., Ivison, R. J., & Blain, A. W. 1997, ApJL, 490, L5
Sobral, D., Best, P. N., Matsuda, Y., et al. 2012, MNRAS, 420, 1926
Sobral, D., Best, P. N., Smail, I., et al. 2014, MNRAS, 437, 3516
Sobral, D., Smail, I., Best, P. N., et al. 2013, MNRAS, 428, 1128
Stanley, F., Harrison, C. M., Alexander, D. M., et al. 2015, MNRAS, 453, 591
Steidel, C. C., Giavalisco, M., Dickinson, M., & Adelberger, K. L. 1996, AJ, 112, 352
Stott, J. P., Sobral, D., Bower, R., et al. 2013, MNRAS, 436, 1130
Swinbank, A. M., Simpson, J. M., Smail, I., et al. 2014, MNRAS, 438, 1267
Swinbank, A. M., Smail, I., Chapman, S. C., et al. 2004, ApJ, 617, 64
Swinbank, A. M., Smail, I., Longmore, S., et al. 2010, Natur, 464, 733
Swinbank, A. M., Smail, I., Sobral, D., et al. 2012, ApJ, 760, 130
Symeonidis, M., Vaccari, M., Berta, S., et al. 2013, MNRAS, 431, 2317
Tacconi, L. J., Neri, R., Genzel, R., et al. 2013, ApJ, 768, 74
Takata, T., Sekiguchi, K., Smail, I., et al. 2006, ApJ, 651, 713
Telford, O. G., Dalcanton, J. J., Skillman, E. D., & Conroy, C. 2016, ApJ, 827, 35
Thomson, A. P., Ivison, R. J., Smail, I., et al. 2012, MNRAS, 425, 2203
Ueda, Y., Watson, M. G., Stewart, I. M., et al. 2008, ApJS, 179, 124
Whitaker, K. E., van Dokkum, P. G., Brammer, G., & Franx, M. 2012, ApJL, 754, L29
Wit, E., van den Heuvel, E., & Romeijn, J. W. 2012, Statistica Neerlandica, 66, 217
Wuyts, S., Förster Schreiber, N. M., Nelson, E. J., et al. 2013, ApJ, 779, 135
Yan, H., & Ma, Z. 2016, ApJL, 820, L16
Zanella, A., Daddi, E., Le Floc'h, E., et al. 2015, Natur, 521, 54