



Universiteit
Leiden
The Netherlands

The Brightest Galaxies in the Dark Ages: Galaxies : Dust Continuum Emission during the Reionization Era

Casey, C.M.; Zavala, J.A.; Spilker, J.; Cunha, E. da; Hodge, J.; Hung, C.L.; ... ; Drew, P.

Citation

Casey, C. M., Zavala, J. A., Spilker, J., Cunha, E. da, Hodge, J., Hung, C. L., ... Drew, P. (2018). The Brightest Galaxies in the Dark Ages: Galaxies : Dust Continuum Emission during the Reionization Era. *Astrophysical Journal*, 862(1), 77. doi:10.3847/1538-4357/aac82d

Version: Not Applicable (or Unknown)

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/68622>

Note: To cite this publication please use the final published version (if applicable).

THE BRIGHTEST GALAXIES IN THE DARK AGES: GALAXIES' DUST CONTINUUM EMISSION DURING THE REIONIZATION ERA

CAITLIN M. CASEY,¹ JORGE A. ZAVALA,¹ JUSTIN SPILKER,¹ ELISABETE DA CUNHA,² JACQUELINE HODGE,³ CHAO-LING HUNG,⁴ JOHANNES STAGUHN,^{5,6} STEVEN L. FINKELSTEIN,¹ AND PATRICK DREW¹¹*Department of Astronomy, The University of Texas at Austin, 2515 Speedway Blvd Stop C1400, Austin, TX 78712*²*Research School of Astronomy and Astrophysics, The Australian National University, Canberra ACT 2611, Australia*³*Leiden Observatory, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands*⁴*Department of Physics, Manhattan College, 4513 Manhattan College Pkwy, Bronx, NY 10471*⁵*NASA Goddard Space Flight Center, Code 665, Greenbelt, MD 20771*⁶*Bloomberg Center for Physics and Astronomy, Johns Hopkins University 3400 N. Charles Street, Baltimore, MD 21218*

ABSTRACT

Though half of cosmic starlight is absorbed by dust and reradiated at long wavelengths ($3\mu\text{m}$ – 3mm), constraints on the infrared through millimeter galaxy luminosity function (the ‘IRLF’) are poor in comparison to the rest-frame ultraviolet and optical galaxy luminosity function, particularly at $z \gtrsim 2.5$. Here we present a backward evolution model for interpreting number counts, redshift distributions, and cross-band flux density correlations in the infrared and submillimeter sky, from $70\mu\text{m}$ – 2mm , using a model for the IRLF out to the epoch of reionization. Mock submillimeter maps are generated by injecting sources according to the prescribed IRLF and flux densities drawn from model spectral energy distributions that mirror the distribution of SEDs observed in $0 < z < 5$ dusty star-forming galaxies (DSFGs). We explore two extreme hypothetical case-studies: a dust-poor early Universe model, where DSFGs contribute negligibly ($<10\%$) to the integrated star-formation rate density at $z > 4$, and an alternate dust-rich early Universe model, where DSFGs dominate $\sim 90\%$ of $z > 4$ star-formation. We find that current submm/mm datasets do not clearly rule out either of these extreme models. We suggest that future surveys at 2mm will be crucial to measuring the IRLF beyond $z \sim 4$. The model framework developed in this paper serves as a unique tool for the interpretation of multiwavelength IR/submm extragalactic datasets and will enable more refined constraints on the IRLF than can be made from direct measurements of individual galaxies’ integrated dust emission.

Keywords: galaxies: evolution — galaxies: starburst — submillimeter: galaxies

1. INTRODUCTION

The census of cosmic star-formation out to the highest redshifts is a central goal of galaxy evolution surveys and yet current measurements are imbalanced, biased towards unobscured star-formation tracers (Madau & Dickinson 2014). Finding the most distant galaxies, formed less than a billion years after the Big Bang, is of fundamental importance in order to observationally test theories of galaxy assembly. This includes constraining the Population III stellar initial mass function, the formation of early dust and metals, and the timescale of dark matter halo collapse. Significant effort and work

has been poured into taking census of galaxies detected via their rest-frame ultraviolet emission (e.g. Schiminovich et al. 2005; Dahlen et al. 2007; Reddy & Steidel 2009; Bouwens et al. 2012; Schenker et al. 2013; Ellis et al. 2013; Coe et al. 2013; Oesch et al. 2013; Finkelstein et al. 2013; Bouwens et al. 2015; Finkelstein et al. 2015). The presence of a strong Lyman break has successfully been used for redshift identification (Steidel et al. 1996) out to $z \sim 11$ (Oesch et al. 2016), revealing a peak in the cosmic star-formation rate density from $z \sim 2 - 4$ and values more consistent with the local Universe at earlier times ($z \sim 7 - 10$). Debates as to the slope of the cosmic star-formation rate density near the Epoch of Reionization (EoR) are forming over ever increasing samples of early-Universe Lyman-break galaxies (Oesch et al. 2013, 2014; McLeod et al. 2015).

While this work in the rest-frame UV, redshifted into the near-IR at $z > 8$, has been pioneering, similar surveys of the early Universe at long wavelengths have not kept pace. And yet, this long wavelength work is necessary in the census of cosmic star-formation, not least because we know roughly half of the energy from the extragalactic background radiation is output at long wavelengths. This is because ultraviolet light from young, massive stars is absorbed by dust and re-radiated. And it is clear that the conditions of the interstellar medium (ISM) and the environments of star-formation have tremendous impact on whether or not galaxies will appear largely unobscured or heavily obscured, thus whether or not they are counted in existing surveys. Due to their very high star-formation rates and thus extreme levels of obscuration (e.g. Brinchmann et al. 2004; Whitaker et al. 2017), dusty star-forming galaxies (DSFGs; Casey, Narayanan, & Cooray 2014a) are largely absent from the optical census of cosmic star-formation. Though there are some DSFGs that may appear in optical surveys as LBGs, often their rest-frame UV colors imply very little dust, thus star-formation rates that are factors ~ 100 times lower than implied by their long-wavelength emission (Casey et al. 2014b). While locally the population of bright DSFGs (SFRs $\gtrsim 100 M_{\odot} \text{ yr}^{-1}$) is negligible, at $z \sim 2 - 3$ the population is over one-thousand-fold more common and becomes the dominant factories of star-formation in the early Universe. Therefore, taking census of the Universe’s star-formation history requires a bolometric approach, analyzing galaxy populations detected via their direct starlight and those via their dust emission.

Galaxy surveys at long wavelengths have naturally been more limited by instrumentation and the additional hurdles involved in identifying galaxies’ redshifts – a characteristic which, for the LBG samples, is inferred directly from the observations used in their selection. From single-dish submillimeter and millimeter surveys, large beamsizes have obfuscated the identification of precise multiwavelength counterparts (Smail et al. 1997; Hughes et al. 1998; Barger et al. 1998; Chapman et al. 2003b). Even when multiwavelength counterparts are identified, redshift confirmation can be extremely challenging with low yields (Chapman et al. 2005; Casey et al. 2012a,b; Danielson et al. 2017; Casey et al. 2017). Only recent wide bandwidth receivers in the millimeter have made it possible to spectroscopically confirm high- z DSFGs without laborious and observationally expensive multiwavelength campaigns (Bradford et al. 2009; Vieira et al. 2013). Spectroscopic follow-up in the mm of the most luminous class of DSFGs has interestingly revealed a prominent population of sources at $3 < z < 7$

(Weiß et al. 2013; Riechers et al. 2013; Strandet et al. 2017; Marrone et al. 2017), but such surveys have not yet become efficient for large samples of less luminous (unlensed) DSFGs.

Knowing the prevalence of dust-obscured star-formation is particularly important at $z > 4$, when cosmic time becomes a constraint on the physical processes involved in producing dust, metals and stars seen in galaxies. For example, Capak et al. (2015) show a marked difference in the dust-to-gas ratio for a population of $z \sim 5$ normal star-forming galaxies; however, their sample was exclusively selected via rest-frame UV and optical surveys, which are biased towards low dust content. Unfortunately, current deep-field *HST* surveys are blind to $z > 4$ sources with $\gtrsim 50 M_{\odot} \text{ yr}^{-1}$ due to pencil-beam sky coverage limiting the dynamic range of observable galaxies. Existing samples of DSFGs at these redshifts are extremely bright ($> 2000 M_{\odot} \text{ yr}^{-1}$) and come from extremely shallow, biased surveys. The South Pole Telescope sample of DSFGs (e.g. Vieira et al. 2013), though less biased with color or SFR, is dominated by gravitationally-lensed sources whose volume density is nearly impossible to measure. Therefore, there is almost no constraint on the contribution of obscured star-formation to the cosmic star-formation rate density at $z \gtrsim 2$, and absolutely no constraint beyond $z \sim 5.5$ (see Figure 1).

Identifying high-redshift obscured galaxies has proven to be particularly challenging. IR color selection, like the technique used to identify *Herschel* 500 μm -risers (Pearson et al. 2013; Dowell et al. 2014; Ivison et al. 2016), seems to provide an effective route to several exciting, high- z discoveries (e.g. Oteo et al. 2017; Zavala et al. 2018), but the nature of source selection and follow-up make it difficult to back out any information on underlying population statistics. On the other hand, deep blank-field ALMA campaigns (Dunlop et al. 2016; Walter et al. 2016), which are not based on color selection, have failed to yield a population of very high-redshift sources.

In this paper, we describe a model for the far-infrared through millimeter emission of galaxies from $z = 0$ to $z \sim 10$ to explain the results of (sub)mm single-dish survey campaigns to-date. An accompanying paper presents results of analysis of the same models on scales observable with sensitive interferometers like ALMA (Casey et al. 2018, submitted). This paper follows other literature works which present similar models of the Universe’s (sub)mm emission, including the Simulated Infrared Dusty Extragalactic Sky (SIDES) and its predecessor models (Béthermin et al. 2012a; Béthermin et al. 2017), as well as the work of Zavala et al. (2014), which

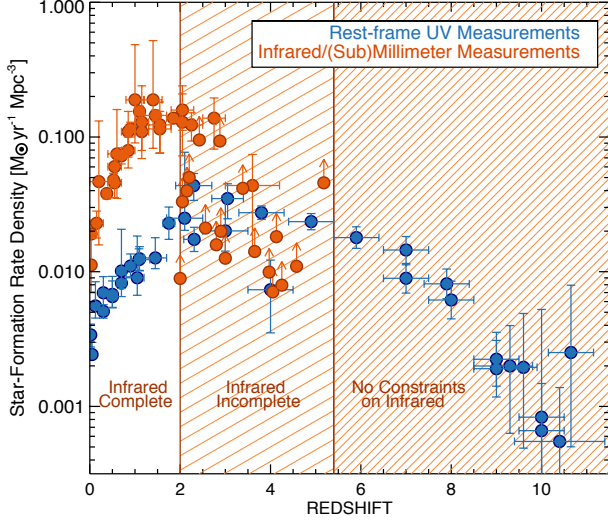


Figure 1. The cosmic star-formation history of the Universe as measured at rest-frame UV wavelengths (blue points; Schiminovich et al. 2005; Dahlen et al. 2007; Reddy & Steidel 2009; Bouwens et al. 2012; Schenker et al. 2013; Ellis et al. 2013; Coe et al. 2013; Oesch et al. 2013; Finkelstein et al. 2013; Bouwens et al. 2015; Finkelstein et al. 2015), and infrared through millimeter measurements (orange points; Le Floc’h et al. 2005; Magnelli et al. 2011; Gruppioni et al. 2013; Casey 2012; Casey et al. 2012a, 2013; Barger et al. 2012; Roseboom et al. 2013; Chapman et al. 2005; Wardlow et al. 2011) from facilities like the *Herschel Space Observatory*, SCUBA, and AzTEC. While far-infrared/(sub)mm surveys (globally referred to as the IR) have mapped obscured star-formation with individual galaxy detections out to $z \sim 7$ (Strandet et al. 2017; Marrone et al. 2017), there are few constraints on their SFRD contribution at $z \gtrsim 2.5$ due to sample incompleteness. Our current understanding of star-formation in the early Universe is severely limited by the lack of IR constraints, particularly beyond $z \sim 4$.

explains the different redshift distributions of (sub)mm-selected populations with a single underlying source population. We explore some of the strengths of each of these models by analyzing differences in the *a priori* assumptions and approaching from a different perspective focused on the total infrared through millimeter galaxy luminosity function (henceforth referred to as the IR luminosity function, or IRLF, in this paper).

We use existing measurements of submm number counts, redshift distributions, and multi-band flux information collated from across the literature to comment on the shape and behavior of the IR luminosity function of galaxies (from dust continuum) out to the epoch of reionization. Two extreme case studies are used to frame this discussion and outline goals of future work. One case study assumes a dust-poor early Universe, similar to existing models used by the rest-frame UV com-

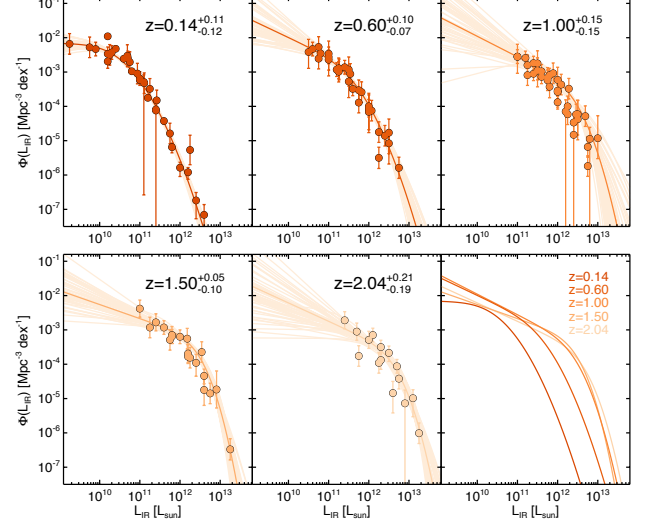


Figure 2. A summary of integrated L_{IR} luminosity functions in the literature, as collated in Casey, Narayanan, & Cooray (2014a). Original data is from: Sanders et al. (2003); Le Floc’h et al. (2005); Casey et al. (2012a); Gruppioni et al. (2013); Magnelli et al. (2011, 2013). Five redshift ranges are shown: $z = 0.14^{+0.11}_{-0.12}$, $0.60^{+0.10}_{-0.07}$, 1.00 ± 0.15 , $1.50^{+0.05}_{-0.10}$, and $2.04^{+0.21}_{-0.19}$, and a sixth panel shows the relative evolution between them. Though these fits highlight a continuous double powerlaw form, the broken double powerlaw is statistically indistinguishable and we adopt it for its simplicity for the rest of this work.

munity, while the other assumes a dust-rich early Universe. It is important to point out that here dust-rich does not refer to the content of all galaxies uniformly, but rather, the abundance of very dust-rich DSFGs relative to UV-bright galaxies. The construction of the model framework and its assumptions are described in § 2, and we compare our results with literature datasets and other models in § 3. The implications of our constraints are discussed fully in § 4 and in § 5 we conclude. We assume a *Planck* cosmology throughout this paper, adopting $H_0 = 67.7 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $\Omega_\lambda = 0.6911$ (Planck Collaboration et al. 2016), and where SFRs are alluded to, we assume a Chabrier IMF (Chabrier 2003).

2. MODEL CONSTRUCTION

Our backward-evolution model provides a prediction of far-infrared and submillimeter flux number counts (from $70\mu\text{m}$ through 2 mm), redshift distributions and overlaps in populations, given a parameterized, evolving galaxy luminosity function. This model is empirically-driven and motivated by existing measurements of the galaxy luminosity function in the infrared and their measured SED characteristics. What follows here is a step-by-step detailed description of the model, begin-

ning with the luminosity function. A table summarizing all of the model assumptions, including the equations described below, is given in Table 2 at the end of the description of each component.

2.1. The IR Luminosity Function

Figure 2 shows a summary of measured galaxy luminosity functions in the IR, as collated in Casey, Narayanan, & Cooray (2014a). There is strong luminosity evolution evident in these data, with the possibility of some minor evolution in galaxy number density. The shape of the IRLF is poorly constrained relative to the rest-frame UV/optical luminosity function of galaxies. We show a continuous double powerlaw fit in Figure 2, but emphasize that there is no statistical difference between adopting a continuous double powerlaw and a broken double powerlaw. A Schechter function is deemed inappropriate for these IR-luminous galaxies because the bright-end falls off gradually and not exponentially. Due to its simplicity and intuitive nature, we adopt a broken double powerlaw model for Φ which is a function of both redshift z and IR luminosity, which we will simply denote L :

$$\Phi(L, z) = \begin{cases} \Phi_* \left(\frac{L}{L_*} \right)^{\alpha_{\text{LF}}} & : L < L_* \\ \Phi_* \left(\frac{L}{L_*} \right)^{\beta_{\text{LF}}} & : L \geq L_* \end{cases} \quad (1)$$

It is clear that L_* evolves strongly with redshift (as shown in Figure 2), and it is possible that Φ_* , α_{LF} and β_{LF} also have some redshift dependence, although there is little data to constrain this currently (though our accompanying ALMA-focused paper addresses possible evolution in α_{LF} with redshift and implications on detections in small ALMA deep fields). The units of Φ are $\text{Mpc}^{-3} \text{dex}^{-1}$ and L and L_* are in L_\odot . We discuss the values and nature of the redshift evolution of these parameters in § 2.6.

2.2. Galaxies' IR Spectral Energy Distributions

Modeling the multi-wavelength (sub)millimeter emission of galaxies requires a keen understanding of their spectral energy distributions (SEDs) in addition to the underlying galaxy luminosity function. Dust radiative transfer models (Silva et al. 1998; Dopita et al. 2005; Siebenmorgen & Krügel 2007) and observations of local IR-luminous galaxies (U et al. 2012) show that the far-infrared/submillimeter SEDs of galaxies are well-represented by a single modified blackbody, with additional emission in the mid-infrared representing the emission of less massive and more concentrated pockets of warm to hot dust in the galaxy's ISM. Emission from polycyclic aromatic hydrocarbons (PAHs) can also

dominate this mid-infrared regime, contributing as much as 10% to the total integrated IR luminosity of galaxies. While many works in the literature use detailed empirically-driven templates (Chary & Elbaz 2001; Dale et al. 2001; Dale & Helou 2002; Draine & Li 2007; Rieke et al. 2009) or energy-balance techniques (Burgarella et al. 2005; da Cunha et al. 2008, 2013; Noll et al. 2009) to model the emission of high- z submm-detected galaxies, the detail of these models goes beyond the constraints of existing data for large DSFG samples.

For the purposes of this paper, we adopt a very simple four parameter mid-infrared powerlaw + modified blackbody (Blain et al. 2003) fit as described in Casey (2012). The free parameters of the model are the luminosity L (the integral under the curve, roughly scaling to its normalization), the dust temperature T_{dust} (related to the wavelength where the SED peaks, λ_{peak}), the mid-infrared powerlaw slope α_{MIR} , and the emissivity spectral index β_E (we give it the subscript to distinguish with β_{LF} , the bright-end slope of the IRLF). For the purposes of our model we fix the latter two parameters to $\alpha_{\text{MIR}} = 2.0$ and $\beta_E = 1.8$ in line with the average constraints from well-characterized galaxies both in the nearby and distant Universe (e.g. Paradis et al. 2010). The adoption of $\alpha_{\text{MIR}} = 2.0$ measured as the median mid-IR slope for GOALS galaxies (U et al. 2012) accounts for both hot dust emission and PAH emission via its integral, though does not directly spectrally model the PAH features because we determine this to only have a significant effect on galaxy observability at rest-frame wavelengths $< 10 \mu\text{m}$, which makes up a negligible fraction of the total power in the bands analyzed in this paper.

Note that the dust temperature, T_{dust} , represents the temperature of the cold dust in the ISM, and the temperature input for the dominating cold-dust modified blackbody component of the SED. Its relationship to the peak wavelength of the SED, λ_{peak} , depends on the adopted dust opacity model; in this work we assume that $\tau = 1$ at $\lambda_{\text{rest}} = 100 \mu\text{m}$ (Conley et al. 2011), whereby the blackbody is optically thick at shorter wavelengths and optically thin at longer wavelengths. This is consistent with observations of DSFGs in the local Universe; there is little evidence to suggest that this would not also hold for DSFGs in the early Universe. Though a different assumption of opacity will have a dramatic impact on the relationship between T_{dust} and λ_{peak} (see Figure 20 of Casey, Narayanan, & Cooray 2014a), we choose to parameterize our model using λ_{peak} instead of T_{dust} . This choice makes our SEDs insensitive to different adopted opacity models. Hence the rest of this

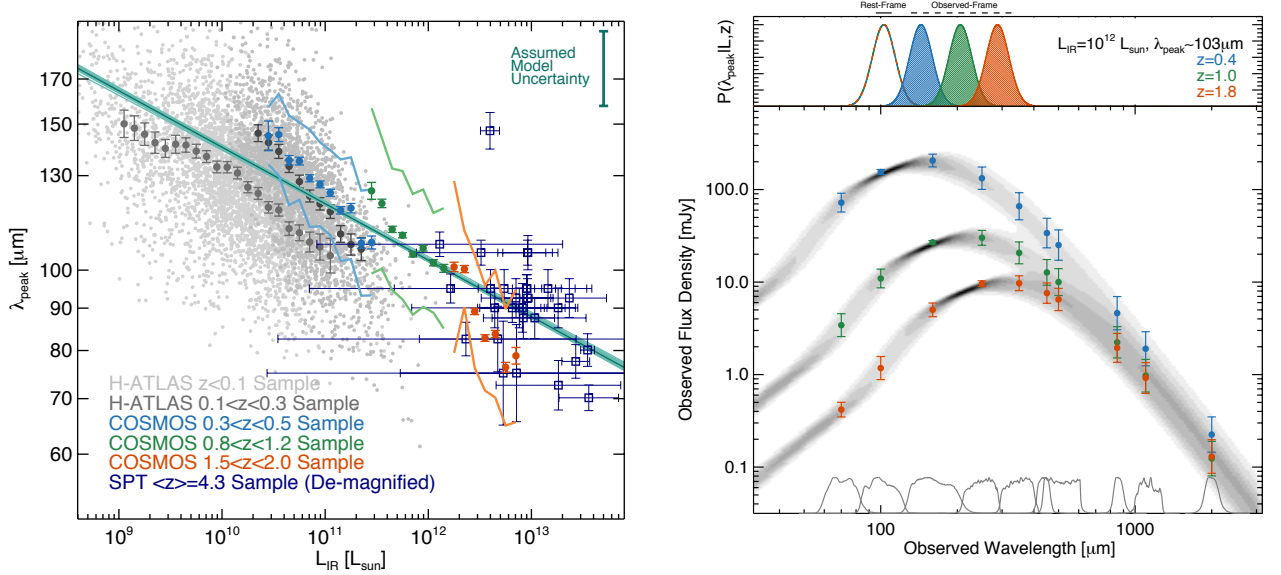


Figure 3. LEFT: The relationship between luminosity and dust temperature, shown here in observable quantities: L_{IR} and rest-frame peak wavelength λ_{peak} of S_{ν} . The local sample (Valiante et al. 2016) is shown as gray points in two redshift bins $z < 0.1$ (lightest gray) and $0.1 < z < 0.3$ (light gray). Darker gray points are the median values of λ_{peak} at a given L_{IR} for each sample. Higher redshift galaxies (Lee et al. 2013) sit on the extension of this relationship toward higher luminosities; 1σ scatter is shown as light blue, green and orange lines. Overplotted are the sample of South Pole Telescope DSFGs with well-measured SEDs and constrained magnification factors (Strandet et al. 2016) with a median redshift of $\langle z \rangle = 4.3$. The adopted model is overplotted as a teal line with associated fit uncertainty, and the scatter about that model used to generate a diversity of SEDs at all redshifts and shown in the upper right. RIGHT: Given a redshift and L_{IR} , here we show an example of how we estimate far-IR through millimeter flux densities to inject into our model maps. The example sources have a fixed $L_{\text{IR}} = 10^{12} L_{\odot}$, but sit at different redshifts, $z = 0.4$, 1.0 and $z = 1.8$. The range of rest-frame peak wavelengths is represented by the open Gaussian distribution at top and the observed-frame λ_{peak} by the filled distributions at each redshift. Below, we generate 1000 SEDs for each probability distribution in λ_{peak} given L_{IR} , and overplot the median predicted flux densities across the far-IR/mm bands. The filter profiles of the bands we use for this simulation are inset and described in the text.

paper discusses the idea of temperature only through the measurable quantity λ_{peak} .

Figure 3 shows the existing empirical constraints on the peak SED wavelength of IR-luminous galaxies as a function of L and z . It has been known for some time that there is a direct correlation between galaxies’ intrinsic IR luminosity (or total star-formation rate) and their observed peak of the IR SED, or dust temperature (Sanders et al. 2003; Chapman et al. 2003a). Galaxies with higher IR luminosities have intrinsically hotter luminosity-weighted dust temperatures, or lower values of λ_{peak} . We show the extensive data from the H-ATLAS survey (Valiante et al. 2016) largely encompassing galaxies from $0 < z < 0.5$ and with *Herschel* SPIRE-detected galaxies in the COSMOS field extending out to $z \sim 2$ (Lee et al. 2013). Though samples at higher redshifts are sparse, we draw on the Strandet et al. (2016) compilation of statistics on the South Pole Telescope (SPT)-detected DSFGs that have well-constrained SEDs and a median redshift of $\langle z \rangle = 4.3$; these high-redshift DSFGs seem to follow the same broad trend, where higher luminosity galaxies have hotter SEDs.

All SEDs of galaxies from the literature are refit using the Casey (2012) SED fitting method to provide uniform analysis of their characteristics. We test for biases introduced by limited band coverage, only including galaxies with sufficiently robust photometric measurements above detection limits where selection biases are negligible (more details are provided in § A.2). The $L_{\text{IR}}-\lambda_{\text{peak}}$ relationship can be modeled by a powerlaw such that:

$$\langle \lambda_{\text{peak}}(L) \rangle = \lambda_0 \left(\frac{L}{L_t} \right)^{\eta} \quad (2)$$

We measure $\lambda_0 = 102.8 \pm 0.4 \mu\text{m}$ at $L_t \equiv 10^{12} L_{\odot}$ and $\eta = -0.068 \pm 0.001$ from the aggregate samples plotted on Figure 3. We do note that the local galaxy sample, particularly at $z < 0.1$ appears offset from the best-fit $L_{\text{IR}}-\lambda_{\text{peak}}$ relationship toward slightly warmer temperatures. We discuss this deviation further in Appendix A.2 but argue here that the shift seen in low-redshift galaxies does not impact either our number counts or our inferred redshift distributions for sources found on 1 deg^2 scales. This is primarily because of the relative rarity of $z \lesssim 0.3$ dust-obscured galaxies relative to $z > 1$ sources.

Physically, this $L_{\text{IR}}-\lambda_{\text{peak}}$ relationship can be thought of as a galaxy-scale Stefan-Boltzmann law for the cold ISM. While a direct translation of Stefan-Boltzmann to $L_{\text{IR}}-\lambda_{\text{peak}}$ space would imply a value of $\eta = -0.25$, variations in galaxy shape (which are certainly not spherical or emitting isotropically), as well as a correlation of galaxies' sizes (or effective surface areas) with luminosity, and the fact that galaxies' dust emission is not a perfect blackbody, lead to the shallower value of $\eta = -0.068 \pm 0.001$. While this provides some context for shallower η slopes, we emphasize that it is an empirically measured quantity as taken from the samples in Figure 3. Folding in the assumed dust emissivity index, $\beta_E = 1.8$, η would change to -0.17 , and radiative transfer modeling suggest similarly intermediate slopes of η , for example $\eta \approx -0.16$ found in [Siebenmorgen & Krügel \(2007\)](#), who use a spherically symmetric model accounting for different star-formation rates, sizes and dust masses. A more in depth analysis of the $L_{\text{IR}}-\lambda_{\text{peak}}$ relationship is needed to understand the physical drivers behind the observed trends and possible evolution, but that is beyond the scope of this work. The exact value of η becomes less important when considering the observed scatter of real galaxies about the relation.

The average model uncertainty adopted, $\sigma_{\langle \log \lambda_{\text{peak}} \rangle} = 0.045$ (corresponding to $\Delta \lambda_{\text{peak}} / \lambda_{\text{peak}} \approx 10\%$), is shown in the upper left of Figure 3 and is derived from the average deviation of individual galaxies about the median $L_{\text{IR}}-\lambda_{\text{peak}}$ relationship. Each galaxy in the model is assigned a dust temperature, or λ_{peak} value, according to the probability density function in $\log \lambda_{\text{peak}}$ for that galaxy's redshift and L .

2.3. Relating luminosity to the SED

The next step in building our model is to generate fake maps of the sky at a variety of far-IR/submm wavelengths, and to do that we must generate a list of input sources drawn from our luminosity functions and use our data constraints to assign a best-guess far-infrared SED. For example, if we are to inject one source into the map with $L_{\text{IR}} = 10^{12} L_{\odot}$, we can predict that its rest-frame peak wavelength is close to $\langle \lambda_{\text{peak}}(L|z) \rangle = 103 \mu\text{m}$. Thus, each injected source is 'assigned' a peak wavelength (and thus far-IR SED) after drawing from a normal probability distribution in $\log \lambda_{\text{peak}}(L)$ with width σ . Such a distribution in rest-frame λ_{peak} is shown in the top right panel of Figure 3 (unfilled). The implied distributions in observed-frame peak wavelength are shown as filled histograms at each redshift. With a peak wavelength drawn from the probability distribution in $\log \lambda_{\text{peak}}$, an SED is constructed using the [Casey \(2012\)](#) analytic approach and flux densities are mea-

sured across the far-infrared through millimeter bands with their filter bandpasses. Our initial phase of modeling includes the following filters: *Herschel* PACS at $70 \mu\text{m}$, $100 \mu\text{m}$, and $160 \mu\text{m}$ ([Poglitsch et al. 2010](#)), *Herschel* SPIRE at $250 \mu\text{m}$, $350 \mu\text{m}$, and $500 \mu\text{m}$ ([Griffin et al. 2010](#)), SCUBA-2 at $450 \mu\text{m}$ and $850 \mu\text{m}$ ([Holland et al. 2013](#)), AzTEC at 1.1 mm ([Wilson et al. 2008](#)), a hypothetical TolTEC filter at 1.4 mm , and GISMO at 2 mm ([Staguhn et al. 2014](#)). We also note that ongoing 2 mm surveys at the IRAM30m telescope have begun with the NIKA-2 instrument ([Catalano et al. 2016](#)); the 2 mm beamsize at IRAM is $16.5''$. This is not sufficiently different than the beamsize with the LMT as to cause differences in the measured 2 mm number counts, but of course does matter in the identification of multiwavelength counterparts. An accompanying paper models emission in ALMA bands 3, 4, 6, and 7. Given the lack of instrumentation available at 3 mm on single-dish facilities to map large areas, plus the need to push deeper at 3 mm to detect galaxies of matched luminosity as those found in 1 mm or 2 mm , we do not model 3 mm single-dish continuum number counts in this paper. However, the accompanying paper analyzing modeled ALMA datasets does find that the 3 mm channel can be quite useful in constraining source densities at high-redshift.

2.4. Impact of the CMB at Long Wavelengths

With the goal of estimating the dust continuum emission of galaxies near the epoch of reionization, it is important to consider the effect of heating from the cosmic microwave background (CMB). [da Cunha et al. \(2013\)](#) explore the impact of the CMB on dust continuum and CO observations in detail and we refer the reader to their paper for more contextual background. Towards higher redshifts, the temperature of the CMB itself was sufficiently large to heat the internal ISM of galaxies forming during that epoch (where $T_{\text{CMB}} \sim T_{\text{dust}}$). This causes a boost in the submillimeter/millimeter output of the ISM, however, observationally this results in a net loss in flux density (compared to the absence of the CMB) because galaxies must always be detected in contrast to the CMB thermal background. To summarize the discussion in [da Cunha et al. \(2013\)](#), we must alter the fitted dust temperature for sources according to this effect, first by adjusting their internal dust-temperature:

$$T_{\text{dust}}(z) = \left((T_{\text{dust}}^{z=0})^{4+\beta_E} + T_{\text{CMB}}^{z=0\ 4+\beta_E} [(1+z)^{4+\beta_E} - 1] \right)^{\frac{1}{4+\beta_E}} \quad (3)$$

Here $T_{\text{dust}}^{z=0}$ represents the dust temperature the galaxy would have at $z = 0$ or in the absence of the CMB, the emissivity spectral index is taken to be $\beta_E = 1.8$, and

$T_{\text{CMB}}^{z=0} = 2.73 \text{ K}$. Then the fraction of the flux density that is observable against the CMB background is:

$$f(z, T_{\text{dust}}) = 1 - \frac{B_{\nu}[T_{\text{CMB}}(z)]}{B_{\nu}[T_{\text{dust}}(z)]} \quad (4)$$

In other words, this is the ratio of the galaxy’s observed flux density against the CMB versus what the galaxy’s flux density would be in the absence of the CMB. Procedurally, we do this by first computing a galaxy’s SED as it would be in the absence of the CMB, and then we fold in this effect by multiplying that flux density by the factor in Eq. 4 that effectively shifts the peak towards warmer temperatures and lower flux densities. Note that this adjustment *does* depend on the input dust temperature of our model, T_{dust} , and not the observable peak wavelength, λ_{peak} ; this means that our assumptions about opacity – that SEDs are optically thick to rest-frame $\sim 100 \mu\text{m}$ – impact the perceived impact of the CMB at high- z . For a galaxy that peaks at rest-frame $100 \mu\text{m}$, the difference between an optically thick blackbody and optically thin blackbody is $\sim 10 \text{ K}$, which translates to about a 10% difference in anticipated impact of the CMB on that galaxy’s SED.

The impact of the CMB is most prominent at $z > 5$ and $\lambda_{\text{obs}} > 1 \text{ mm}$. The effect is not uniform for all galaxies at this epoch, however, as some will have intrinsically warmer temperatures than others. If the $L_{\text{IR}}\text{--}\lambda_{\text{peak}}$ trend seen in Figure 3 holds (in some form) at high-redshift, then this will result in the lowest luminosity galaxies falling below the detection limit out of our survey, leaving only the brighter galaxies with intrinsically warmer temperatures to be detected. This has some important implications on the search for dust continuum emitters towards the EoR, which is discussed more in the context of our results in § 4.

Because this paper primarily focuses on galaxies above L_{\star} , detectable with single-dish submillimeter facilities on deg^2 scales, we favor the opacity model that includes self-absorption on the Wien tail as indicated in Table 2. We note that an optically thin assumption would only alter the resulting flux densities (after correction for the CMB) a small amount for these characteristically luminous sources, since they are likely to be significantly hotter than the CMB at most redshifts. For example, the CMB will result in a $\approx 30\%$ flux density reduction for $10^{12} L_{\odot}$ sources at $z \sim 6$, and the difference in the deficit between the optically thin model and general opacity model is of order 10%.

2.5. Impact of AGN Dust Heating and Synchrotron Emission

One real effect that is not explicitly baked into the model is the impact of AGN. The shortest-wavelength

bands, $70\text{--}160 \mu\text{m}$, are significantly dominated by DSFGs containing AGN at $z \approx 1\text{--}2$ (Kartaltepe et al. 2012; Kocevski et al. 2012; Brown et al. 2018). Additional dust-heating by AGN in the vicinity of the central dust torus to temperatures of a few $\sim 100 \text{ K}$ typically flatten out the mid-infrared spectrum, to $\alpha_{\text{MIR}} \approx 1\text{--}1.5$. This additional emission is not added into our model directly, but needs to be accounted for after the fact. To do this, we use measurements of $0 < z < 2$ AGN luminosity functions (Lacy et al. 2015) as measured in the mid-infrared, and randomly draw sources at the same redshifts and source densities and reassign their flux densities to account for shallower mid-infrared slopes (which we assign to be $\alpha_{\text{MIR}} = 1.5$ at $L_{\text{IR}} = 10^{11} L_{\odot}$, up to $\alpha_{\text{MIR}} = 1$ at $L_{\text{IR}} = 10^{13} L_{\odot}$). This effectively provides a boost of order $1.1\text{--}2.0\times$ to the flux densities in the *Herschel* PACS bands, and does not impact any of the longer-wavelength bands.

While AGN might also be thought to possibly contribute to sources detected at long-wavelengths ($> 1 \text{ mm}$) through radio-loud synchrotron emission, the number counts generated from such sources should be quite low in surveys $\sim 1 \text{ deg}^2$ of the depths we explore (de Zotti et al. 2005; Tucci et al. 2011). Such radio-loud quasars would become much more dominant at higher flux densities covering much larger areas, like those explored by the South Pole Telescope. We exclude such sources from our model since we are primarily focused on exploring the prevalence of DSFGs in the $\sim \text{mJy}$ flux density regime.

2.6. Redshift Evolution of the model

We build the majority of uncertainty of our model into the galaxy luminosity function, such that its evolution with redshift is unconstrained beyond $z \gtrsim 2.5$, but it must be modeled in order to reproduce millimeter deep field number counts, redshift distributions, and correlation of flux densities for sources between different selection wavelengths.

Our model posits that the evolution of L_{\star} follows:

$$L_{\star}(z) \propto \begin{cases} (1+z)^{\gamma_1} & : z \ll z_{\text{turn}} \\ (1+z)^{\gamma_2} & : z \gg z_{\text{turn}} \end{cases} \quad (5)$$

And similarly, that the evolution of Φ_{\star} follows:

$$\Phi_{\star}(z) \propto \begin{cases} (1+z)^{\psi_1} & : z \ll z_{\text{turn}} \\ (1+z)^{\psi_2} & : z \gg z_{\text{turn}} \end{cases} \quad (6)$$

To achieve these conditions with a smooth transition at a ‘turnover’ redshift, z_{turn} , we gradually transition from one redshift dependence to the other over a redshift interval that has thickness z_{w} . For example, L_{\star} might

evolve like $(1+z)^{\gamma_1}$ up to $z \sim 1.5$, and then transition to $(1+z)^{\gamma_2}$ gradually by a redshift of $z \sim 3.5$ (in this example, $z_{\text{turn}} = 2.1$ and $z_w = 2.0$ are adopted as appropriate ballpark estimates for one of our two models). We parameterize this in terms of x such that:

$$\begin{aligned} x &\equiv \log_{10}(1+z) \\ x_t &\equiv \log_{10}(1+z_{\text{turn}}) \\ x_w &\equiv \frac{z_w}{\ln(10)(1+z_{\text{turn}})} \end{aligned} \quad (7)$$

and then L_\star evolves with x like:

$$\begin{aligned} \log_{10} L_\star(x) &= \frac{(\gamma_2 - \gamma_1)x_w}{2\pi} \left[\ln \left(\cosh \left(\pi \frac{x - x_t}{x_w} \right) \right) \right. \\ &\quad \left. - \ln \left(\cosh \left(-\pi \frac{x_t}{x_w} \right) \right) \right] \\ &\quad + \frac{(\gamma_2 - \gamma_1)}{2} x + \log_{10}(L_0) \end{aligned} \quad (8)$$

Similarly,

$$\begin{aligned} \log_{10} \Phi_\star(x) &= \frac{(\psi_2 - \psi_1)x_w}{2\pi} \left[\ln \left(\cosh \left(\pi \frac{x - x_t}{x_w} \right) \right) \right. \\ &\quad \left. - \ln \left(\cosh \left(-\pi \frac{x_t}{x_w} \right) \right) \right] \\ &\quad + \frac{(\psi_2 - \psi_1)}{2} x + \log_{10}(\Phi_0) \end{aligned} \quad (9)$$

This functional form follows the same structure as is often adopted by the rest-frame UV community in analyzing the luminosity function for Lyman-break galaxies (LBGs). Figure 4 shows measured constraints for both L_\star and Φ_\star in the rest-frame UV with best-fit values of L_0 , Φ_0 , γ_2 , γ_1 , ψ_2 , ψ_1 , z_{turn} , and z_{width} for those UV measurements as parameterized above. In contrast, we show various measurements of the IR values of L_\star and Φ_\star from the literature, which show a very different (and less well-constrained) evolutionary path. We overplot the adopted evolutionary curves for the models in this paper in orange.

The redshift dependence of SED characteristics, or the $L_{\text{IR}}-\lambda_{\text{peak}}$ relationship shown in Figure 3, is a bit more difficult to constrain given the lack of complete samples in the early Universe, and the introduction of potential dust-temperature biases. Given the consistency of SEDs across $0.3 < z < 5$ we proceed with a non-evolving $L_{\text{IR}}-\lambda_{\text{peak}}$ relationship though we discuss possible caveats of this assumption in § A.2. This is, by design, open to revision if it is later determined that high- z DSFG SEDs do evolve with redshift or exhibit some other bulk characteristics or trends with higher quality data.

2.7. Generating Source Maps

Sources are injected into a series of maps of fixed solid angle; for this paper we generate 1 deg^2 maps with a

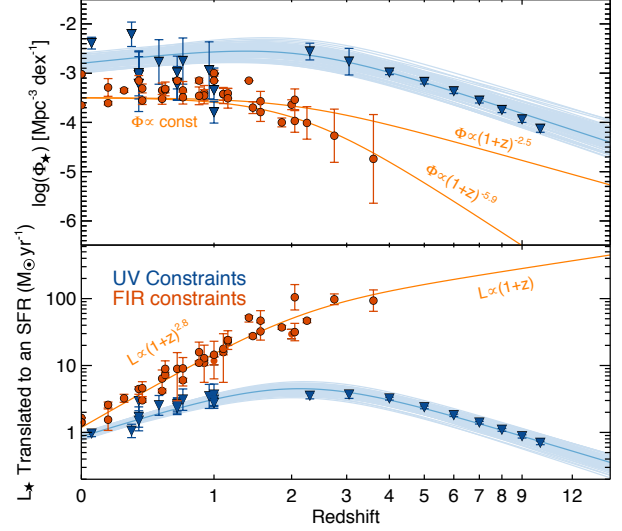


Figure 4. A comparison of the luminosity function parameters L_\star and Φ_\star from the rest-frame UV community, and from the FIR/submm community. Data relevant to the rest-frame UV luminosity function is gathered from Arnouts et al. (2005), Reddy & Steidel (2009), and Finkelstein (2016), where M_\star has been translated from a magnitude to a SFR for direct comparison to the IR data. Data from the FIR/submm community comes from Le Floc’h et al. (2005), Caputi et al. (2007), Goto et al. (2010), Magnelli et al. (2011), Magnelli et al. (2013), and Gruppioni et al. (2013). Values for IR data have been renormalized to match at $z \sim 0$. The shaded blue region shows a range of plausible models for the UV LF of the form shown in Eq’s 8 & 9 using $z_{\text{turn}} = 3.5$. The adopted parameterizations in this paper are shown in light orange. The primary difference in proposed outcomes is the high-redshift evolution of Φ_\star , either evolving steeply $\propto (1+z)^{-5.9}$ (Model A) or more gradually $\propto (1+z)^{-2.5}$ (Model B).

$0.5''$ pixel scale but this is easily adjusted to test observational setups different from those described herein. Sources are injected with uniformly random positions and with a surface density determined by the projection of the galaxy luminosity function and flux densities from inferred SEDs. The effect of the CMB heating of high- z galaxies’ ISM is taken into account, impacting the injected sources’ final observed flux densities. Each filter has its own map, and though the positions are conserved from wavelength to wavelength, no clustering is taken into account; we compare our model predictions to the SIDES project, which does incorporate clustering from semi-analytic models in § 3. Input flux densities, positions and redshifts are recorded for later use. After sources have been injected, the maps are convolved with the filter beam. The beam is taken from real data maps by stacking hundreds of significant detections at each wavelength observed with each facility (note that

Table 1. Characteristics of Observational Setup

PASSBAND	INSTRUMENT/ TELESCOPE	BEAMSIZE FWHM ["]	RMS [mJy]
70 μ m	PACS (<i>Herschel</i>)	5	0.4
100 μ m	PACS (<i>Herschel</i>)	7	0.4
160 μ m	PACS (<i>Herschel</i>)	12	0.9
250 μ m	SPIRE (<i>Herschel</i>)	18	5.8
350 μ m	SPIRE (<i>Herschel</i>)	25	6.3
500 μ m	SPIRE (<i>Herschel</i>)	36	6.8
450 μ m	SCUBA-2 (JCMT)	7	1.0
850 μ m	SCUBA-2 (JCMT)	15	0.8
1100 μ m	AzTEC (32 m LMT)	8.5	0.3
1400 μ m	TOLTEC(50 m LMT)	6.9	0.3
2000 μ m	GISMO/TOLTEC (50 m LMT)	9.9	0.1

Notes. This table summarizes the different observational setups we test for our 1 deg² simulations from 70 μ m–2 mm using various past/existing instruments. The simulations at 1.4 mm do not explicitly simulate observations from any existing instrument, though will be analogous to future surveys from the TolTEC instrument at the LMT; in this paper, they serve as a good analogue to unlensed South Pole Telescope-detected 1.4 mm-selected sources, which are magnified by factors of $\mu = 5$ –20.

the beams are not well represented by a 2D Gaussian, as described further in [Coppin et al. 2015](#)). We then generate a noise map by convolving the beam with a standard normal distribution of pixel values, and rescaling the resulting noise map to the appropriate RMS. This noise map is then added to the beam-convolved map with source injections. Maps are then renormalized so the mode is equal to zero; this adjustment is only significant for the mock *Herschel* SPIRE maps but is in line with the instrument’s flux calibration procedure ([Griffin et al. 2010](#)). The details of the noise and beam characterization are given in Table 1 which summarizes observationally-driven model inputs. Example cutouts from the fake maps are shown in Figure 5 (the differences between the two models highlighted in the figure are described in the next section).

Sources are identified in the mock maps by first constructing signal-to-noise maps (‘SNR’ map), by dividing the simulated map by the instrumental noise as quoted in Table 1. All significant peaks in the SNR map are then identified with a ‘region grow’ algorithm in IDL. Sources’ positions and flux densities are then reported as corresponding to the point of their peak signal-to-noise. The threshold for detection, or the lower limit of SNR is initially set to 3.5, although we conservatively limit our analysis to $>5\sigma$ sources when discussing source

redshift distributions. No adjustments for confusion or Eddington boosting are made, as all comparisons with the literature are made against raw quantities.

2.8. Two Case Studies: Impact of Parameters

Table 3 provides a list of all of the tunable parameters of the model; the reader should consult this table in conjunction with Table 2 for a complete understanding of the model construction and parameter space. Fifteen different parameters are listed, and though all could theoretically be left open, most are already constrained well by existing datasets, while others are relatively unconstrained and are the focus of our study. Those that are well constrained are so noted in the table; the justification of their choice values and the impact of changing their values is discussed further in Appendix A.

Some of the most impactful parameters that are fixed for this model are $\gamma_1 = 2.8$, which traces the evolution of L_\star towards much higher luminosities from $z = 0$ to $z \sim 2$ as shown in Figure 4. Similarly, the corresponding number density Φ_\star does not evolve over the same interval, so we fix $\psi_1 = 0$. A simple set of tests – sampling different potential values of z_{turn} and z_w – reveal that the turnover redshift must be close to $z = 2$, otherwise the measured number counts comparison will be off substantially, underestimated if $z_{\text{turn}} \ll 2$ and overestimated if $z_{\text{turn}} \gg 2$. We explicitly choose the values of z_{turn} (either =1.8 or =2.1, depending on high- z evolutionary parameters), and $z_w = 2.0$ so that the measured number density evolution of LIRGs, ULIRGs, and the total IR contribution to the star-formation rate density is well matched to data (see Figure 6).

The most highly uncertain quantities (γ_2 and ψ_2) describe the evolution of L_\star and Φ_\star beyond $z \gtrsim 2$, where measurements are sparse. In this paper we present two case studies, adopting dramatically different values for ψ_2 , signifying either a dust-poor early Universe, or extremely dust-rich early Universe. Both of these models adopt $\gamma_2 = 1$, asserting that L_\star continues to evolve upwards toward higher redshifts. A positive value of γ_2 is chosen for three reasons: adopting $\gamma_2 \leq 0$ underpredicts IR number counts above ~ 1 mJy regardless of adopted evolution of source number density (ψ_2), a reversal might also imply evolution back towards warmer dust temperatures at high-redshift which is not seen for SPT-detected galaxies (contradicting the claims of [Faisst et al. 2017](#)), and adopting $\gamma_2 = 1$ neatly results in L_\star consistent with L_\star of the quasar luminosity function at $z > 4$ ([Hopkins et al. 2007](#)). This positive value for γ_2 is also consistent with reports in the literature of a dramatically-bright L_\star value towards the epoch of reionization ([Cowie et al. 2017](#)). The latter result is in line

Table 2. Summary of Model Assumptions

Name	Equation	Description
(Eq 1) Galaxy Lumi- nosity Function	$\Phi(L, z) = \begin{cases} \Phi_*(z) \left(\frac{L}{L_*(z)} \right)^{\alpha_{\text{LF}}} (z) & : L < L_*(z) \\ \Phi_*(z) \left(\frac{L}{L_*(z)} \right)^{\beta_{\text{LF}}} (z) & : L \geq L_*(z) \end{cases}$	We adopt a luminosity function model that is a broken powerlaw with a faint-end slope, α_{LF} , a bright-end slope, β_{LF} , the characteristic luminosity at the knee of the luminosity function, L_* (given in L_\odot) and characteristic number density Φ_* (given in $\text{Mpc}^{-3} \text{dex}^{-1}$). In principle, all four parameters of the luminosity function (α_{LF} , β_{LF} , L_* , and Φ_*) can be redshift dependent.
(Casey 2012 Eq 3) Form of Dust SED	$S_\nu(T_{\text{dust}}) = C_1 \frac{(1 - e^{-\tau(\nu)})\nu^3}{e^{h\nu/kT_{\text{dust}}} - 1} + C_2 \nu^{-\alpha_{\text{MIR}}} e^{-(\nu_c/\nu)^2}$	Analytic approximation for a sources' flux density (in mJy) as a function of dust temperature (T_{dust}) and frequency (ν), in the form of a modified blackbody added to a mid-infrared powerlaw, following the methodology given in Casey (2012). Here, $\tau(\nu) = (\nu/\nu_0)^{\beta_E}$, where β_E is the spectral emissivity index, and $\nu_0 \approx 3 \text{ THz}$. The slope of the mid-infrared powerlaw is α_{MIR} , and the coefficients C_1 and C_2 are fixed with respect to one another, and set so that the integral under this curve between 8-1000 μm is L in L_\odot . ν_c is the frequency at which the powerlaw and modified blackbody contribute equally, and is a fixed function of T_{dust} , ν_0 , and β_E .
(Eq 2) λ_{peak} of SED	$\langle \lambda_{\text{peak}}(L) \rangle = \lambda_0 \left(\frac{L}{L_t} \right)^\eta$	The <i>median</i> rest-frame wavelength at which a dust SED will peak given its luminosity, L . We measure no significant redshift evolution in this relation beyond $z \sim 0.3$, and this is based on the observed empirical relationship shown in Figure 3. Here, λ_{peak} relates to the model's input dust temperature, T , via $\lambda_{\text{peak}} \approx b/T^{0.9}$, where $b = 2.898 \times 10^3 \mu\text{m K}$, Wien's Displacement Constant. Note this is an approximation and not exact (and not $=b/T$) because the opacity of the model shifts the peak of the SED towards longer wavelengths than the peak of a perfect blackbody (see Casey et al. 2014a, Figure 20). The λ_{peak} for any one galaxy is assigned assuming a Gaussian probability distribution in $\log_{10}(\langle \lambda_{\text{peak}} \rangle)$ with width σ . L_t is fixed to $10^{12} L_\odot$ and holds no physical meaning.
(Eq 3) T change due to CMB	$T'_{\text{dust}}(z) = ((T)^{4+\beta_E} + T_{\text{CMB}}^{4+\beta_E} [(1+z)^{4+\beta_E} - 1])^{\frac{1}{4+\beta_E}}$	Here T_{dust} is the intrinsic dust temperature of the galaxy as it would be at $z = 0$, i.e. the same as T_{dust} from Casey 2012 Eq 3 above. $T_{\text{CMB}}^{z=0} = 2.725 \text{ K}$, β_E is the emissivity spectral index, and $T_{\text{dust}}(z)$ is the adjusted temperature of the galaxy taking into consideration heating from the CMB. This temperature is then used to infer the fraction of flux at any frequency ν that would be observable at the given redshift.
(Eq 4) $S_\nu(T)$ change due to CMB	$f(z, T'_{\text{dust}}) = 1 - \frac{B_\nu[T_{\text{CMB}}(z)]}{B_\nu[T'_{\text{dust}}(z)]}$	The fraction of flux density S of $S_\nu(T_{\text{dust}})$ as given in Casey 2012 Eq 3, i.e. $S_{\text{obs}} = f(z, T'_{\text{dust}}) S_\nu(T_{\text{dust}})$, that would be detectable by an observer at frequency ν , redshift z , and adjusted dust temperature T_{dust} from Eq 3. B_ν is the Planck Function dependent on temperature. Here, the CMB temperature follows the redshift dependence $T_{\text{CMB}}(z) = T_{\text{CMB}}^{z=0} (1+z)$, where $T_{\text{CMB}}^{z=0} = 2.725 \text{ K}$.
(Eq 7) Clarifying Definitions	$\begin{aligned} x &\equiv \log_{10}(1+z) \\ x_t &\equiv \log_{10}(1+z_{\text{turn}}) \\ x_w &\equiv \frac{z_w}{\ln(10)(1+z_{\text{turn}})} \end{aligned}$	Simple definitions to clarify the evolution of $\log_{10}(L_*)$ and $\log_{10}(\Phi_*)$ in Eq 8 and Eq 9. z_{turn} is the adopted turnover redshift while z_w is the width in redshift over which the transition from one state to the other happens. x_t is a direct mapping of z_{turn} , while x_w is a mapping of z_w .
(Eq 8) L_* evolution	$\begin{aligned} \log_{10} L_*(x) = & \frac{(\gamma_2 - \gamma_1)x_w}{2\pi} \left[\ln \left(\cosh \left(\pi \frac{x - x_t}{x_w} \right) \right) \right. \\ & \left. - \ln \left(\cosh \left(-\pi \frac{x_t}{x_w} \right) \right) \right] \\ & + \frac{(\gamma_2 - \gamma_1)}{2} x + \log_{10}(L_0) \end{aligned}$	The evolution of the knee of the luminosity function L_* with redshift is assumed to evolve as $(1+z)^{\gamma_1}$, with a possible redshift turnover or 'reversal' happening at a redshift of z_{turn} such that at higher redshifts the relation evolves with a different slope, γ_2 .
(Eq 9) Φ_* evolution	$\begin{aligned} \log_{10} \Phi_*(x) = & \frac{(\psi_2 - \psi_1)x_w}{2\pi} \left[\ln \left(\cosh \left(\pi \frac{x - x_t}{x_w} \right) \right) \right. \\ & \left. - \ln \left(\cosh \left(-\pi \frac{x_t}{x_w} \right) \right) \right] \\ & + \frac{(\psi_2 - \psi_1)}{2} x + \log_{10}(\Phi_0) \end{aligned}$	The evolution of the characteristic number density of the luminosity function Φ_* with redshift is assumed to evolve as $(1+z)^{\psi_1}$, with a possible redshift turnover or 'reversal' happening at a redshift of z_{turn} such that at higher redshifts the relation evolves with a different slope, ψ_2 .

Table 3. Parameter Definitions and Adopted Values

NAME	DESCRIPTION	MODEL A	MODEL B	QUALITY OF CONSTRAINTS ^a
<i>– Luminosity Function Parameters –</i>				
L_0	Knee of the IR luminosity function at $z = 0$, in L_\odot .	1.3×10^{11}	1.3×10^{11}	SECURE
Φ_0	Characteristic Number Density of the IR luminosity function at $z = 0$, in $\text{Mpc}^{-3} \text{dex}^{-1}$.	3.2×10^{-4}	3.2×10^{-4}	SECURE
α_{LF}	Best-fit faint-end slope of the IR luminosity function from $z = 0$ to $z = 2.5$.	−0.6	−0.6	MINOR IMPACT
β_{LF}	Best-fit bright-end slope of the IR luminosity function from $z = 0$ to $z = 2.5$.	−3.0	−3.0	SECURE
<i>– Rest-Frame SED Parameters –</i>				
α_{MIR}	Mid-Infrared Powerlaw Slope.	2.0	2.0	SECURE
β_E	Emissivity Spectral Index.	1.8	1.8	SECURE
<i>– Peak of SED Parameters –</i>				
$\log \lambda_0$	λ_0 is the average rest-frame wavelength of S_ν at $L_t = 10^{12} L_\odot$.	2.012	2.012	SECURE
η	The slope of the $L_{\text{IR}}\text{--}\lambda_{\text{peak}}$ relation, as shown in Figure 3.	−0.068	−0.068	MINOR IMPACT
σ	Standard deviation of $\log(\lambda_{\text{peak}})$ at any given luminosity L .	0.045	0.045	MINOR IMPACT
<i>– Parameters describing Redshift Evolution –</i>				
γ_1	At $z \ll z_{\text{turn}}$, γ_1 describes the redshift evolution of L_\star , such that $L_\star \propto (1+z)^{\gamma_1}$	2.8	2.8	SECURE
γ_2	At $z \gg z_{\text{turn}}$, γ_2 describes the redshift evolution of L_\star , such that $L_\star \propto (1+z)^{\gamma_2}$	1.0	1.0	UNKNOWN
ψ_1	At $z \ll z_{\text{turn}}$, ψ_1 describes the redshift evolution of Φ_\star , such that $\Phi_\star \propto (1+z)^{\psi_1}$	0.0	0.0	SECURE
ψ_2	At $z \gg z_{\text{turn}}$, ψ_2 describes the redshift evolution of Φ_\star , such that $\Phi_\star \propto (1+z)^{\psi_2}$	−5.9	−2.5	UNKNOWN
z_{turn}	The ‘turning point’ redshift at which L_\star and Φ_\star are transitioning in their evolution.	2.1	1.8	SECURE
z_w	The redshift interval over which the evolution shifts exponents (e.g. γ_1 to γ_2).	2.0	2.0	SECURE

Notes. ^a We classify the level at which a parameter is already constrained by data in three classes: secure, minor impact or unknown. Secure means that the parameter is directly measurable with existing data. Minor impact means that the parameter is perhaps not very well known, but that changes to this variable (within reason) would not dramatically impact our measured results in this paper. Variables that are unknown are those which have no constraints. Appendix A expands on how well each of these parameters is known and how changes to their values impact the results. Parameters which are considered fixed as part of the SED (ν_c , C_1 and C_2) are fixed functions of α_{MIR} , T_{dust} , and L and therefore not given in this table. See Casey (2012) for details.

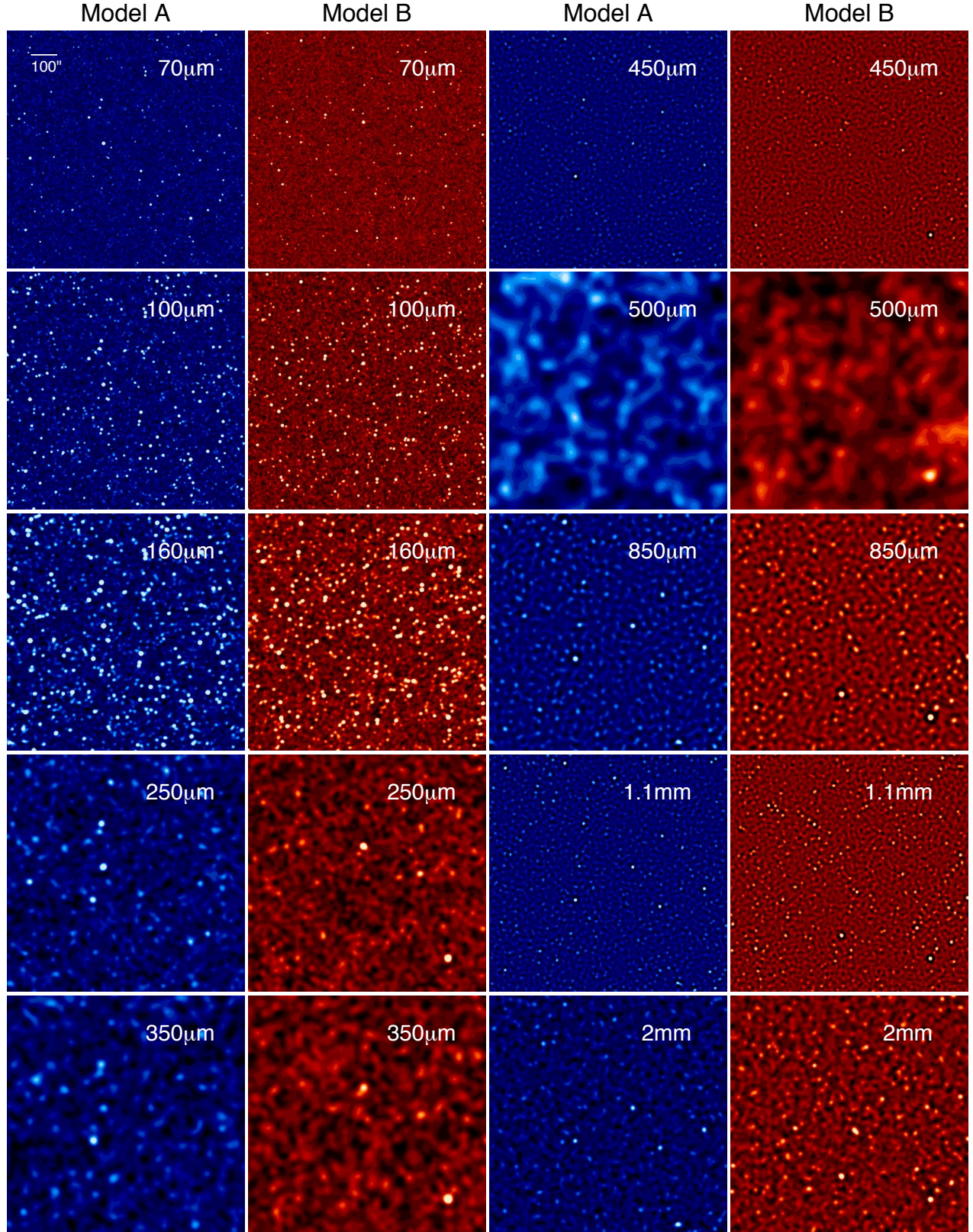


Figure 5. Simulated $900'' \times 900''$ signal-to-noise map cutouts of mock 1 deg^2 simulations, following the luminosity prescriptions described for Model A (the dust-poor early Universe) and Model B (the dust-rich early Universe). Our cutouts include mock *Herschel* PACS 70–160 μm , SPIRE 250–500 μm , SCUBA-2 450 μm and 850 μm , AzTEC 1.1 mm (with a 32 m diameter LMT) and GISMO 2.0 mm (with a 50 m diameter LMT). The hypothetical 1.4 mm TolTEC maps are not shown but are similar to the 1.1 mm and 2.0 mm maps.

with what might be expected from hierarchical formation and cosmic downsizing, suggesting a close relationship between the most massive starbursting galaxies, the supermassive black holes that grow at their centers, and the assertion that both live in some of the most overdense regions of the early Universe. A convergence of L_* values between IR luminosity functions and quasar luminosity functions, but a lack of agreement between number densities at those luminosities, hints at the possible factor of ~ 10 difference in the quasar and DSFG lifetimes. Below we discuss the physical context of our adopted values of ψ_2 for our case studies: models A and B, and we illustrate the differences in implied star-formation rate density between them in Figure 6. A more thorough discussion of alternate values of γ_2 are given in § A.1. It should be noted that both models are in agreement with the measured total energy output of the cosmic infrared background (CIB; Puget et al. 1996; Fixsen et al. 1998; Dwek et al. 1998). The CIB is dominated by sources at $z < 2$ by nature of the dominant source of emission at $\lambda \lesssim 500 \mu\text{m}$ (Viero et al. 2015), where the integrated background is unconstrained at longer wavelengths due to the brightness of the CMB.

2.8.1. Model A: A Dust-Poor Early Universe

Model A adopts $\psi_2 = -5.9$, suggesting a steep number density evolution for IR-luminous galaxies from the epoch of reionization to $z \sim 2$. In other words, this model suggests DSFGs are extremely rare in the early Universe ($z > 4$) such that their contribution to cosmic star-formation is negligible compared to much more numerous Lyman-break galaxies at the same epoch. Model A suggests DSFGs only come to dominate cosmic star-formation globally for a billion years or so near $z \sim 2$ and are very rare in both the early Universe and the local Universe. The adopted value of $\psi_2 = -5.9$ originates from the measured number density evolution of bright $M_{\text{UV}} \approx -21$ galaxies from $4 < z < 8$, which follows $\Phi \propto (1+z)^{-5.9}$ (Finkelstein et al. 2015). Adopting the same type of evolution for UV-luminous galaxies and IR-luminous galaxies might be quite appropriate if they occupy dark matter halos of similar masses and grow on similar timescales. Note that model A, or slight variants thereof, represent the currently accepted paradigm where the dust-formation timescale (primarily via AGB stars but also via supernovae) is longer than the formation timescale of the first UV-bright galaxies; it is often the adopted evolutionary scenario in the rest-frame UV literature (Bouwens et al. 2009, 2015, 2016; Finkelstein et al. 2015; Oesch et al. 2013).

2.8.2. Model B: A Dust-Rich Early Universe

Model B adopts $\psi_2 = -2.5$, asserting a gradual evolution in the number density of DSFGs from very high-redshifts up to $z \sim 2$. The implications of this value are that DSFGs would play an increasingly important role in cosmic star-formation towards higher redshifts, and this star-formation would be distributed far less ‘evenly’ than is suggested by rest-frame UV surveys. In other words, this model suggests that most of cosmic star-formation at early times was isolated to rare starbursts with very high star-formation rates, rather than more homogeneously distributed in lower-luminosity UV-bright galaxies. As shown in Figure 6, this model suggests that DSFGs would dominate cosmic star-formation (at roughly $\gtrsim 1/2$ of total) from $1.5 < z < 6.5$. Furthermore, at $z > 6.5$, DSFGs might dominate all star-formation by factors > 10 higher than UV-luminous galaxies.

Note that this model is designed to be somewhat extreme and thus provocative, because there has been no evidence to-date that DSFGs dominate cosmic star-formation at these very high redshifts. In fact, fewer than a dozen DSFGs have yet been found at these epochs ($z > 5$)! Despite this, we adopt this extreme case to illustrate a few points. The first is that such an extreme history has not yet been ruled out by existing surveys. The lack of DSFG identifications at high- z is, in large part, an observational limitation (as discussed extensively in Chapters 4 and 6 of Casey, Narayanan, & Cooray 2014a, also see e.g. Marrone et al. 2017). The second point worth illustrating – as the next section will detail – is that such dramatically different number density evolution at high- z (as constructed by Models A and B) result in largely the same observable characteristics in IR datasets from $70 \mu\text{m}$ – 1 mm , with only subtle differences. As we will show, the way of distinguishing between high-redshift DSFG models relies almost exclusively on the 1.4 mm and 2 mm bands, for which there is only sparse data existing to-date.

One obvious caveat to Model B is the lack of clarity on when the first DSFGs might turn on. This is not a problem in Model A because that model asserts that DSFGs are extremely rare at early times. In our implementation of model B we have not explicitly set a cutoff redshift above which DSFGs do not exist, though it is likely the case that such a limit exists in reality. We find that this has a relatively minor impact on the conclusions we reach for 1 deg^2 fields as the number of very high- z sources is not a dominant source of emission at any wavelength. However, it could impact results for $> 1 \text{ mm}$ surveys on larger scales ($\gg 1 \text{ deg}^2$), and a more

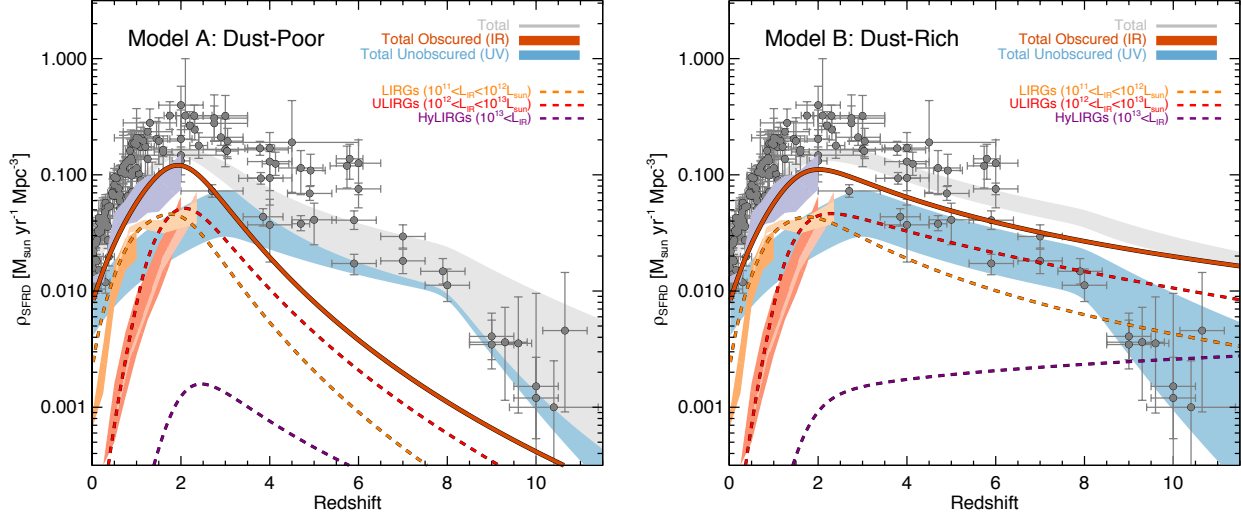


Figure 6. The implied star-formation rate densities for our two model universes. The thick lines denote the adopted models while the points and shaded regions represent existing measurements. Model A, the “dust-poor” model (described in § 2.8.1), asserts that the contribution of heavily obscured galaxies to cosmic star-formation is negligible in comparison to the contribution of UV-bright, unobscured galaxies at $z \gtrsim 4$. Model B, the “dust-rich” model (described in § 2.8.2), posits a very different history of cosmic star-formation, whereby dust-obscured galaxies dominate cosmic star-formation at $z \gtrsim 4$ over UV-bright galaxies by over a factor of ten. Measurements from the literature are shown as gray points (Madau & Dickinson 2014), though only UV-based measurements have been made beyond $z \approx 3.5$. The total contribution from unobscured sources is noted as a blue shaded region, the model IR contributions are shown as a solid orange line, and the total of these two components shown as a gray shaded region. The breakdown in contribution from LIRGs ($10^{11} < L_{\text{IR}} < 10^{12} L_{\odot}$), ULIRGs ($10^{12} < L_{\text{IR}} < 10^{13} L_{\odot}$), and HyLIRGs ($L_{\text{IR}} > 10^{13} L_{\odot}$) are shown as dashed gold, red, and purple lines, respectively. The light orange, red and purple shaded regions represent some of the best to-date measurements of the LIRG, ULIRG and total IR contributions to the SFRD (Murphy et al. 2011; Magnelli et al. 2011, 2013).

elaborate model could be implemented to account for this.

3. COMPARISON WITH DATA & MODELS

We quantify the appropriateness of models A and/or B using all available data constraints, including number counts, redshift distributions, and relative detection rates and measured flux densities of sources across the many IR/submm bands. We also describe alternate models that simulate the submm sky to provide some context in interpreting the dominant sources of emission and relative importance and constraints (or lack thereof) of galaxies in the early Universe.

3.1. Comparison with Alternate Models

The SIDES model (or the Simulated Infrared Dusty Extragalactic Sky; Bethérmin et al. 2017) is an update of the 2SFM (two star-formation modes Béthermin et al. 2012a; Sargent et al. 2012) galaxy evolution model to analyze the impact of clustering on IR map analysis. Our model differs from the 2SFM (Béthermin et al. 2012b) model at a fundamental level in its treatment of the underlying galaxy population: 2SFM builds galaxies’ SEDs from their position on the “main sequence” rela-

tion (the relationship between galaxies’ stellar mass and star-formation rate Noeske et al. 2007) and our model builds them up from their IR luminosity (or SFR) only. This difference would not necessarily result in a discrepancy, but the former assumes galaxies SEDs are linked more fundamentally to their specific SFR in two modes, either on the main sequence, or in a starbursting phase; the SIDES model has slightly warmer dust in $z > 2$ galaxies than the 2SFM model. Galaxies with higher specific SFR are asserted to have much warmer dust SEDs which is traced to harder radiation fields ($\langle U \rangle$). In this paper, we argue against a bi-modal population, and that luminosity, or SFR alone, is more fundamentally linked to a galaxy’s SED (and thus the nature of their IR flux densities) than their specific SFR. When separated by stellar mass, for example, there is little evidence that galaxies sitting on the high-mass end of the main sequence have colder temperatures than those at similar SFRs that are elevated above the main sequence at lower masses (U et al. 2012), and instead, their dust temperatures track very tightly with IR luminosity.

Beyond the scope of 2SFM, SIDES (Bethérmin et al. 2017) incorporates clustering and its effects in source multiplicity by using abundance matching to populate

dark matter halos in a dark-matter only large volume simulation using stellar mass abundance matching. Then, those galaxies are represented by IR SEDs according to the 2SFM model in order to explain discrepancies between high- and low-resolution IR number counts. This is not within the scope of this paper, though we do incorporate the effects of angular resolution. Our model does not account for source clustering. This makes it difficult to characterize the entire nature of DSFG multiples (sources that break into several components with higher resolution, and whether or not those multiples are physically associated, e.g. Hodge et al. 2013; Hayward et al. 2013; Hill et al. 2017); though it would, in principle be possible to estimate the multiple fraction of DSFGs caused by chance projections in this model, it is beyond the scope of this paper. Of course the issue of multiplicity is important for our understanding of the physical drivers of DSFGs and thus provides important insight which the SIDES model addresses.

At a more fundamental level, the luminosity function in SIDES/2SFM is fixed at high-redshift according to the evolution of Φ_* modeled in Sargent et al. (2012), which leans heavily on galaxy luminosity functions measured at wavelengths shortward of $24\mu\text{m}$. Specifically, at high-redshifts, the Béthermin et al. (2012a) model assumes the IR contribution to the SFRD follows the shape of the measured rest-frame UV evolution to $z \sim 6$, that the main sequence (or galaxies' specific star-formation rates) is fixed at $z > 2.5$, and that the decreasing contribution of IR luminous galaxies at very early times is due entirely to increased rarity (i.e. a drop in Φ_*). The updated SIDES model has a similar drop at the highest redshift despite continued luminosity function at $z > 2.5$ (following Schreiber et al. 2015), which is a direct consequence of the evolving stellar mass function in the model (Davidzon et al. 2017). In this manner, the SIDES model follows the high-redshift evolution of our Model A, the dust-poor early Universe. Because the goals of our work are to place constraints on the evolution of the IR luminosity function itself, our model is, and needs to be, constructed in a very different way to provide insight for quantities that are otherwise fixed in SIDES.

We also compare redshift distributions with the Zavala et al. (2014) model, which focuses exclusively on reproducing (sub)mm galaxy redshift distributions from a single underlying high- z population that is detected at 1.1 mm, but does not go so far as to model the underlying IR luminosity function.

3.2. Number Counts Comparison

The data used as a baseline for comparison come from the existing rich literature of FIR/submm number counts. Specifically, we draw on datasets at $70\mu\text{m}$ (Dole et al. 2004; Béthermin et al. 2010; Berta et al. 2011), $100\mu\text{m}$ (Héraudeau et al. 2004; Rodighiero & Franceschini 2004; Kawara et al. 2004; Berta et al. 2011; Magnelli et al. 2013), $160\mu\text{m}$ (Dole et al. 2004; Kawara et al. 2004; Béthermin et al. 2010; Berta et al. 2011; Magnelli et al. 2013), $250\mu\text{m}$, $350\mu\text{m}$ and $500\mu\text{m}$ (Patanchon et al. 2009; Oliver et al. 2010; Clements et al. 2010; Béthermin et al. 2010, 2012a), $450\mu\text{m}$ (Smail et al. 2002; Geach et al. 2013; Casey et al. 2013; Chen et al. 2013; Wang et al. 2017; Zavala et al. 2017), $850\mu\text{m}$ (Blain et al. 1999; Scott et al. 2002; Chapman et al. 2002; Smail et al. 2002; Cowie et al. 2002; Borys et al. 2003; Webb et al. 2003; Barnard et al. 2004; Coppin et al. 2006; Scott et al. 2006; Knudsen et al. 2008; Beelen et al. 2008; Weiß et al. 2009; Karim et al. 2013; Casey et al. 2013; Chen et al. 2013; Simpson et al. 2015; Geach et al. 2016), and 1.1 mm (Perera et al. 2008; Austermann et al. 2010; Scott et al. 2010; Hatsukade et al. 2011; Aretxaga et al. 2011; Scott et al. 2012; Oteo et al. 2016; Aravena et al. 2016; Fujimoto et al. 2016). Though there have been some initial estimates of number counts at 2 mm (e.g. Staguhn et al. 2014), they are not yet robust enough to place useful constraints on our model. Nevertheless, we include them for a qualitative comparison.

Figure 7 shows differential number counts from $70\mu\text{m}$ –2 mm taken from the above literature measurements against both of our models' number counts output, as well as the output from the SIDES model. There is generally good order-of-magnitude agreement of all models in all bands, despite the very different model assumptions on which each is built.

At shorter wavelengths, $\lambda_{\text{obs}} < 500\mu\text{m}$, the output of our two case studies (blue as model A and orange as model B) are indistinguishable from one-another. This is largely due to the fact that all emission at these bands comes from galaxies at $z < 2$ where our models are identical. In both cases, the *Herschel* PACS wavelengths, $70\mu\text{m}$ – $160\mu\text{m}$, have required the adjustment for AGN contribution to the mid-infrared powerlaw, as discussed in § 2.8. No such adjustment is needed for the mock *Herschel* SPIRE bands or ground-based submm wavelengths where AGN are not a dominant emission mechanism.

At wavelengths beyond $\lambda > 200\mu\text{m}$, both of our models match the galaxy number counts well, though as selection wavelength increases, the separation between Model A and Model B become more distinct, with Model A systematically providing lower number count predictions than Model B. Both $850\mu\text{m}$ and 1.1 mm provide

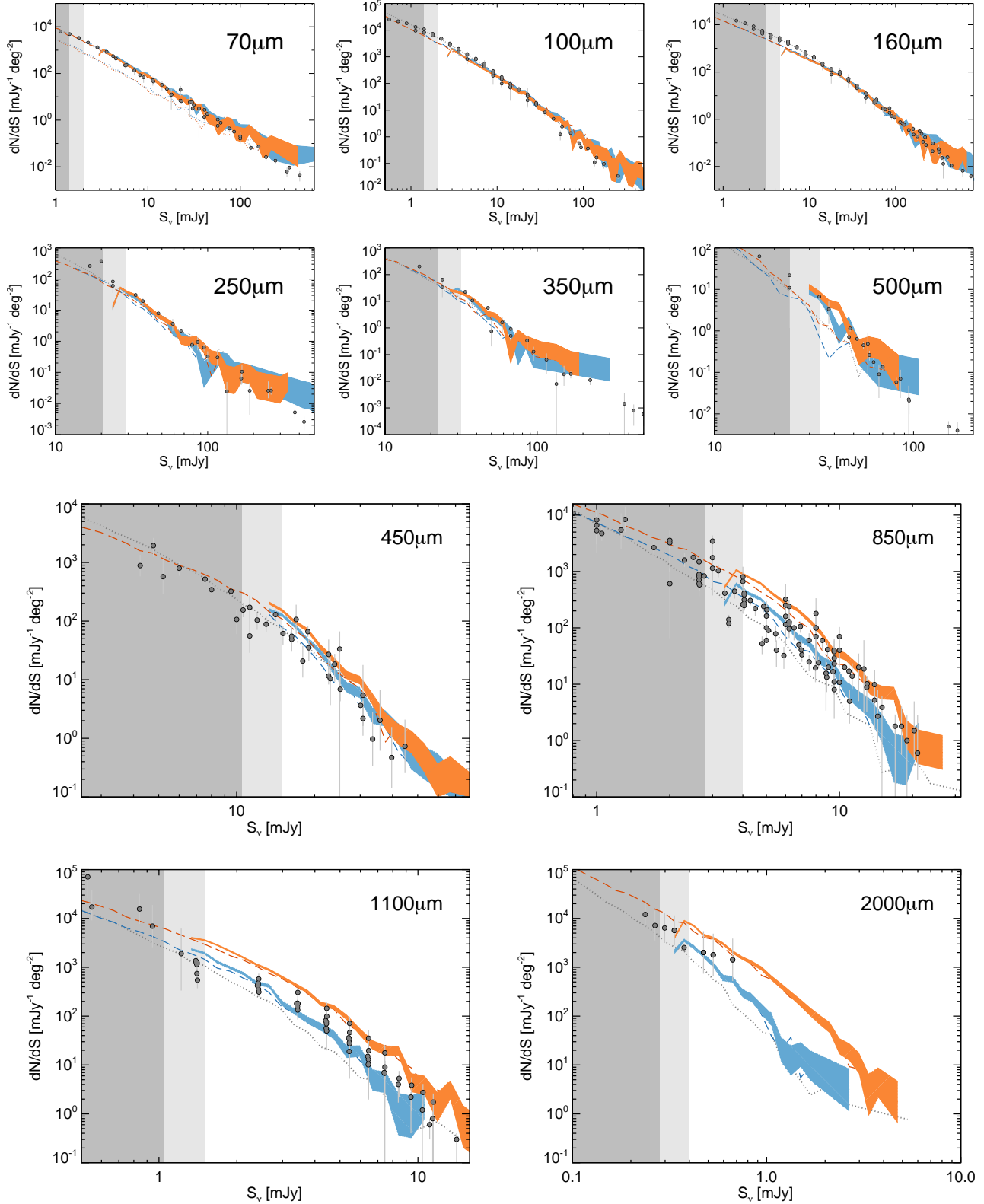


Figure 7. The resulting differential number counts of our two case-study models from $70\mu\text{m}$ – 2mm . Data points (gray) from the literature are pooled from many sources referenced in the text and summarized in Casey et al. (2014a). The results of the SIDES model (Bethérmin et al. 2017) are shown as a dotted gray line while the results of the dust-poor model (Model A) are shown in blue, and the dust-rich model (Model B) are shown in orange. The injected source counts are the darker, long-dashed curves without uncertainty. The extracted source counts, after degrading to the beamsizes and RMS of typical observations, are shown as the lighter blue and orange, with shaded uncertainty as measured from a 1 deg^2 simulated map. Dark gray background denotes flux densities at $<3.5\sigma$ for our simulation while light gray background denotes flux densities at $3.5 < \sigma < 5.0$. The $70\mu\text{m}$ panel also shows the model predictions without the included AGN component (dotted blue and orange lines).

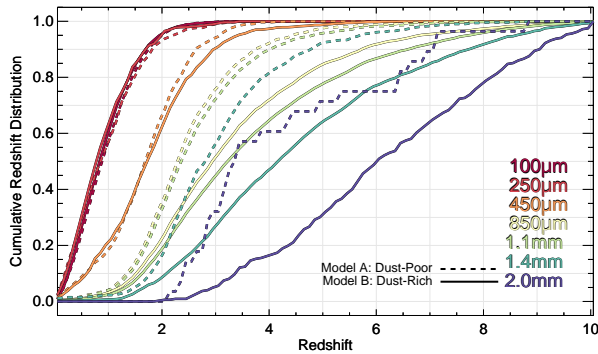


Figure 8. The predictive cumulative redshift distribution of modeled selection wavelengths for our two case studies. Predicted distributions for a dust-poor Universe (Model A) are shown as dashed lines and a dust-rich Universe (Model B) are shown as solid lines. The sources included on this plot are only those detected above a 5σ detection threshold in each of our 1 deg^2 simulated maps, with noise and beamsizes characteristics given in Table 1. Median redshift for the samples track with selection wavelength, and the two models become sufficiently distinct at selection wavelengths long-ward of $\sim 1\text{ mm}$.

the most discriminating power between the models while still having sufficient data available to constrain number counts. However, the spread on data measurements is somewhat extraordinary, varying by up to an order of magnitude for $\sim\text{mJy}$ sources. Both Model A and B predict $850\mu\text{m}$ and 1.1 mm number counts well within the constraints of existing data (though perhaps model A is slightly favored at 1.1 mm if you consider the extension towards the faint end, sub-mJy sources). The measurements of 2 mm number counts to-date are too few to meaningfully comment on which model is favored; the GISMO Deep Field (Staguhn et al. 2014) only contains seven sources due to its small size. However, it is clear that of all the wavelength regimes, 2 mm has the most discriminating power between hypothetical high- z models and should be a priority for future observational efforts to constrain the high- z IR luminosity function.

We note that at these long wavelengths, the SIDES simulation (dotted line) underpredicts the number density of $\sim 1\text{ mJy}$ $850\mu\text{m}$ – 1.1 mm sources in comparison to both our Models A and B and literature measurements. We attribute this to their adoption of a UV-like L_\star and Φ_\star evolution towards the highest-redshift epochs.

3.3. Redshift Distributions

Due to the very negative K-correction in the submillimeter (Casey, Narayanan, & Cooray 2014a), sources’ flux densities are largely independent of redshift, and so redshift distributions break down another important dimension of our mock data, giving an independent mea-

surement of the quality of our models that is distinct from the comparison of number counts. Redshifts for extracted sources are determined by first identifying all possible sources within a beamsize (as stated in Table 1) that could contribute flux to a given identified source in the output map. The injected source that contributed the most input flux density at the selection wavelength within the beamsize is then marked as the primary source. This method is imperfect, as it assumes that the redshift of the brightest source within a beam provides the best representation of a given detected source, and it neglects the impact of source multiplicity caused by line-of-sight projections (Hayward et al. 2013). However, we find that this method provides a fair representation of the statistical populations of galaxies detected at each of the selection wavelengths (in particular the observational setups with higher angular resolution). This method is also most analogous to the observational methods used in the majority of redshift survey work carried out to-date on single-dish submm galaxy populations (Chapman et al. 2003b, 2005; Swinbank et al. 2004; Wardlow et al. 2011; Casey et al. 2012a,b, 2017; Danielson et al. 2017).

Figure 8 illustrates the predicted cumulative redshift distributions from $70\mu\text{m}$ – 2 mm for models A (dashed lines) and B (solid lines). The importance of the long-wavelength regime ($\lambda_{\text{obs}} \gtrsim 1\text{ mm}$) for picking out high- z DSFGs becomes quite clear on this plot. As might be self-evident, a dust-poor early Universe should have far fewer high-redshift detectable sources than a dust-rich early Universe, and so Model A’s results are more skewed toward lower-redshift distributions. For example, at $870\mu\text{m}$ (the selection wavelength of the ALESS sample, one of the best-studied, uniformly-selected samples DSFGs in the literature; Hodge et al. 2013; Danielson et al. 2017), the predicted median redshift for the brightest, most robust subset ($S_{870} > 3.5\text{ mJy}$) is $\langle z_{870} \rangle = 2.4$ for Model A vs $\langle z_{870} \rangle = 3.1$ for Model B. The predicted median redshift for a $S_{1.1\text{ mm}} > 1.5\text{ mJy}$ sample is $\langle z_{1.1} \rangle = 2.4$ for Model A and $\langle z_{1.1} \rangle = 3.4$ for Model B.

Intriguingly, measurements of redshift distributions from the literature do not completely favor model A and rule out model B. The direct comparisons of redshift distributions for measured samples (at matching flux density cuts) is shown in Figure 9. The three data samples given are the most robust, most complete subsamples of spectroscopically-confirmed DSFGs in the literature. At $870\mu\text{m}$, we compare against the LABOCA-selected (Weiß et al. 2009), ALMA followed-up (Hodge et al. 2013) sample whose redshift survey is described in Danielson et al. (2017). At 1.1 mm , we compare against

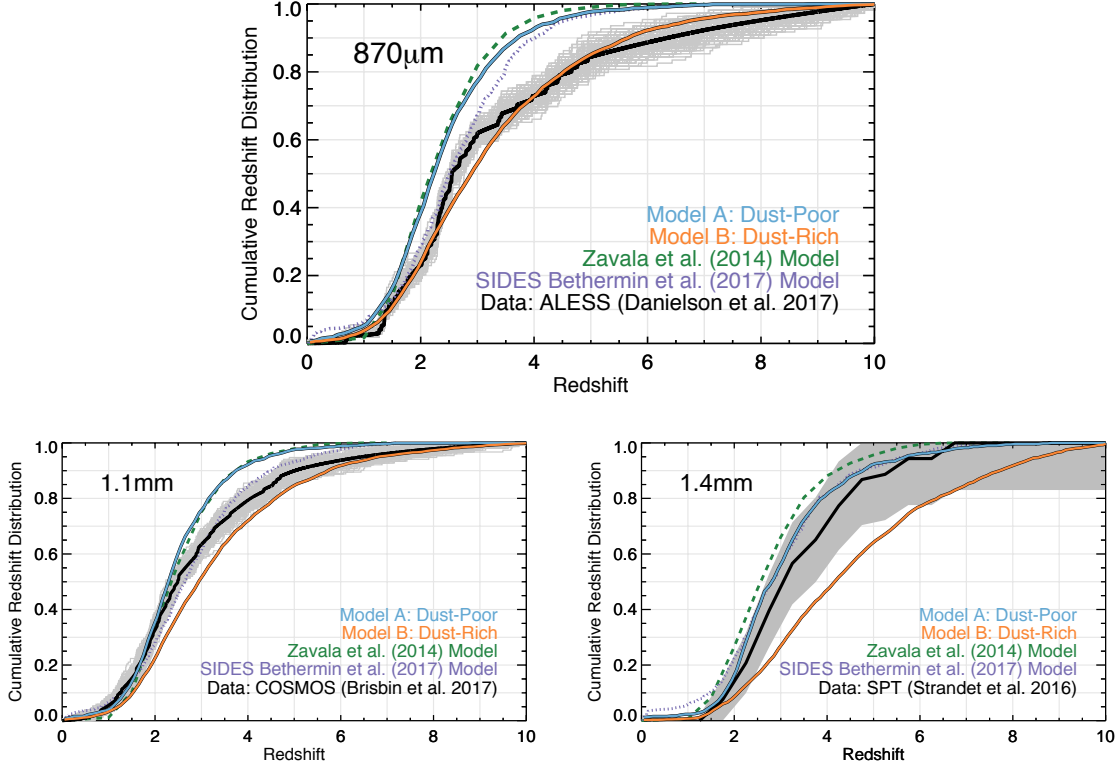


Figure 9. Direct comparisons of our model-predicted redshift distributions and data. Here we compare against three independent and robust datasets on redshifts of (sub)millimeter galaxies that are the most complete to-date: the ALESS spectroscopic survey of LABOCA-selected $870\mu\text{m}$ sources [Danielson et al. \(2017\)](#), the combined spectroscopic and photometric survey of AzTEC/ALMA-followed up sources in COSMOS at 1.1 mm from [Brisbin et al. \(2017\)](#), and the South Pole Telescope 1.4 mm-selected SMGs described in [Strandet et al. \(2016\)](#). Gray regions denote uncertainties in the measured distributions. Model A is denoted by the blue line and Model B is the orange line (note that these curves are not identical to those in Figure 8 due to different flux density limits). The dashed green line shows a comparison to the [Zavala et al. \(2014\)](#) model redshift distribution, and the predictions of SIDES ([Bethérmin et al. 2017](#)) are shown as a dotted purple line.

AzTEC-selected ([Aretxaga et al. 2011](#)) sources in the COSMOS field that have ALMA or other interferometric followed-up ([Smolčić et al. 2012](#); [Brisbin et al. 2017](#)). Both the $870\mu\text{m}$ and 1.1 mm samples are hybrid combinations of spectroscopic and photometric redshifts, with a few sources lacking redshifts completely. To accurately model the uncertainty on the aggregate redshift distribution for the whole sample, we assign each source a probability density function in redshift according to measurement uncertainties (either spectroscopic or photometric); sources without constraints are assumed to sit at $z > 1$ with probability density function mirroring a step-function. We then generate many realizations of the measured cumulative redshift distribution by drawing from each source’s individual probability density function.

The last data sample we use for comparison is the SPT, 1.4 mm-selected sample of lensed SMGs discussed in [Strandet et al. \(2016\)](#). While this sample is the most spectroscopically complete of any DSFG sample, it has

the added complication in that relatively small and it is almost entirely comprised of strongly-lensed systems, which is a natural consequence of its high flux density cut, $S_{1.4} > 20 \text{ mJy}$ ([Weiß et al. 2013](#)). The measured median magnification factor for the SPT sample is $\langle\mu\rangle \approx 6$ ([Hezaveh et al. 2013](#); [Spilker et al. 2016](#)), and so the equivalent unlensed flux density limit would be around $S_{1.4} \gtrsim 3.3 \text{ mJy}$. To mimic this selection in our models, we select 1.4 mm-detected sources assuming a $6.9''$ beamsize FWHM above a flux density of 2 mJy (which accounts for additional uncertainty in the magnification factor). The SPT sample also has a redshift bias that excludes sources at $z \ll 1.5$ ([Hezaveh & Holder 2011](#)) due to the low probability of galaxy-galaxy lenses. While this second effect could be substantial, the lack of $z < 1.5$ sources in our 1.4 mm-selected sample to begin with implies this effect is negligible.

At all three wavelengths, we also include comparisons to the SIDES model output and the model redshift distributions of [Zavala et al. \(2014\)](#). While none of the

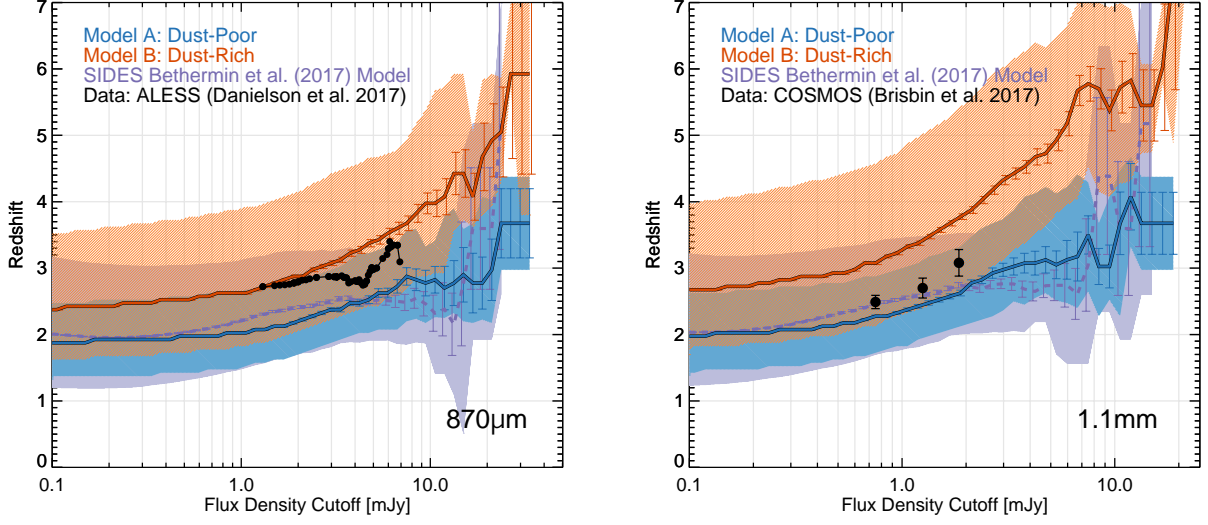


Figure 10. The dependence on flux density cutoff and that galaxy sample’s median redshift for our Model A (blue), Model B (orange), and the SIDES model (purple). The shaded regions enclose the inner 50% of sources from each model, and the bootstrap-measured uncertainty on the median is shown via the colored error bars; these become dominant in the flux density regime where there are very few sources per square degree. We compare to data from both ALESS (Danielson et al. 2017) and COSMOS (Brisbin et al. 2017) at $870\mu\text{m}$ and 1.1mm , respectively.

models – including those from this paper – provide perfect matches to the redshift distributions found by the data, the comparison is illuminating as to the nature of high- z DSFG prevalence. For example, the measured median redshift for ALESS $870\mu\text{m}$ -selected galaxies is $\langle z_{870} \rangle = 2.4 \pm 0.1$, or $2.6^{+0.2}_{-0.1}$ after accounting for unconfirmed sources and likely high- z solutions for those sources. Both the Zavala et al. and our dust-poor Model A have median redshifts of 2.2 falling short of the median measured redshift. While the median of SIDES output is the closest to data, the SIDES model fails to catch the measured high-redshift tail of the population. Model B, while overestimating the median of the full sample, accurately catches this high-redshift tail of the population. A similar phenomenon is seen in the comparisons of data and models at 1.1mm , though more subtle. As pointed out in Brisbin et al. (2017), the measured redshift distribution at 1.1mm is very dependent on the exact flux cutoff (they measure $\langle z \rangle = 2.18 \pm 0.09$ for sources with $S_{1.2} < 1.25\text{mJy}$ and $\langle z \rangle = 3.08 \pm 0.17$ for sources with $S_{1.2} > 1.8\text{mJy}$). Both the ALESS and COSMOS samples exhibit significant high- z tails, with 23% of $870\mu\text{m}$ sources above $z > 3$ and 27% of 1.1mm sources above $z > 3$. At 1.4mm , the small number statistics on the measured redshift distribution implies more uncertainty and difficulty in distinguishing between favored models, though model A is slightly more favored (with model B in less than 1σ tension with data).

Figure 10 shows a more detailed breakdown of the flux-density dependence on samples’ median redshifts.

As a function of the lower-limit flux cutoff of a survey, we compute the median redshift for all galaxies in our model detected above that threshold, and also compare against the SIDES model input. The Danielson et al. and Brisbin et al. samples are used for comparison. The interpretation of this analysis is not straightforward: the data are in slight tension with the SIDES output and fall between our two extreme model case studies, likely pointing to a true value of $-5.9 < \psi_2 < -2.5$. We wish to highlight that these results, instead of adequately distinguishing between conflicting models, fail to rule out even extreme models for the prevalence of DSFGs in the early Universe. We discuss a possible path forward with 2mm observations in § 4.3.

3.4. Correlation of Flux Densities across Bands

Beyond number counts and redshift distributions, we can also compare flux densities across (sub)mm bands by measuring their relative correlation. For example, are sources’ detected at $850\mu\text{m}$ also detected at $450\mu\text{m}$, and are their flux densities proportional to one another? To measure such a correlation for any two bands, we construct lists of corresponding flux densities from the union of detected sources in either band. For sources detected in both bands, the extracted peak flux densities in each band is recorded with a positional match accuracy less than the smallest beamsize; for single-band detections, we measure the flux density in the other band at the position of the extracted source. From this sample, we measure the Pearson’s correlation coefficient $\rho(\lambda_1, \lambda_2)$

such that:

$$\rho(\lambda_1, \lambda_2) = \frac{\text{cov}(\lambda_1, \lambda_2)}{\sigma_{\lambda_1} \sigma_{\lambda_2}} \quad (10)$$

where $\text{cov}(\lambda_1, \lambda_2)$ is the covariance of measured flux densities at observed λ_1 and λ_2 , and σ_λ is the standard deviation of flux densities for either sample. Values near one represent perfect correlation, zero represent no correlation, and negative one represents perfect anti-correlation. One could imagine that a universe with very diverse dust SEDs could result in lower cross-band correlation than a universe where galaxies' dust SEDs are remarkably similar. In this sense, measuring the correlation across bands in real data samples and in models gives a measure of how homogeneous or diverse galaxies' dust SEDs are. For the models, this makes use of extracted sources only, after maps have been degraded to the spatial resolution and instrumental noise of real observations.

Extracting such correlations from real data is somewhat challenging for large samples, since very few regions of the sky have been mapped deeply in many bands across the far-infrared/submm. For its superb multi-wavelength coverage, we draw on maps already in hand in the COSMOS field, from *Herschel* PACS, SPIRE, SCUBA-2, and AzTEC data. We use the compilation of data from Lee et al. (2013), Casey et al. (2013) and Aretxaga et al. (2011), in addition, drawing on the SCUBA-2 data for the EGS described in Zavala et al. (2017).

Figure 11 shows the results of our comparison of flux densities in each band pairing for which we have adequate data. We compare the correlation of the same band fluxes from the SIDES model output catalog (Bethérmin et al. 2017), and both the output from our Model A, the dust-poor model, and Model B, the dust-rich model. We note that the predictions of SIDES suggest significantly more correlation across bands than exists in measured maps. The output of our simulations, both A and B, are less correlated across bands than SIDES. In some cases, both of our models overestimate the correlation (at short wavelengths $< 250 \mu\text{m}$), likely due to the lack of inclusion of an array of mid-infrared powerlaw slopes. At longer wavelengths, our models are more aligned with observed correlation. Because this measured correlation coefficient is a tracer of SED diversity and not particularly sensitive to source number density, both Models A and B largely follow the same trends in cross-band correlation with the exception of the longest wavelength bands, where we see the divergence of model predictions between Model A and Model B predominantly because of the growing dominance of higher-redshift sources in Model B at $\lambda \gtrsim 850 \mu\text{m}$.

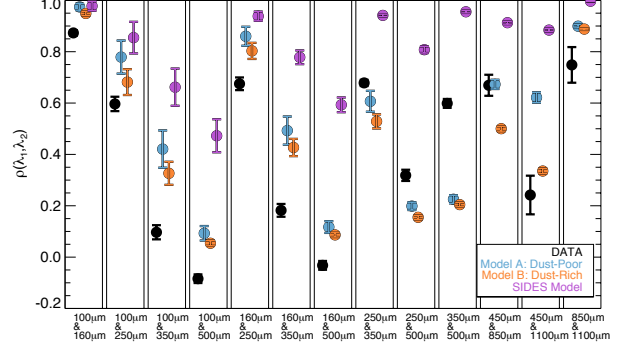


Figure 11. The correlation coefficient, $\rho(\lambda_1, \lambda_2)$ between flux densities of detected sources at λ_1 and at λ_2 from real data in the COSMOS field (black points; Aretxaga et al. 2011; Lee et al. 2013; Casey et al. 2013), from the output of our models: model A (blue) and model B (orange), as well as the SIDES model (purple; Bethérmin et al. 2017). Values close to one indicate strong correlation between sources' measured flux densities in the two bands, while zero represents no correlation, and a negative value would indicate anti-correlation. We find that SIDES is generally much more correlated across all bands than our simulations as well as real data.

4. DISCUSSION

This paper presents a model to interpret observations of the IR/submm sky. With mock observations we have compared the results of two case-study models with extremely different assumptions about the shape and evolution of the IR luminosity function beyond $z > 2$. Model A has assumed the early Universe is dust-poor, with DSFGs contributing very little to cosmic star-formation at $z > 4$, while Model B has assumed the opposite: that DSFGs are so dominant at $z > 4$ that they render the contribution from LBGs negligible. The physical implications of these models differ wildly.

4.1. Broad Implications for DSFGs in the Early Universe

In the case of Model A, star formation at early times is dominated by UV-luminous LBGs, and dust-rich systems may be exceedingly rare. This model is aligned with some literature results that suggest a diminished role of dust-enrichment in 'normal' galaxies at $z \gtrsim 5$ (Capak et al. 2015; Bouwens et al. 2016). This model could also be in agreement with the relatively few DSFGs discovered above $z > 5$ to-date, though their volume density has not been directly constrained (Strandet et al. 2017). Furthermore, the measurements of the star-formation rate density of the Universe from deep optical and near-infrared measurements, as shown in Figure 1, would hold as universally true to all galaxy types.

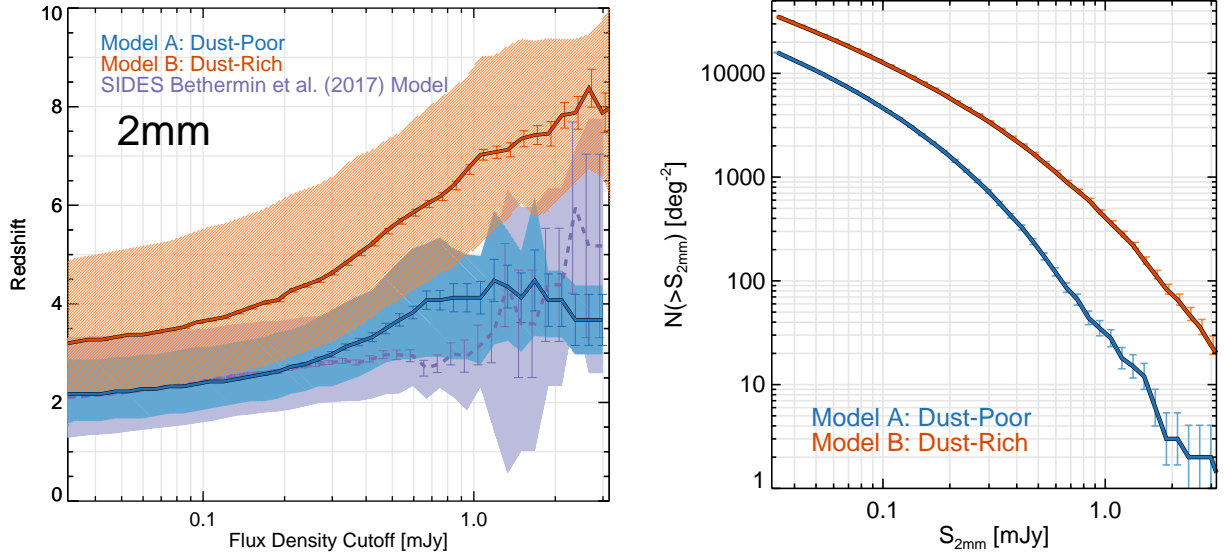


Figure 12. The combined predicted characteristics of future 2 mm DSFG surveys. At left, we show the flux-dependent redshift distributions (as shown in Figure 10 at other wavelengths) between model A (dust-poor; blue) and model B (dust-rich; orange). The most dramatic difference in median redshifts by model are for sources with flux densities 0.3–1 mJy. At right, we show the predicted cumulative number counts at 2 mm for both models, which should be easily distinguishable in surveys $\sim 0.1 \text{ deg}^2$ down to flux densities of $\sim 0.5 \text{ mJy}$ (at 5σ).

In the case of model B, DSFGs would be far more prevalent in the early Universe, dominating cosmic star-formation ($>90\%$ of the total). This would physically imply that dust production mechanisms post-Big Bang would need to be particularly efficient and likely form via supernovae (Matsuura et al. 2011; Dwek et al. 2014) combined with low destruction rates, rather than from coagulation in the upper atmospheric winds of AGB stars, or coagulation or accretion of dust in the ISM (Matsuura et al. 2006, 2009; Jones et al. 2013). Recent simulations of the first stellar production of metals could produce some dust grains as early as $z \sim 25$, though it appears the dust at that epoch does not significantly impact the observable characteristics of typical galaxies (Jaacks et al. 2018). It is yet unclear, from a simulator’s perspective, how early Universe DSFGs might form rapidly. Furthermore, DSFGs tend to be quite massive galaxies, so their formation on short $<1 \text{ Gyr}$ timescales may be needed to directly constrain cosmological models of massive halo growth (Wechsler et al. 2002). If Model B were correct, it would mean that most of early Universe star-formation has not been accounted for in the census of galaxies at $z > 5$. DSFGs at this epoch would be entirely obscured, in stark contrast to the population of LBGs seen at the same epoch which appear to be much more dust-poor than their analogues at lower redshift.

Current datasets do not clearly rule out either of these two simple yet extreme models. Though number counts are largely invariant with model assumptions and lack constraining power (because at most wavelengths they are dominated by sources at $z < 2$, Lagache et al. 2004), it is the deficit of sources at long wavelengths ($\lambda \gtrsim 850 \mu\text{m}$) in the dust-poor model A and SIDES model that hint towards a possible higher prevalence of DSFGs at $z > 2$ as present in Model B. Unfortunately, direct sample measurements of the IRLF from well characterized DSFGs only confirm that constraints peter out beyond $z \sim 2.5$, largely due to the lack of completeness (at $\lambda \lesssim 850 \mu\text{m}$). Measured galaxy redshift distributions at $870 \mu\text{m}$ – 1.4 mm provide the cleanest contrast between dust-poor and dust-rich models, while neither model is clearly favored; where some measures (median redshift) might favor model A, other measures (the distribution of sources in the highest-redshift tail) might favor model B. It is also clear that any sample incompleteness in redshift surveys would severely hamper our ability to draw meaningful conclusions, and most redshift surveys of DSFGs are far too incomplete to be of use (i.e. the gray regions on Figure 9, already large for plotted samples, would be much larger for any samples suffering from more incompleteness). In that sense, current datasets have very little discriminating power even with extremely different input assumptions.

Though our model provides some broad context for interpretation of high- z DSFGs, it does nothing to explain the physical origins of such systems. As discussed extensively in Chapter 10 of Casey, Narayanan, & Cooray (2014a), cosmological simulations of DSFGs are especially challenging, both from the perspectives of semi-analytic model (SAM) N-body simulations (e.g. Baugh et al. 2005; Lacey et al. 2008; Somerville et al. 2012; Hayward et al. 2013) and hydrodynamic cosmological simulations (e.g. Springel 2005; Vogelsberger et al. 2014a). They are particularly challenging due to the high computational cost of dust radiative transfer (required for shorter wavelengths where dust emission is not optically-thin) in a cosmological context. In the case of SAMs, dust prescriptions are analytically described from galaxies’ halo characteristics and no direct modeling of ISM processes is included; there are few calibrations against observational data to refine these dust prescriptions, and we know they under-predict the number of DSFGs across most redshift regimes. Cosmological hydrodynamic models, like the recent Illustris and Illustris TNG models (Vogelsberger et al. 2014a,b, 2018; Weinberger et al. 2017) have still not been able to directly implement dust prescriptions. One hydrodynamic simulation that uses smooth particle hydrodynamics (SPH) and does model dust emission by linking it to metal abundances is the BLUETIDES simulation (Wilkins et al. 2017a,b); to date, BLUETIDES has run down to $z = 8$ with a hopeful expansion down to $z = 6$ in the near future.

Wilkins et al. (2017a) presents results on the $z \geq 8$ dust-obscured galaxy population from BLUETIDES, and remarkably come to a conclusion that would largely be aligned with our Model B, the dust-rich Universe. They find that approximately 90% of star-formation in high-mass galaxies $> 10^{10} M_{\odot}$ at $z > 8$ is already obscured by dust. This is in-line with mass-dependent obscuration fractions measured at much later epochs (Whitaker et al. 2017), as well as some high- z work on the mass-dependent dust content of UV-selected galaxies, which is found to be constant from $4 < z < 7$ (Finkelstein et al. 2012, also see Bowler et al. 2018). It counters the argument that galaxies are less dust-rich beyond $z > 4$ (Bouwens et al. 2016). Wilkins et al. highlight the importance of pushing submm surveys into new parameter space, in the direction of directly constraining the prevalence of DSFGs at such high-redshifts.

4.2. Impact of Individual Model Assumptions

Though we caution that neither model A or model B is clearly favored by existing datasets, suggesting that model B is not cleanly ruled out is a somewhat bold claim, and should be immediately followed-up with an

analysis of each individual parameter’s impact on that conclusion. For example, if we were to assume a different history to the evolution of dust SEDs, how would the conclusion that DSFGs are prominent at $z > 4$ possibly change? We explore some implications here with more detailed analysis presented in Appendix A.

Though we understand the shape and diversity of dust-rich galaxies’ IR SEDs well from $0 < z < 2$, there could be reason to doubt our SED model assumptions hold for galaxies that formed in the first few billion years after the Big Bang. For example, early Universe galaxies are likely far less metal-enriched than those at low-redshift, and they might have had a different (possibly top-heavy) initial mass function (Baugh et al. 2005; Swinbank et al. 2008). These effects of metal-enrichment and possible varying IMF (linked to formation of stars in super-star clusters, SSCs; McKee & Tan 2003; Bastian et al. 2010; Portegies Zwart et al. 2010) may result in very different geometric distribution of dust and gas in the ISM of the galaxies. It is this geometry – densely-packed dust or diffuse dust, patchy or smooth – that directly impacts the mass of dust heated to different temperatures, which in turn, shapes the galaxy’s IR SED. It could be that the more primordial dust-rich galaxies were much more densely-packed, hot-dust dominated than the mature, massive and extended systems seen at $z \sim 2$.

If the $z \gtrsim 4$ DSFGs are intrinsically hotter than DSFGs at $z \sim 2$ (as explicitly suggested by Faisst et al. 2017), then we may expect the $L_{\text{IR}}-\lambda_{\text{peak}}$ relationship at $z \sim 5$ to evolve significantly away from the one for $z \sim 2$. This would translate to SEDs shifted to shorter rest-frame wavelengths, and a reduction of flux for wavebands still probing the Rayleigh-Jeans tail, at $\lambda_{\text{obs}} > 1 \text{ mm}$. Implementing this shift into our model explicitly results in the number counts of both model A and B to shift downward by ≈ 0.3 dex at $\lambda_{\text{obs}} \gtrsim 850 \mu\text{m}$, and a contraction of redshift distributions towards lower redshifts (median redshifts a factor of $\sim 1.5\times$ lower), more discrepant with measurements. It is for this reason – in addition to the argument that the dust-temperatures of the SPT samples are well aligned with colder temperatures¹ – that we do not think DSFGs near the epoch of reionization are hot.

¹ The SPT samples could be biased towards cold temperatures with respect to the median at high- z , but as discussed in the Appendix, if DSFGs are much hotter at high- z than the relation shown in Figure 3, then generating number counts at $850 \mu\text{m}$ or 1.1 mm that reproduce measurements would require an even higher DSFG contribution to cosmic star-formation than is shown for Model B.

The impact of the CMB heating is very significant and should not be lost on the reader due to its somewhat straightforward implementation in our model. In the absence of CMB heating, at high- z , galaxies' dust temperatures would be physically cooler, their flux densities would be brighter against the background, and thus much easier to measure. Its absence would make the detection of DSFGs at early epochs significantly easier. For example, while our current dust-rich model B, we estimate ~ 3000 sources deg^{-2} above a 2 mm detection threshold of 0.3 mJy, and without the CMB effects and the same underlying model we would detect $\gtrsim 4500$ –6000 sources deg^{-2} at the same threshold. Indeed, the CMB effect is of crucial importance to the interpretation of the prevalence of high- z DSFGs, without it we might be easily lead to faulty conclusions, and its impact has made it far more difficult to discern the underlying IR luminosity function at early epochs.

A more subtle effect not yet discussed in this paper is the assumption of $\beta_E = 1.8$ for the dust emissivity index of galaxies (without any intrinsic variation). β_E effectively impacts the slope of the Rayleigh-Jeans tail (higher values resulting in steeper SEDs), and measures between $\beta = 1 - 2$ in star-forming galaxies (Hildebrand 1983; Dunne & Eales 2001), and is thought to anticorrelate with dust temperature according to lab experiments and detailed measurements from nearby galaxies (Lisenfeld et al. 2000; Shetty et al. 2009; Tabatabaei et al. 2014). The emissivity index is not thought to vary significantly with metallicity (Rémy-Ruyer et al. 2013), although the environments of DSFGs in the early Universe could be quite different from the local environments in which these measurements are made. Lower values of β_E imply the presence of cooler dust grains, and at the highest redshifts where CMB heating is significant, it is the emission from these cool grains that will be affected more significantly than warmer grains. This would effectively steepen the Rayleigh-Jeans tail. In comparison to the other model effects, this effect is relatively subtle and would not be discernible from existing, or even near-future, datasets.

4.3. Motivating Future Observations

This work has highlighted the importance and urgency of designing specific submm observational programs that directly address the relative ubiquity of dust-rich starbursts towards the epoch of reionization. While existing observational datasets have been partially constraining, hinting at a more dust-rich $z > 4$ Universe than previously thought, though perhaps not as dust-rich as our extreme Model B assumptions, the constraints are weak. The strongest among them come from the measured

redshift distributions of samples of 30–100 DSFGs, as shown in Figure 9. Figure 8 makes clear where the next generation of measurements needs to come from: the highest-redshift DSFGs will be found at 2 mm at depths that can be reached by single-dish facilities over large solid angles $\sim 1 \text{ deg}^2$.

We show the predictive median redshifts (and inner 50% of redshift distributions) for our models at 2 mm in Figure 12, mirroring the format of Figure 10. This shows that the 0.3–1 mJy range of sources at 2 mm are key to distinguishing between such drastically different evolutionary models, when shorter wavelengths are not as constraining. Where SIDES predicts a median redshift of $z \sim 2.9$ for 0.7 mJy sources, Model A predicts $z \approx 3.9$, and Model B suggests a median redshift of $z \approx 6.2$, a dramatic difference! Figure 12, right panel, gives the expected corresponding cumulative number counts per square degree, which is also clearly distinguishable between our two model case-studies, with ≈ 200 or ≈ 1500 sources found above > 0.5 mJy per square degree. Even a survey of much more modest size, $\sim 0.1 \text{ deg}^2$ which could be completed with ALMA, should be able to easily distinguish between these models, containing between ~ 20 –150 sources. Such a survey, would have the added benefit of immediate multiwavelength characterization due to the unambiguous counterparts identified in interferometric data.

Statistical samples of 2 mm-selected DSFGs, followed-up for spectroscopic redshifts – mirroring the existing samples of ~ 100 at $870 \mu\text{m}$ and 1.1 mm – will be critical to discerning between differing hypotheses for the $z > 4$ IR luminosity function. Note that this will not be particularly straightforward to obtain, especially the spectroscopic redshifts for the 2 mm samples. Already spectroscopic follow-up of DSFGs has proven itself one of the most difficult steps of DSFG study over the past 20 years, and it is all the more difficult for higher-redshift DSFGs than lower-redshift DSFGs. Nevertheless, this is what is needed to further constrain the IRLF. It is likely that the vast majority of detected sources will require redshift confirmations in the millimeter via detections of transitions of CO or [C II], most efficiently carried out by ALMA.

Once such complete, spectroscopic datasets are in hand, the backward evolution model described in this paper will be a uniquely useful tool toward making constraints without requiring the full detailed multiwavelength characterization needed for direct luminosity function measurements. Long-term, single-dish facilities like the Large Millimeter Telescope (LMT) and the IRAM 30-m telescope will play essential roles in pushing the largest statistical samples of 2 mm-detected

sources as those pioneered with the GISMO instrument (Staguhn et al. 2014, Magnelli et al., in preparation). Once such samples are in-hand and secured with redshifts, the *James Webb Space Telescope* will play an essential role in illuminating their physical drivers and characteristics, including metal content and stellar/gas kinematics.

5. CONCLUSIONS

This paper has presented a new model for interpreting the bulk infrared/submillimeter characteristics of the extragalactic sky. Our goal in designing this model is to infer constraints on the prevalence and characteristics of the DSFG population at $z \gtrsim 4$ using bulk statistical measurements. The model is built directly from assumptions of the infrared galaxy luminosity function and from known characteristics of the aggregate properties of galaxies’ dust-generated spectral energy distributions. The free parameters of the model are constrained from direct measurements at $z \lesssim 2$, and logical inference out to $z \sim 4$.

To illustrate how poor our constraints on the high- z DSFG population currently are, we use this framework to construct two hypothetical universes. The first (Model A) is dust-poor at high-redshift; this model assumes the DSFG population peaks at $z \sim 2$, and that UV-luminous sources dominate cosmic star-formation over DSFGs at $z \gtrsim 4$. Model A is a reflection of what is often stated in the rest-frame UV extragalactic literature: the DSFGs are too rare to contribute significantly to cosmic star-formation in the first few billion years. The second model (Model B) is dust-rich at high-redshift; this model asserts that the DSFG population peaks around $z \sim 2$ but that its number densities are not much lower at $z \gtrsim 4$. In other words, Model B assumes DSFGs are the dominant source of cosmic star-formation (consisting $\sim 90\%$ of the total) in the first few billion years instead of faint UV-luminous galaxies.

Our comparisons to data from the literature – from number counts, redshift distributions and cross-band flux correlations – suggest that, of our two extreme models, neither can be ruled out by current datasets. This, in particular, is due to the vast majority of constraining datasets existing at submillimeter wavelengths (< 1 mm) that only reliably inform measurements of the IR luminosity function at $z \lesssim 2.5$. In contrast, data at millimeter wavelengths, which should be more direct probes of the high-redshift Universe, are quite limited.

It is clear that more long-wavelength surveys, in particular those at 2 mm, are desperately needed to constrain the evolution of the IR luminosity function beyond current constraints at $z \approx 2.5$. While both 1.1 mm

and 2.0 mm surveys contain very high-redshift DSFGs that will enable these constraints, 2.0 mm surveys will more easily distinguish the highest-redshift sources due to a lack of ‘contaminating’ lower-redshift DSFGs, enabling a much more swift characterization of the high- z DSFG population. We propose that the model designed for this paper can be a uniquely useful tool for the interpretation of DSFG samples across all IR/submm wavebands, where perhaps directly-constraining the galaxy luminosity function is not plausible.

The authors wish to thank the anonymous referee for a constructive report which has improved the paper. We also wish to thank the Aspen Center for Physics for hosting two summer workshops, “The Obscured Universe: Dust and Gas in Distant Starburst Galaxies” in summer 2013 and “New Frontiers in Far-infrared and Submillimeter Astronomy” in summer 2016, whose stimulating conversations led to this work. The Aspen Center for Physics is supported by National Science Foundation grant PHY-1066293. CMC thanks the National Science Foundation for support through grant AST-1714528, and additionally CMC and JAZ thank the University of Texas at Austin College of Natural Sciences for support. JS thanks the McDonald Observatory at the University of Texas at Austin for support through a Smith Fellowship. EdC gratefully acknowledges the Australian Research Council for funding support as the recipient of a Future Fellowship (FT150100079). JAH acknowledges support of the VIDI research programme with project number 639.042.611, which is (partly) financed by the Netherlands Organisation for Scientific Research (NWO). SLF acknowledges support from an NSF AAG award AST-1518183.

This work relies in part on the COSMOS survey dataset; COSMOS is based on observations with the NASA/ESA *Hubble Space Telescope*, obtained at the Space Telescope Science Institute, which is operated by AURA Inc, under NASA contract NAS 5-26555; also based on data collected at: the Subaru Telescope, which is operated by the National Astronomical Observatory of Japan; the *XMM-Newton* and *Herschel Space Observatory*, ESA science missions with instruments and contributions directly funded by ESA Member States and NASA; the European Southern Observatory, Chile; Kitt Peak National Observatory, Cerro Tololo Inter-American Observatory, and the National Optical Astronomy Observatory, which are operated by AURA, Inc., under cooperative agreement with the NSF; the National Radio Astronomy Observatory which is a facility of the NSF operated under cooperative agreement by Associated Universities, Inc.; the Canada-France-Hawaii

Telescope operated by the National Research Council of Canada, the Centre National de la Recherche Scientifique de France and the University of Hawaii; and the James Clerk Maxwell Telescope, which is operated by the East Asian Observatory on behalf of The National Astronomical Observatory of Japan, Academia Sinica Institute of Astronomy and Astrophysics, the Korea Astronomy and Space Science Institute, the Operation, Maintenance and Upgrading Fund for Astronomical Telescopes and Facility Instruments, budgeted from

the Ministry of Finance (MOF) of China and administered by the Chinese Academy of Sciences (CAS). Many of the datasets this paper and analysis rely on were obtained on the summit of Maunakea on the island of Hawai'i. The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Maunakea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain.

REFERENCES

- Aravena, M., Decarli, R., Walter, F., et al. 2016, ArXiv e-prints. <https://arxiv.org/abs/1607.06772>
- Arétxaga, I., Wilson, G. W., Aguilar, E., et al. 2011, MNRAS, 415, 3831, doi: [10.1111/j.1365-2966.2011.18989.x](https://doi.org/10.1111/j.1365-2966.2011.18989.x)
- Arnouts, S., Schiminovich, D., Ilbert, O., et al. 2005, ApJL, 619, L43, doi: [10.1086/426733](https://doi.org/10.1086/426733)
- Austermann, J. E., Dunlop, J. S., Perera, T. A., et al. 2010, MNRAS, 401, 160, doi: [10.1111/j.1365-2966.2009.15620.x](https://doi.org/10.1111/j.1365-2966.2009.15620.x)
- Barger, A. J., Cowie, L. L., Sanders, D. B., et al. 1998, Nature, 394, 248, doi: [10.1038/28338](https://doi.org/10.1038/28338)
- Barger, A. J., Wang, W.-H., Cowie, L. L., et al. 2012, ApJ, 761, 89, doi: [10.1088/0004-637X/761/2/89](https://doi.org/10.1088/0004-637X/761/2/89)
- Barnard, V. E., Vielva, P., Pierce-Price, D. P. I., et al. 2004, MNRAS, 352, 961, doi: [10.1111/j.1365-2966.2004.07985.x](https://doi.org/10.1111/j.1365-2966.2004.07985.x)
- Bastian, N., Covey, K. R., & Meyer, M. R. 2010, ARA&A, 48, 339, doi: [10.1146/annurev-astro-082708-101642](https://doi.org/10.1146/annurev-astro-082708-101642)
- Baugh, C. M., Lacey, C. G., Frenk, C. S., et al. 2005, MNRAS, 356, 1191, doi: [10.1111/j.1365-2966.2004.08553.x](https://doi.org/10.1111/j.1365-2966.2004.08553.x)
- Beelen, A., Omont, A., Bavouzet, N., et al. 2008, A&A, 485, 645, doi: [10.1051/0004-6361/200809500](https://doi.org/10.1051/0004-6361/200809500)
- Berta, S., Magnelli, B., Nordon, R., et al. 2011, A&A, 532, A49, doi: [10.1051/0004-6361/201116844](https://doi.org/10.1051/0004-6361/201116844)
- Béthermin, M., Dole, H., Cousin, M., & Bavouzet, N. 2010, A&A, 516, A43, doi: [10.1051/0004-6361/200913910](https://doi.org/10.1051/0004-6361/200913910)
- Béthermin, M., et al. 2012a, A&A, 542
- Béthermin, M., Daddi, E., Magdis, G., et al. 2012b, ApJL, 757, L23, doi: [10.1088/2041-8205/757/2/L23](https://doi.org/10.1088/2041-8205/757/2/L23)
- Bethermin, M., Wu, H.-Y., Lagache, G., et al. 2017, ArXiv e-prints. <https://arxiv.org/abs/1703.08795>
- Blain, A. W., Barnard, V. E., & Chapman, S. C. 2003, MNRAS, 338, 733, doi: [10.1046/j.1365-8711.2003.06086.x](https://doi.org/10.1046/j.1365-8711.2003.06086.x)
- Blain, A. W., Smail, I., Ivison, R. J., & Kneib, J. 1999, MNRAS, 302, 632, doi: [10.1046/j.1365-8711.1999.02178.x](https://doi.org/10.1046/j.1365-8711.1999.02178.x)
- Borys, C., Chapman, S., Halpern, M., & Scott, D. 2003, MNRAS, 344, 385, doi: [10.1046/j.1365-8711.2003.06818.x](https://doi.org/10.1046/j.1365-8711.2003.06818.x)
- Bouwens, R. J., Illingworth, G. D., Franx, M., & Ford, H. 2007, ApJ, 670, 928, doi: [10.1086/521811](https://doi.org/10.1086/521811)
- Bouwens, R. J., Illingworth, G. D., Franx, M., et al. 2009, ApJ, 705, 936, doi: [10.1088/0004-637X/705/1/936](https://doi.org/10.1088/0004-637X/705/1/936)
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2012, ApJ, 754, 83, doi: [10.1088/0004-637X/754/2/83](https://doi.org/10.1088/0004-637X/754/2/83)
- . 2015, ApJ, 803, 34, doi: [10.1088/0004-637X/803/1/34](https://doi.org/10.1088/0004-637X/803/1/34)
- Bouwens, R. J., Aravena, M., Decarli, R., et al. 2016, ApJ, 833, 72, doi: [10.3847/1538-4357/833/1/72](https://doi.org/10.3847/1538-4357/833/1/72)
- Bowler, R. A. A., Bourne, N., Dunlop, J. S., McLure, R. M., & McLeod, D. J. 2018, ArXiv e-prints. <https://arxiv.org/abs/1802.05720>
- Bradford, C. M., Aguirre, J. E., Aikin, R., et al. 2009, ApJ, 705, 112, doi: [10.1088/0004-637X/705/1/112](https://doi.org/10.1088/0004-637X/705/1/112)
- Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151, doi: [10.1111/j.1365-2966.2004.07881.x](https://doi.org/10.1111/j.1365-2966.2004.07881.x)
- Brisbin, D., Miettinen, O., Aravena, M., et al. 2017, ArXiv e-prints. <https://arxiv.org/abs/1708.05748>
- Brown, A., Nayyeri, H., Cooray, A., et al. 2018, ArXiv e-prints. <https://arxiv.org/abs/1801.02233>
- Burgarella, D., Buat, V., & Iglesias-Páramo, J. 2005, MNRAS, 360, 1413, doi: [10.1111/j.1365-2966.2005.09131.x](https://doi.org/10.1111/j.1365-2966.2005.09131.x)
- Capak, P. L., Carilli, C., Jones, G., et al. 2015, ArXiv e-prints. <https://arxiv.org/abs/1503.07596>
- Caputi, K. I., Lagache, G., Yan, L., et al. 2007, ApJ, 660, 97, doi: [10.1086/512667](https://doi.org/10.1086/512667)
- Casey, C. M. 2012, MNRAS, 425, 3094, doi: [10.1111/j.1365-2966.2012.21455.x](https://doi.org/10.1111/j.1365-2966.2012.21455.x)
- Casey, C. M., Narayanan, D., & Cooray, A. 2014a, PhR, 541, 45, doi: [10.1016/j.physrep.2014.02.009](https://doi.org/10.1016/j.physrep.2014.02.009)
- Casey, C. M., Scoville, N. Z., et al. 2014b, ApJ, 796, 95
- Casey, C. M., Chapman, S. C., Beswick, R. J., et al. 2009, MNRAS, 399, 121, doi: [10.1111/j.1365-2966.2009.15291.x](https://doi.org/10.1111/j.1365-2966.2009.15291.x)
- Casey, C. M., Berta, S., Béthermin, M., et al. 2012a, ApJ, 761, 140, doi: [10.1088/0004-637X/761/2/140](https://doi.org/10.1088/0004-637X/761/2/140)

- . 2012b, *ApJ*, 761, 139, doi: [10.1088/0004-637X/761/2/139](https://doi.org/10.1088/0004-637X/761/2/139)
- Casey, C. M., Chen, C.-C., Cowie, L. L., et al. 2013, *MNRAS*, 436, 1919, doi: [10.1093/mnras/stt1673](https://doi.org/10.1093/mnras/stt1673)
- Casey, C. M., Cooray, A., Killi, M., et al. 2017, *ApJ*, 840, 101, doi: [10.3847/1538-4357/aa6cb1](https://doi.org/10.3847/1538-4357/aa6cb1)
- Catalano, A., Adam, R., Ade, P., et al. 2016, *ArXiv e-prints*. <https://arxiv.org/abs/1605.08628>
- Chabrier, G. 2003, *PASP*, 115, 763, doi: [10.1086/376392](https://doi.org/10.1086/376392)
- Chapman, S. C., Blain, A. W., Ivison, R. J., & Smail, I. R. 2003a, *Nature*, 422, 695
- Chapman, S. C., Blain, A. W., Smail, I., & Ivison, R. J. 2005, *ApJ*, 622, 772, doi: [10.1086/428082](https://doi.org/10.1086/428082)
- Chapman, S. C., Helou, G., Lewis, G. F., & Dale, D. A. 2003b, *ApJ*, 588, 186, doi: [10.1086/374038](https://doi.org/10.1086/374038)
- Chapman, S. C., Lewis, G. F., Scott, D., Borys, C., & Richards, E. 2002, *ApJ*, 570, 557, doi: [10.1086/339498](https://doi.org/10.1086/339498)
- Chapman, S. C., Smail, I., Blain, A. W., & Ivison, R. J. 2004, *ApJ*, 614, 671, doi: [10.1086/423833](https://doi.org/10.1086/423833)
- Chary, R., & Elbaz, D. 2001, *ApJ*, 556, 562, doi: [10.1086/321609](https://doi.org/10.1086/321609)
- Chen, C.-C., Cowie, L. L., Barger, A. J., et al. 2013, *ApJ*, 762, 81, doi: [10.1088/0004-637X/762/2/81](https://doi.org/10.1088/0004-637X/762/2/81)
- Clements, D. L., Dunne, L., & Eales, S. 2010, *MNRAS*, 403, 274, doi: [10.1111/j.1365-2966.2009.16064.x](https://doi.org/10.1111/j.1365-2966.2009.16064.x)
- Coe, D., Zitrin, A., Carrasco, M., et al. 2013, *ApJ*, 762, 32, doi: [10.1088/0004-637X/762/1/32](https://doi.org/10.1088/0004-637X/762/1/32)
- Conley, A., et al. 2011, *ApJL*, 732, L35, doi: [10.1088/2041-8205/732/2/L35](https://doi.org/10.1088/2041-8205/732/2/L35)
- Coppin, K., Chapin, E. L., Mortier, A. M. J., et al. 2006, *MNRAS*, 372, 1621, doi: [10.1111/j.1365-2966.2006.10961.x](https://doi.org/10.1111/j.1365-2966.2006.10961.x)
- Coppin, K. E. K., Geach, J. E., Almaini, O., et al. 2015, *MNRAS*, 446, 1293, doi: [10.1093/mnras/stu2185](https://doi.org/10.1093/mnras/stu2185)
- Cowie, L. L., Barger, A. J., Hsu, L.-Y., et al. 2017, *ApJ*, 837, 139, doi: [10.3847/1538-4357/aa60bb](https://doi.org/10.3847/1538-4357/aa60bb)
- Cowie, L. L., Barger, A. J., & Kneib, J.-P. 2002, *AJ*, 123, 2197, doi: [10.1086/339978](https://doi.org/10.1086/339978)
- da Cunha, E., Charlot, S., & Elbaz, D. 2008, *MNRAS*, 388, 1595
- da Cunha, E., Groves, B., Walter, F., et al. 2013, *ApJ*, 766, 13, doi: [10.1088/0004-637X/766/1/13](https://doi.org/10.1088/0004-637X/766/1/13)
- Dahlen, T., Mobasher, B., Dickinson, M., et al. 2007, *ApJ*, 654, 172, doi: [10.1086/508854](https://doi.org/10.1086/508854)
- Dale, D. A., & Helou, G. 2002, *ApJ*, 576, 159, doi: [10.1086/341632](https://doi.org/10.1086/341632)
- Dale, D. A., Helou, G., Contursi, A., Silbermann, N. A., & Kolhatkar, S. 2001, *ApJ*, 549, 215, doi: [10.1086/319077](https://doi.org/10.1086/319077)
- Danielson, A. L. R., Swinbank, A. M., Smail, I., et al. 2017, *ApJ*, 840, 78, doi: [10.3847/1538-4357/aa6caf](https://doi.org/10.3847/1538-4357/aa6caf)
- Davidzon, I., Ilbert, O., Laigle, C., et al. 2017, *A&A*, 605, A70, doi: [10.1051/0004-6361/201730419](https://doi.org/10.1051/0004-6361/201730419)
- de Zotti, G., Ricci, R., Mesa, D., et al. 2005, *A&A*, 431, 893, doi: [10.1051/0004-6361:20042108](https://doi.org/10.1051/0004-6361:20042108)
- Dole, H., Le Floc'h, E., Pérez-González, P. G., et al. 2004, *ApJS*, 154, 87, doi: [10.1086/422472](https://doi.org/10.1086/422472)
- Dopita, M. A., Groves, B. A., Fischera, J., et al. 2005, *ApJ*, 619, 755, doi: [10.1086/423948](https://doi.org/10.1086/423948)
- Dowell, C. D., Conley, A., Glenn, J., et al. 2014, *ApJ*, 780, 75, doi: [10.1088/0004-637X/780/1/75](https://doi.org/10.1088/0004-637X/780/1/75)
- Draine, B. T., & Li, A. 2007, *ApJ*, 657, 810, doi: [10.1086/511055](https://doi.org/10.1086/511055)
- Dunlop, J. S., McLure, R. J., Biggs, A. D., et al. 2016, *ArXiv e-prints*. <https://arxiv.org/abs/1606.00227>
- Dunne, L., & Eales, S. A. 2001, *MNRAS*, 327, 697, doi: [10.1046/j.1365-8711.2001.04789.x](https://doi.org/10.1046/j.1365-8711.2001.04789.x)
- Dwek, E., Staguhn, J., Arendt, R. G., et al. 2014, *ApJL*, 788, L30, doi: [10.1088/2041-8205/788/2/L30](https://doi.org/10.1088/2041-8205/788/2/L30)
- Dwek, E., Arendt, R. G., Hauser, M. G., et al. 1998, *ApJ*, 508, 106, doi: [10.1086/306382](https://doi.org/10.1086/306382)
- Eales, S., Lilly, S., Webb, T., et al. 2000, *AJ*, 120, 2244, doi: [10.1086/316823](https://doi.org/10.1086/316823)
- Ellis, R. S., McLure, R. J., Dunlop, J. S., et al. 2013, *ApJL*, 763, L7, doi: [10.1088/2041-8205/763/1/L7](https://doi.org/10.1088/2041-8205/763/1/L7)
- Faisst, A. L., Capak, P. L., Yan, L., et al. 2017, *ArXiv e-prints*. <https://arxiv.org/abs/1708.07842>
- Finkelstein, S. L. 2016, *PASA*, 33, e037, doi: [10.1017/pasa.2016.26](https://doi.org/10.1017/pasa.2016.26)
- Finkelstein, S. L., Papovich, C., Salmon, B., et al. 2012, *ApJ*, 756, 164, doi: [10.1088/0004-637X/756/2/164](https://doi.org/10.1088/0004-637X/756/2/164)
- Finkelstein, S. L., Papovich, C., Dickinson, M., et al. 2013, *Nature*, 502, 524, doi: [10.1038/nature12657](https://doi.org/10.1038/nature12657)
- Finkelstein, S. L., Ryan, Jr., R. E., Papovich, C., et al. 2015, *ApJ*, 810, 71, doi: [10.1088/0004-637X/810/1/71](https://doi.org/10.1088/0004-637X/810/1/71)
- Fixsen, D. J., Dwek, E., Mather, J. C., Bennett, C. L., & Shafer, R. A. 1998, *ApJ*, 508, 123, doi: [10.1086/306383](https://doi.org/10.1086/306383)
- Fujimoto, S., Ouchi, M., Ono, Y., et al. 2016, *ApJS*, 222, 1, doi: [10.3847/0067-0049/222/1/1](https://doi.org/10.3847/0067-0049/222/1/1)
- Geach, J. E., Chapin, E. L., Coppin, K. E. K., et al. 2013, *MNRAS*, 432, 53, doi: [10.1093/mnras/stt352](https://doi.org/10.1093/mnras/stt352)
- Geach, J. E., Dunlop, J. S., Halpern, M., et al. 2016, *ArXiv e-prints*. <https://arxiv.org/abs/1607.03904>
- Goto, T., Takagi, T., Matsuhara, H., et al. 2010, *A&A*, 514, A6+, doi: [10.1051/0004-6361/200913182](https://doi.org/10.1051/0004-6361/200913182)
- Griffin, M. J., Abergel, A., Abreu, A., et al. 2010, *A&A*, 518, L3, doi: [10.1051/0004-6361/201014519](https://doi.org/10.1051/0004-6361/201014519)
- Gruppioni, C., et al. 2013, *MNRAS*, 432, 23
- Hatsukade, B., et al. 2011, *MNRAS*, 411, 102
- Hayward, C. C., Behroozi, P. S., Somerville, R. S., et al. 2013, *MNRAS*, 434, 2572, doi: [10.1093/mnras/stt1202](https://doi.org/10.1093/mnras/stt1202)

- Héraudeau, P., Oliver, S., del Burgo, C., et al. 2004, MNRAS, 354, 924, doi: [10.1111/j.1365-2966.2004.08259.x](https://doi.org/10.1111/j.1365-2966.2004.08259.x)
- Hezaveh, Y. D., & Holder, G. P. 2011, ApJ, 734, 52, doi: [10.1088/0004-637X/734/1/52](https://doi.org/10.1088/0004-637X/734/1/52)
- Hezaveh, Y. D., Marrone, D. P., Fassnacht, C. D., et al. 2013, ApJ, 767, 132, doi: [10.1088/0004-637X/767/2/132](https://doi.org/10.1088/0004-637X/767/2/132)
- Hildebrand, R. H. 1983, QJRAS, 24, 267
- Hill, R., Chapman, S. C., Scott, D., et al. 2017, ArXiv e-prints. <https://arxiv.org/abs/1710.02231>
- Hodge, J. A., Karim, A., Smail, I., et al. 2013, ApJ, 768, 91, doi: [10.1088/0004-637X/768/1/91](https://doi.org/10.1088/0004-637X/768/1/91)
- Holland, W. S., et al. 2013, MNRAS, 430, 2513
- Hopkins, P. F., Richards, G. T., & Hernquist, L. 2007, ApJ, 654, 731, doi: [10.1086/509629](https://doi.org/10.1086/509629)
- Hughes, D. H., Serjeant, S., Dunlop, J., et al. 1998, Nature, 394, 241, doi: [10.1038/28328](https://doi.org/10.1038/28328)
- Ivison, R. J., Lewis, A. J. R., Weiss, A., et al. 2016, ApJ, 832, 78, doi: [10.3847/0004-637X/832/1/78](https://doi.org/10.3847/0004-637X/832/1/78)
- Jaacks, J., Finkelstein, S. L., & Bromm, V. 2018, MNRAS, doi: [10.1093/mnras/sty049](https://doi.org/10.1093/mnras/sty049)
- Jones, A. P., Fanciullo, L., Köhler, M., et al. 2013, A&A, 558, A62, doi: [10.1051/0004-6361/201321686](https://doi.org/10.1051/0004-6361/201321686)
- Karim, A., Swinbank, A. M., Hodge, J. A., et al. 2013, MNRAS, 432, 2, doi: [10.1093/mnras/stt196](https://doi.org/10.1093/mnras/stt196)
- Kartaltepe, J. S., Dickinson, M., Alexander, D. M., et al. 2012, ApJ, 757, 23, doi: [10.1088/0004-637X/757/1/23](https://doi.org/10.1088/0004-637X/757/1/23)
- Kawara, K., Matsuhara, H., Okuda, H., et al. 2004, A&A, 413, 843, doi: [10.1051/0004-6361:20031532](https://doi.org/10.1051/0004-6361:20031532)
- Kirkpatrick, A., Pope, A., Sajina, A., et al. 2017, ApJ, 843, 71, doi: [10.3847/1538-4357/aa76dc](https://doi.org/10.3847/1538-4357/aa76dc)
- Knudsen, K. K., van der Werf, P. P., & Kneib, J.-P. 2008, MNRAS, 384, 1611, doi: [10.1111/j.1365-2966.2007.12820.x](https://doi.org/10.1111/j.1365-2966.2007.12820.x)
- Kocevski, D. D., Faber, S. M., Mozena, M., et al. 2012, ApJ, 744, 148, doi: [10.1088/0004-637X/744/2/148](https://doi.org/10.1088/0004-637X/744/2/148)
- Lacey, C. G., Baugh, C. M., Frenk, C. S., et al. 2008, MNRAS, 385, 1155, doi: [10.1111/j.1365-2966.2008.12949.x](https://doi.org/10.1111/j.1365-2966.2008.12949.x)
- Lacy, M., Ridgway, S. E., Sajina, A., et al. 2015, ApJ, 802, 102, doi: [10.1088/0004-637X/802/2/102](https://doi.org/10.1088/0004-637X/802/2/102)
- Lagache, G., Dole, H., Puget, J.-L., et al. 2004, ApJS, 154, 112, doi: [10.1086/422392](https://doi.org/10.1086/422392)
- Le Floc'h, E., Papovich, C., Dole, H., et al. 2005, ApJ, 632, 169, doi: [10.1086/432789](https://doi.org/10.1086/432789)
- Lee, N., Sanders, D. B., Casey, C. M., et al. 2013, ApJ, 778, 131, doi: [10.1088/0004-637X/778/2/131](https://doi.org/10.1088/0004-637X/778/2/131)
- Lisenfeld, U., Isaak, K. G., & Hills, R. 2000, MNRAS, 312, 433, doi: [10.1046/j.1365-8711.2000.03150.x](https://doi.org/10.1046/j.1365-8711.2000.03150.x)
- Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415, doi: [10.1146/annurev-astro-081811-125615](https://doi.org/10.1146/annurev-astro-081811-125615)
- Magnelli, B., Elbaz, D., Chary, R. R., et al. 2011, A&A, 528, A35, doi: [10.1051/0004-6361/200913941](https://doi.org/10.1051/0004-6361/200913941)
- Magnelli, B., Popesso, P., Berta, S., et al. 2013, A&A, 553, A132, doi: [10.1051/0004-6361/201321371](https://doi.org/10.1051/0004-6361/201321371)
- Marrone, D. P., Spilker, J. S., Hayward, C. C., et al. 2017, ArXiv e-prints. <https://arxiv.org/abs/1712.03020>
- Matsuura, M., Wood, P. R., Sloan, G. C., et al. 2006, MNRAS, 371, 415, doi: [10.1111/j.1365-2966.2006.10664.x](https://doi.org/10.1111/j.1365-2966.2006.10664.x)
- Matsuura, M., Barlow, M. J., Zijlstra, A. A., et al. 2009, MNRAS, 396, 918, doi: [10.1111/j.1365-2966.2009.14743.x](https://doi.org/10.1111/j.1365-2966.2009.14743.x)
- Matsuura, M., Dwek, E., Meixner, M., et al. 2011, Science, 333, 1258, doi: [10.1126/science.1205983](https://doi.org/10.1126/science.1205983)
- McKee, C. F., & Tan, J. C. 2003, ApJ, 585, 850, doi: [10.1086/346149](https://doi.org/10.1086/346149)
- McLeod, D. J., McLure, R. J., Dunlop, J. S., et al. 2015, MNRAS, 450, 3032, doi: [10.1093/mnras/stv780](https://doi.org/10.1093/mnras/stv780)
- McLure, R. J., Dunlop, J. S., Bowler, R. A. A., et al. 2013, MNRAS, 432, 2696, doi: [10.1093/mnras/stt627](https://doi.org/10.1093/mnras/stt627)
- Murphy, E. J., Chary, R.-R., Dickinson, M., et al. 2011, ApJ, 732, 126, doi: [10.1088/0004-637X/732/2/126](https://doi.org/10.1088/0004-637X/732/2/126)
- Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, ApJL, 660, L43, doi: [10.1086/517926](https://doi.org/10.1086/517926)
- Noll, S., Burgarella, D., Giovannoli, E., et al. 2009, A&A, 507, 1793, doi: [10.1051/0004-6361/200912497](https://doi.org/10.1051/0004-6361/200912497)
- Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2013, ApJ, 773, 75, doi: [10.1088/0004-637X/773/1/75](https://doi.org/10.1088/0004-637X/773/1/75)
- . 2014, ApJ, 786, 108, doi: [10.1088/0004-637X/786/2/108](https://doi.org/10.1088/0004-637X/786/2/108)
- Oesch, P. A., Brammer, G., van Dokkum, P. G., et al. 2016, ApJ, 819, 129, doi: [10.3847/0004-637X/819/2/129](https://doi.org/10.3847/0004-637X/819/2/129)
- Oliver, S., Frost, M., Farrah, D., et al. 2010, MNRAS, 405, 2279, doi: [10.1111/j.1365-2966.2010.16643.x](https://doi.org/10.1111/j.1365-2966.2010.16643.x)
- Oteo, I., Zwaan, M. A., Ivison, R. J., Smail, I., & Biggs, A. D. 2016, ApJ, 822, 36, doi: [10.3847/0004-637X/822/1/36](https://doi.org/10.3847/0004-637X/822/1/36)
- Oteo, I., Ivison, R. J., Dunne, L., et al. 2017, ArXiv e-prints. <https://arxiv.org/abs/1709.02809>
- Pannella, M., Carilli, C. L., Daddi, E., et al. 2009, ApJL, 698, L116, doi: [10.1088/0004-637X/698/2/L116](https://doi.org/10.1088/0004-637X/698/2/L116)
- Pannella, M., Elbaz, D., Daddi, E., et al. 2015, ApJ, 807, 141, doi: [10.1088/0004-637X/807/2/141](https://doi.org/10.1088/0004-637X/807/2/141)
- Paradis, D., Veneziani, M., Noriega-Crespo, A., et al. 2010, A&A, 520, L8, doi: [10.1051/0004-6361/201015301](https://doi.org/10.1051/0004-6361/201015301)
- Patanchon, G., Ade, P. A. R., Bock, J. J., et al. 2009, ApJ, 707, 1750, doi: [10.1088/0004-637X/707/2/1750](https://doi.org/10.1088/0004-637X/707/2/1750)
- Pearson, E. A., Eales, S., Dunne, L., et al. 2013, MNRAS, 435, 2753, doi: [10.1093/mnras/stt1369](https://doi.org/10.1093/mnras/stt1369)
- Perera, T. A., Chapin, E. L., Austermann, J. E., et al. 2008, MNRAS, 391, 1227, doi: [10.1111/j.1365-2966.2008.13902.x](https://doi.org/10.1111/j.1365-2966.2008.13902.x)

- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, *A&A*, 594, A13, doi: [10.1051/0004-6361/201525830](https://doi.org/10.1051/0004-6361/201525830)
- Poglitsch, A., Waelkens, C., Geis, N., et al. 2010, *A&A*, 518, L2, doi: [10.1051/0004-6361/201014535](https://doi.org/10.1051/0004-6361/201014535)
- Portegies Zwart, S. F., McMillan, S. L. W., & Gieles, M. 2010, *ARA&A*, 48, 431, doi: [10.1146/annurev-astro-081309-130834](https://doi.org/10.1146/annurev-astro-081309-130834)
- Puget, J.-L., Abergel, A., Bernard, J.-P., et al. 1996, *A&A*, 308, L5
- Reddy, N., Dickinson, M., Elbaz, D., et al. 2012, *ApJ*, 744, 154, doi: [10.1088/0004-637X/744/2/154](https://doi.org/10.1088/0004-637X/744/2/154)
- Reddy, N. A., & Steidel, C. C. 2009, *ApJ*, 692, 778, doi: [10.1088/0004-637X/692/1/778](https://doi.org/10.1088/0004-637X/692/1/778)
- R  my-Ruyer, A., Madden, S. C., Galliano, F., et al. 2013, *A&A*, 557, A95, doi: [10.1051/0004-6361/201321602](https://doi.org/10.1051/0004-6361/201321602)
- Riechers, D. A., Bradford, C. M., Clements, D. L., et al. 2013, *Nature*, 496, 329, doi: [10.1038/nature12050](https://doi.org/10.1038/nature12050)
- Rieke, G. H., Alonso-Herrero, A., Weiner, B. J., et al. 2009, *ApJ*, 692, 556, doi: [10.1088/0004-637X/692/1/556](https://doi.org/10.1088/0004-637X/692/1/556)
- Rodighiero, G., & Franceschini, A. 2004, *A&A*, 419, L55, doi: [10.1051/0004-6361:20040144](https://doi.org/10.1051/0004-6361:20040144)
- Roseboom, I. G., Dunlop, J. S., Cirasuolo, M., et al. 2013, *MNRAS*, 436, 430, doi: [10.1093/mnras/stt1577](https://doi.org/10.1093/mnras/stt1577)
- Sanders, D. B., Mazzarella, J. M., Kim, D.-C., Surace, J. A., & Soifer, B. T. 2003, *AJ*, 126, 1607, doi: [10.1086/376841](https://doi.org/10.1086/376841)
- Sargent, M. T., B  thermin, M., Daddi, E., & Elbaz, D. 2012, *ApJL*, 747, L31, doi: [10.1088/2041-8205/747/2/L31](https://doi.org/10.1088/2041-8205/747/2/L31)
- Schenker, M. A., Robertson, B. E., Ellis, R. S., et al. 2013, *ApJ*, 768, 196, doi: [10.1088/0004-637X/768/2/196](https://doi.org/10.1088/0004-637X/768/2/196)
- Schiminovich, D., Ilbert, O., Arnouts, S., et al. 2005, *ApJL*, 619, L47, doi: [10.1086/427077](https://doi.org/10.1086/427077)
- Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, *A&A*, 575, A74, doi: [10.1051/0004-6361/201425017](https://doi.org/10.1051/0004-6361/201425017)
- Scott, K. S., Yun, M. S., Wilson, G. W., et al. 2010, *MNRAS*, 405, 2260, doi: [10.1111/j.1365-2966.2010.16644.x](https://doi.org/10.1111/j.1365-2966.2010.16644.x)
- Scott, K. S., Wilson, G. W., Aretxaga, I., et al. 2012, *MNRAS*, 423, 575, doi: [10.1111/j.1365-2966.2012.20905.x](https://doi.org/10.1111/j.1365-2966.2012.20905.x)
- Scott, S. E., Dunlop, J. S., & Serjeant, S. 2006, *MNRAS*, 370, 1057, doi: [10.1111/j.1365-2966.2006.10478.x](https://doi.org/10.1111/j.1365-2966.2006.10478.x)
- Scott, S. E., Fox, M. J., Dunlop, J. S., et al. 2002, *MNRAS*, 331, 817, doi: [10.1046/j.1365-8711.2002.05193.x](https://doi.org/10.1046/j.1365-8711.2002.05193.x)
- Shetty, R., Kauffmann, J., Schnee, S., & Goodman, A. A. 2009, *ApJ*, 696, 676, doi: [10.1088/0004-637X/696/1/676](https://doi.org/10.1088/0004-637X/696/1/676)
- Siebenmorgen, R., & Kr  gel, E. 2007, *A&A*, 461, 445, doi: [10.1051/0004-6361:20065700](https://doi.org/10.1051/0004-6361:20065700)
- Silva, L., Granato, G. L., Bressan, A., & Danese, L. 1998, *ApJ*, 509, 103, doi: [10.1086/306476](https://doi.org/10.1086/306476)
- Simpson, J. M., Smail, I., Swinbank, A. M., et al. 2015, *ApJ*, 807, 128, doi: [10.1088/0004-637X/807/2/128](https://doi.org/10.1088/0004-637X/807/2/128)
- . 2017, *ApJ*, 839, 58, doi: [10.3847/1538-4357/aa65d0](https://doi.org/10.3847/1538-4357/aa65d0)
- Smail, I., Ivison, R. J., & Blain, A. W. 1997, *ApJL*, 490, L5+, doi: [10.1086/311017](https://doi.org/10.1086/311017)
- Smail, I., Ivison, R. J., Blain, A. W., & Kneib, J. 2002, *MNRAS*, 331, 495, doi: [10.1046/j.1365-8711.2002.05203.x](https://doi.org/10.1046/j.1365-8711.2002.05203.x)
- Smol    , V., Aravena, M., Navarrete, F., et al. 2012, *A&A*, 548, A4, doi: [10.1051/0004-6361/201219368](https://doi.org/10.1051/0004-6361/201219368)
- Somerville, R. S., Gilmore, R. C., Primack, J. R., & Dom  nguez, A. 2012, *MNRAS*, 423, 1992, doi: [10.1111/j.1365-2966.2012.20490.x](https://doi.org/10.1111/j.1365-2966.2012.20490.x)
- Song, M., Finkelstein, S. L., Ashby, M. L. N., et al. 2016, *ApJ*, 825, 5, doi: [10.3847/0004-637X/825/1/5](https://doi.org/10.3847/0004-637X/825/1/5)
- Spilker, J. S., Marrone, D. P., Aravena, M., et al. 2016, *ApJ*, 826, 112, doi: [10.3847/0004-637X/826/2/112](https://doi.org/10.3847/0004-637X/826/2/112)
- Springel, V. 2005, *MNRAS*, 364, 1105, doi: [10.1111/j.1365-2966.2005.09655.x](https://doi.org/10.1111/j.1365-2966.2005.09655.x)
- Staguhn, J. G., Kov  cs, A., Arendt, R. G., et al. 2014, *ApJ*, 790, 77, doi: [10.1088/0004-637X/790/1/77](https://doi.org/10.1088/0004-637X/790/1/77)
- Steidel, C. C., Gialalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996, *ApJL*, 462, L17, doi: [10.1086/310029](https://doi.org/10.1086/310029)
- Strandet, M. L., Weiss, A., Vieira, J. D., et al. 2016, *ApJ*, 822, 80, doi: [10.3847/0004-637X/822/2/80](https://doi.org/10.3847/0004-637X/822/2/80)
- Strandet, M. L., Weiss, A., De Breuck, C., et al. 2017, *ApJL*, 842, L15, doi: [10.3847/2041-8213/aa74b0](https://doi.org/10.3847/2041-8213/aa74b0)
- Swinbank, A. M., Smail, I., Chapman, S. C., et al. 2004, *ApJ*, 617, 64, doi: [10.1086/425171](https://doi.org/10.1086/425171)
- Swinbank, A. M., Lacey, C. G., Smail, I., et al. 2008, *MNRAS*, 391, 420, doi: [10.1111/j.1365-2966.2008.13911.x](https://doi.org/10.1111/j.1365-2966.2008.13911.x)
- Tabatabaei, F. S., Braine, J., Xilouris, E. M., et al. 2014, *A&A*, 561, A95, doi: [10.1051/0004-6361/201321441](https://doi.org/10.1051/0004-6361/201321441)
- Tucci, M., Toffolatti, L., de Zotti, G., & Mart  nez-Gonz  lez, E. 2011, *A&A*, 533, A57, doi: [10.1051/0004-6361/201116972](https://doi.org/10.1051/0004-6361/201116972)
- U, V., Sanders, D. B., Mazzarella, J. M., et al. 2012, *ApJS*, 203, 9, doi: [10.1088/0067-0049/203/1/9](https://doi.org/10.1088/0067-0049/203/1/9)
- Valiante, E., Smith, M. W. L., Eales, S., et al. 2016, *MNRAS*, 462, 3146, doi: [10.1093/mnras/stw1806](https://doi.org/10.1093/mnras/stw1806)
- Vieira, J. D., Marrone, D. P., Chapman, S. C., et al. 2013, *Nature*, 495, 344, doi: [10.1038/nature12001](https://doi.org/10.1038/nature12001)
- Viero, M. P., Moncelsi, L., Quadri, R. F., et al. 2015, *ApJL*, 809, L22, doi: [10.1088/2041-8205/809/2/L22](https://doi.org/10.1088/2041-8205/809/2/L22)
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014a, *MNRAS*, 444, 1518, doi: [10.1093/mnras/stu1536](https://doi.org/10.1093/mnras/stu1536)
- . 2014b, *Nature*, 509, 177, doi: [10.1038/nature13316](https://doi.org/10.1038/nature13316)
- Vogelsberger, M., Marinacci, F., Torrey, P., et al. 2018, *MNRAS*, 474, 2073, doi: [10.1093/mnras/stx2955](https://doi.org/10.1093/mnras/stx2955)
- Walter, F., et al. 2016, *ApJ*

- Wang, W.-H., Lin, W.-C., Lim, C.-F., et al. 2017, ArXiv e-prints. <https://arxiv.org/abs/1707.00990>
- Wardlow, J. L., Smail, I., Coppin, K. E. K., et al. 2011, MNRAS, 415, 1479, doi: [10.1111/j.1365-2966.2011.18795.x](https://doi.org/10.1111/j.1365-2966.2011.18795.x)
- Webb, T. M., Eales, S., Foucaud, S., et al. 2003, ApJ, 582, 6, doi: [10.1086/344608](https://doi.org/10.1086/344608)
- Wechsler, R. H., Bullock, J. S., Primack, J. R., Kravtsov, A. V., & Dekel, A. 2002, ApJ, 568, 52, doi: [10.1086/338765](https://doi.org/10.1086/338765)
- Weinberger, R., Springel, V., Hernquist, L., et al. 2017, MNRAS, 465, 3291, doi: [10.1093/mnras/stw2944](https://doi.org/10.1093/mnras/stw2944)
- Weiß, A., Ivison, R. J., Downes, D., et al. 2009, ApJL, 705, L45, doi: [10.1088/0004-637X/705/1/L45](https://doi.org/10.1088/0004-637X/705/1/L45)
- Weiß, A., De Breuck, C., Marrone, D. P., et al. 2013, ApJ, 767, 88, doi: [10.1088/0004-637X/767/1/88](https://doi.org/10.1088/0004-637X/767/1/88)
- Whitaker, K. E., Pope, A., Cybulski, R., et al. 2017, ArXiv e-prints. <https://arxiv.org/abs/1710.06872>
- Whitaker, K. E., Franx, M., Leja, J., et al. 2014, ApJ, 795, 104, doi: [10.1088/0004-637X/795/2/104](https://doi.org/10.1088/0004-637X/795/2/104)
- Wilkins, S. M., Feng, Y., Di Matteo, T., et al. 2017a, ArXiv e-prints. <https://arxiv.org/abs/1710.01976>
- . 2017b, MNRAS, 469, 2517, doi: [10.1093/mnras/stx841](https://doi.org/10.1093/mnras/stx841)
- Wilson, C. D., Petitpas, G. R., Iono, D., et al. 2008, ApJS, 178, 189, doi: [10.1086/590910](https://doi.org/10.1086/590910)
- Zavala, J. A., Aretxaga, I., & Hughes, D. H. 2014, MNRAS, 443, 2384, doi: [10.1093/mnras/stu1330](https://doi.org/10.1093/mnras/stu1330)
- Zavala, J. A., Aretxaga, I., Geach, J. E., et al. 2017, MNRAS, 464, 3369, doi: [10.1093/mnras/stw2630](https://doi.org/10.1093/mnras/stw2630)
- Zavala, J. A., Montaña, A., Hughes, D. H., et al. 2018, Nature Astronomy, 2, 56, doi: [10.1038/s41550-017-0297-8](https://doi.org/10.1038/s41550-017-0297-8)

APPENDIX

A. DATA CONSTRAINTS ON THE FIXED PARAMETERS

Table 3 provides a list of the fifteen tunable parameters of our model; because the number of parameters is so high, and it might not be immediately obvious to the reader what the impact of a change in one of the variables might mean for our analysis, we provide this Appendix to describe how well each parameter is known and how a change in that parameter might change our results. Furthermore, in this work we emphasize that the set of variables we have chosen is not necessarily the best or the only set of parameters that would give a satisfactory result. Similarly, many of the chosen values for our parameters are intrinsically tied, and so we discuss some of those relationships here. We split the variables into three classes: secure (of which there are 10), minor impact (3) and unknown (2). We present discussions of these parameters in two subsections: luminosity function parameters and SED parameters. We weave in discussion of the uncertainty in the redshift evolution of these parameters in each respective section.

A.1. *Security of Luminosity Function Parameters*

We have characterized three of four luminosity function parameters – L_0 , Φ_0 and β_{LF} – as secure in Table 3. This is based entirely on the measured IR luminosity function (as shown in Figure 2) at $z = 0 - 0.5$. The bright-end of the luminosity function is well characterized with $\beta_{\text{LF}} = -3.00 \pm 0.15$. The exact values of L_0 and Φ_0 are correlated with γ_1 and ψ_1 respectively. We find that the combination of $L_0 = 10^{11.1} L_\odot$, $\Phi_0 = 10^{-3.5} \text{ Mpc}^{-3} \text{ dex}^{-1}$, and $\gamma_1 = 2.8$ and $\psi_1 = 0$ to provide satisfactory fits to: the measured IRLF at $z \lesssim 2$ shown in Figure 2, the reports of measured values of L_\star and Φ_\star from the IR literature as shown in Figure 4, and the inferred LIRG, ULIRG, and total IR contributions to the cosmic star-formation rate densities from $0 < z < 2$ as shown in Figure 6. We caution that this combination of values is not absolutely unique in providing an adequate fit to all existing data; for example, an adjustment of Φ_0 and γ_1 upward and L_0 and ψ_1 downward could produce similar results in deg^2 -scale maps. Such a manipulation would potentially (but not necessarily) change the simulated output for a shallow survey conducted on $100\text{-}10^4 \text{ deg}^2$ -scales, where sources in the local Universe ($z < 0.5$) are the dominant population. The goal is to choose parameters which adequately represent the underlying number densities of galaxies at different luminosities well, and a range of parameter sets can do this within the measurement uncertainties.

The most uncertain parameter of the luminosity function is the faint-end slope of the luminosity function and whether or not it evolves. The choice of α_{LF} matters a great deal to the interpretation of the IRLF and its relationship to the UV luminosity function, and indeed, changing α_{LF} from the initially adopted value of -0.6 can significantly impact the resulting number counts and redshift distribution analysis. Our choice of $\alpha_{\text{LF}} = -0.6$ comes from the average inferred faint-end slope measured from the five luminosity function bins shown in Figure 2.

Physically, it follows that this faint-end slope should be much flatter than the faint end slope of the UV luminosity function, given the increased obscuration of galaxies at high masses and luminosities (Pannella et al. 2009, 2015; Whitaker et al. 2014, 2017). Generally, IR-luminous galaxies are more massive than UV-luminous galaxies and proportionally more obscured. A low-mass galaxy is less likely to be detectable via its dust emission, so most of its energy will be emitted directly through unobscured channels, such that low-mass, IR-luminous galaxies are relatively rare in comparison. If one were to apply the average obscuration as a function of stellar mass to the stellar mass function, then assume a main sequence translation of the star-formation rates of those galaxies, the resulting UV-luminous and IR-luminous luminosity functions would resemble measurements: a Schechter-like luminosity function for UV-selected galaxies, with a faint-end slope of the luminosity function that is much steeper than the IR luminosity function. The IR luminosity function would also exhibit a shallower fall-off at the bright-end, in comparison to the exponential fall-off for the UV luminosity function. This motivates a boundary condition on $\alpha_{\text{LF}} \lesssim -1.5$, which reassuringly means that sources right above our detection threshold will not overwhelmingly dominate the number counts of detectable galaxies.

Figure 13 shows the detection limits of our mock maps, as listed in Table 1, illustrated as a boundary in $L_{\text{IR}}\text{-}z$ against our chosen modeling of $L_\star(z)$. This gives us additional intuition for the impact of the choice of α_{LF} at each modeled wavelength. The wavelengths least impacted by the choice of α_{LF} are $250\mu\text{m}$, $350\mu\text{m}$, $500\mu\text{m}$, and $850\mu\text{m}$. In other words, any adjustment in the faint end slope is unlikely to impact the extracted number counts at those wavelengths. At the other wavelengths, we note that changes in $\alpha_{\text{LF}} = -0.5$ to $\alpha_{\text{LF}} = -0.8$ results in changes of order $\sim 0.05\text{-}0.20 \text{ dex}$ in the extracted number counts at the 5σ detection threshold, with reduced impact towards brighter sources. The wavelength

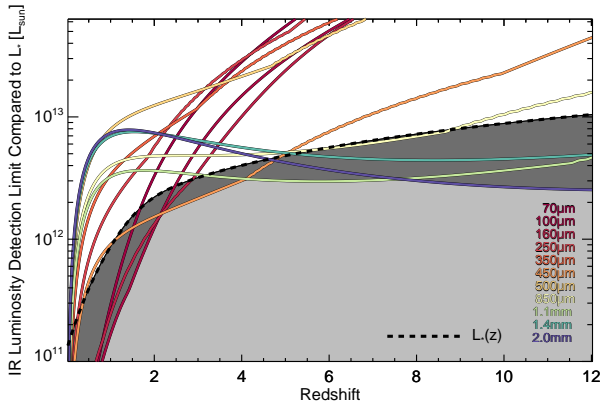


Figure 13. An illustration of the $L_{\text{IR}}-z$ space probed by each wavelength modeled in this paper at the specified RMS listed in Table 1. The colored lines indicate the median 5σ detection limit in L_{IR} as a function of redshift at the given wavelength. The thick dashed line shows our modeled evolution of L_* as adopted in this paper. Below this threshold (gray shaded area), the faint-end slope of the luminosity function, α_{LF} , determines source density, while above it the bright-end slope, β_{LF} , dominates. The dark gray shaded regions highlight where sources are detectable in our mock maps below L_* . This affects the $70\mu\text{m}$ – $160\mu\text{m}$ PACS bands at $z < 1$, the $450\mu\text{m}$ band moderately at $2 < z < 4$, and the 1.1 mm – 2 mm bands at $z > 4$.

most impacted by the choice of the faint-end slope, especially over the range of redshifts we care about, is 1.1 mm . Thus, the best lever-arm for calibrating the faint-end slope comes from the deepest 1.1 mm surveys to-date. Since this regime is especially suited for observations of ALMA, conducted on much smaller angular scales to much greater depth, we defer the reader to our accompanying paper on modeling the emission in ALMA deep fields for an in-depth discussion of constraints on α_{LF} . This includes the possibility that it evolves with redshift, whereas a steepening of the UV luminosity function towards very high redshifts (Bouwens et al. 2007, 2015; Reddy & Steidel 2009; McLure et al. 2013; Finkelstein et al. 2015; Song et al. 2016) might actually correspond to a flattening of the faint end of the IRLF at high-redshifts.

The one parameter that is unknown in our model but not explicitly allowed to vary is the evolutionary form of L_* beyond $z \sim 2$, i.e. the value of $\gamma_2 = 1$. We motivate the choice of $\gamma_2 = 1$ in the text, but wish to highlight here that we arrived at that choice based on testing a number of possible values. We found that all choices of $\gamma_2 \leq 0$ result in the underestimation of galaxy number counts at all wavelengths significantly dominated by sources at $z > 2$. Figure 14 summarizes the tension with observations if $\gamma_2 = 0$ is adopted; negative values of γ_2

show even more tension with observations. The only scenario in which the number counts of a $\gamma_2 \leq 0$ simulation match observations is one where Φ_* does not drop hardly at all past $z \sim 2$ (i.e. $\psi_2 > -1$); however, this possibility is now misaligned with other known characteristics of high-redshift galaxies. Mainly, this would call for the majority of $SFR \approx 10\text{--}50\text{ M}_\odot\text{ yr}^{-1}$ galaxies to be luminous at mm wavelengths, which other works have shown is not the case (Reddy et al. 2012; Bouwens et al. 2016).

If we instead suggest very high values of γ_2 , in excess of ~ 1.5 , we quickly run into an unphysical regime. In other words, L_* luminosities would be brighter than any observed galaxies in the Universe (including the brightest quasars) at somewhat modest redshifts, $z \sim 5$. Aside from the physical tension this causes, this would also introduce extremely bright (but rare) galaxies into the maps that do not exist in real maps. Our choice of $\gamma_2 = 1$ allows for continued modest evolution towards high- z , leading to an alignment of L_* values for the IRLF and the quasar LF at $z \sim 7$ (Hopkins et al. 2007).

It should also be made clear that our adopted values of z_{turn} are different for Model A and Model B. This is done explicitly so that the implied contribution of LIRGs, ULIRGs and total-IR to the cosmic star-formation rate density is well aligned with measurements at $z \lesssim 2.5$. If we adopted the same value, e.g. $z_{\text{turn}} = 2.1$ for both, then the number of $2 < z < 3$ DSFGs would be overpredicted in Model B and misaligned with the SFRD data measurements. This is simply a characteristic of the mathematical form we have adopted for the evolution of L_* and Φ_* . As is seen in Figure 6, the differences between the two models’ turnover redshift is not significant and designed to be very gradual.

A.2. Security of the SED Parameters

Our assumed SED model is extremely important in that it is the link between the luminosity function we wish to constrain and the observables we can constrain. Figure 3 provides the primary motivation for our SED model and is based on samples of galaxies with well-constrained SEDs from which L_{IR} and λ_{peak} can both be reliably measured. There are a few aspects of the SED model, that might come into question for a reader: the choice of mid-infrared powerlaw slope and emissivity spectral index, relationship between dust temperature and the shape of the SED, the possible bias in plotted samples (and the possibility that either very cold or very warm sources are systematically excluded), and whether or not there is any intrinsic redshift evolution underlying the model, or not.

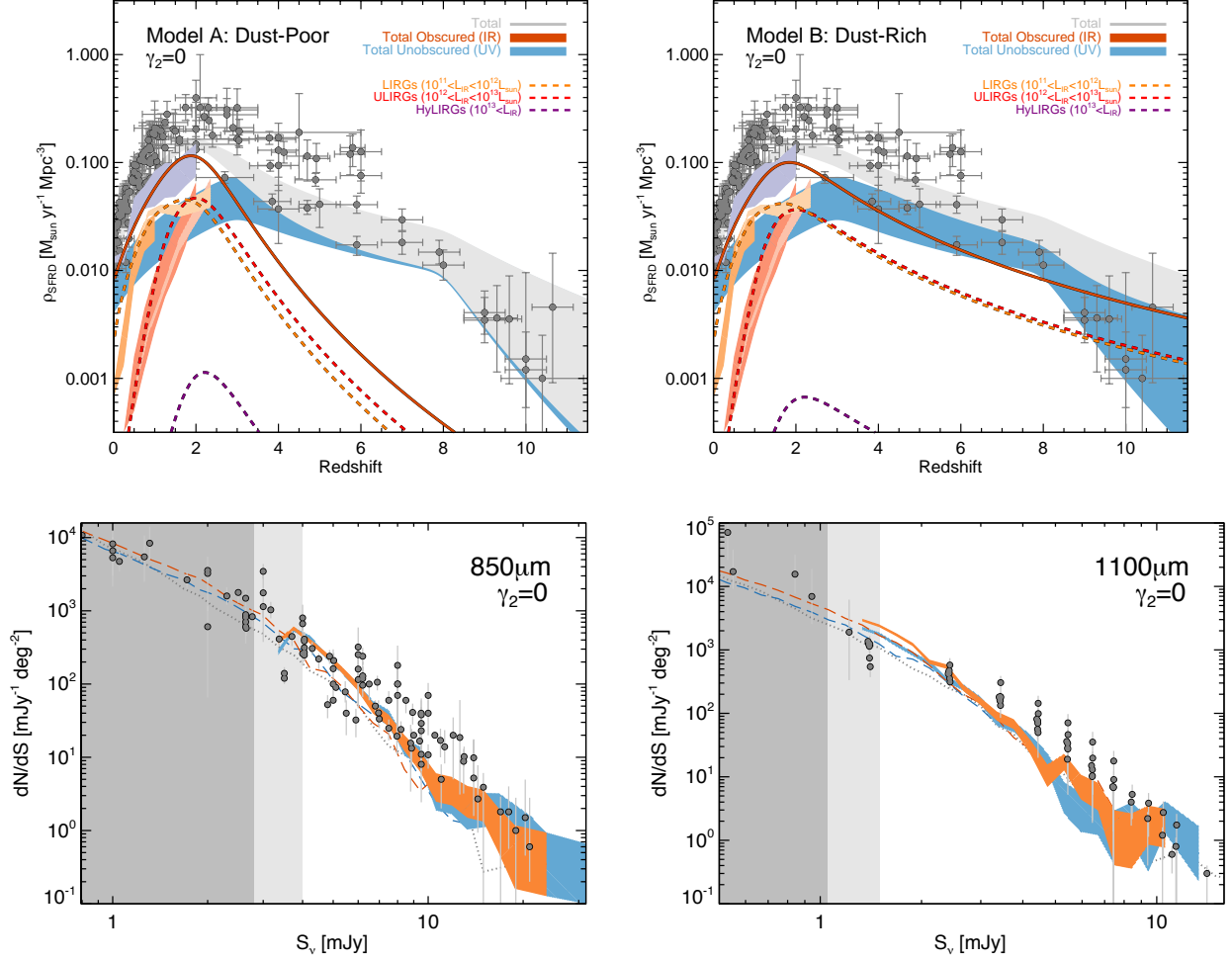


Figure 14. This figure summarizes the results of our simulation if an alternate value of γ_2 is adopted, i.e. if $\gamma_2 = 0$. γ_2 regulates the evolution of L_\star beyond the turnover redshift, $z \sim 2$. Our adopted value in the main text of the paper, $\gamma_2 = 1$, assumes further upward evolution of L_\star towards high-redshifts. At top, we note the subtle differences that $\gamma_2 = 0$ would cause in the contribution of different luminosity classes of DSFGs to cosmic star-formation (reference this against Figure 6). While the bottom two panels show the extracted number counts of the $\gamma_2 = 0$ simulations, highlighting the underdensity seen at both $850\mu\text{m}$ and 1.1 mm compared to the results of Figure 7. Matching number counts at 1.1 mm in particular would require almost no drop in the prevalence of DSFGs towards the highest redshifts ($\psi_2 > -1.5$).

The emissivity spectral index and mid-infrared power-law slope of galaxies’ rest-frame SEDs is fixed to $\beta_E = 1.8$ and $\alpha_{\text{MIR}} = 2.5$ to reflect measured characteristics of the best-measured DSFGs at low- and high-redshift, and the SEDs for dusty galaxies are assumed to be optically thick near the peak of the modified blackbody, such that $\tau = 1$ at $100\mu\text{m}$ in the rest-frame (a plausible assumption for the type of extreme star-formation galaxies that would be detectable in our simulated maps). We use the characteristics of Figure 3 to determine at which rest-frame wavelength a given galaxies’ SED is likely to peak (i.e. the parameters λ_0 , η , and $\sigma_{(\log \lambda_{\text{peak}})}$); this would hold whether or not the SED is assumed to be optically thick since it is defined in terms of the observable, λ_{peak} , and not the physical parameter, T_{dust} . We emphasize to

the reader again that the relationship between T_{dust} and λ_{peak} is highly dependent on the opacity assumptions for galaxies’ SEDs, and so we urge the community to appreciate that observations largely only constrain λ_{peak} , not the dust temperature of the ISM.

The bias of possible single-wavelength selection techniques has been a significant concern for any detailed study of the $L_{\text{IR}}-\lambda_{\text{peak}}$ relationship (Eales et al. 2000; Chapman et al. 2004; Casey et al. 2009; Kirkpatrick et al. 2017). The canonical $850\mu\text{m}$ SMG selection was originally thought to be strongly biased towards colder dust temperatures (Chapman et al. 2004; Casey et al. 2009) than what one might expect from the average DSFG population, given the prevalence of warmer-dust DSFGs in the nearby Universe. Indeed, *Herschel*

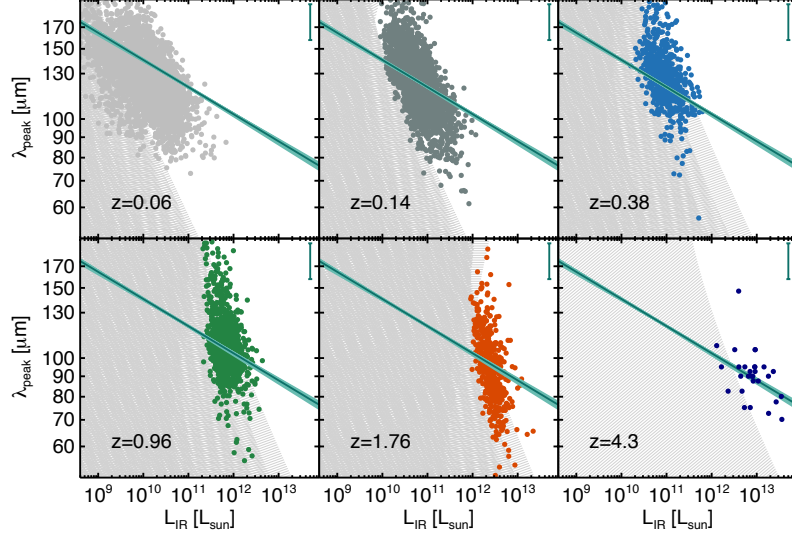


Figure 15. The data samples from Figure 3 split into six redshift bins and compared against their initial selection criteria. The hashed gray regions denote regions of parameter space that are less accessible to inclusion in the illustrated sample, as sources in that regime might fall below the detection limit of the survey. These detection boundaries are a function of dust temperature (or λ_{peak}), though generally selection with *Herschel*-SPIRE is not strongly dependent on λ_{peak} at redshifts where it brackets the peak of the dust SED.

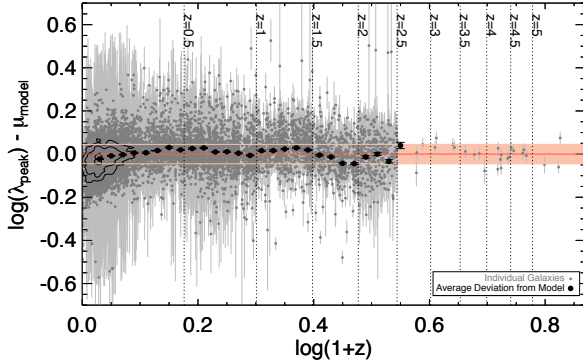


Figure 16. The deviation of individual sources from the model fit as a function of redshift for all sources shown on Figure 3. This explores possible deviations at certain redshift regimes, though we find no global evidence for an evolution of this relationship. Black points represent the median deviation from the model as a function of redshift, and the pink band represents the modeled 1σ spread in SEDs around the model.

selected galaxies revealed some warmer SEDs for DSFGs at similar redshifts, but the *Herschel*-detected sample also tends to sit at higher intrinsic luminosities, which could be attributed to the $L_{\text{IR}}-\lambda_{\text{peak}}$ relationship, and not a clear bias of $850\mu\text{m}$ surveys towards colder temperatures. In recent years multiple samples of high- z DSFGs selected at many wavelengths have demonstrated that higher redshift galaxies tend to be intrinsically colder at a fixed luminosity than their $z = 0$

counterparts (Simpson et al. 2017; Kirkpatrick et al. 2017). Still, the aggregate properties of the rest-frame SED shape of DSFGs beyond $z \sim 2$ are largely unconstrained. However, it is certainly reassuring that the highest-redshift, most complete sample of DSFGs studied to-date (with median redshift ~ 4.3), selected from the South Pole Telescope at 1.4 mm have SEDs that follow our adopted trend for lower redshift DSFGs $z \lesssim 2$ (see Figure 3; also Strandet et al. 2016).

We model the impact of the selection wavelengths and point source depth on this relation in Figure 15, which highlights the regions of parameter space that are mostly inaccessible to the sample at the given redshift due to its selection wavelength. The most severe temperature-dependent selection occurs for the H-ATLAS sample in the lowest redshift bin, where the SPIRE filters preferentially allow for detection of colder galaxies than warmer galaxies. However, it is in this bin that we actually observe a systematic median SED at *warmer* temperatures than the global best-fit (teal line). At other epochs, the selection is less biased with temperature overall. In all cases, we test to see if there could actually be no correlation between $L_{\text{IR}}-\lambda_{\text{peak}}$ and if the perceived relation is driven by selection effects; we find that our results are statistically inconsistent with this hypothesis, and the measurement of η as negative is very significant.

We explored possible redshift evolution of this $L_{\text{IR}}-\lambda_{\text{peak}}$ relationship but failed to find evidence for redshift evolution in all but the lowest redshift bin. Note that

this result does not appear to be discrepant with the findings of Kirkpatrick et al. (2017) who suggest there is evolution; a comparison with their dataset also suggests the bulk of said evolution is at $z < 0.5$ with little evidence for evolution beyond $z > 0.5$. One could invoke evolution in $L_{\text{IR}}-\lambda_{\text{peak}}$ which mimics the evolution of L_{\star} if a steeper value of η is adopted. Though this could match the measured temperatures for detectable galaxies, this invokes unrealistically cold SEDs for the vast majority of galaxies at low luminosities. Such cold temperatures artificially boost the long wavelength flux densities because their dependence on the SED dust temperature is very strong. Further investigation is needed to test whether or not the evolution at very low redshifts is real or if it is a different manifestation of a selection effect. However, it should be noted that these low redshift galaxies contribute negligibly to the maps generated in this analysis due to the rarity of DSFGs at the epoch overall (see Figures 6 and 8).

Figure 16 plots the difference in the modeled average peak wavelength λ_{peak} vs. redshift for all sources plotted on Figure 3. The y-axis represents the distance from the model teal line for each source. From this plot we determine there is no strong evidence for redshift evolution of the relationship.