

ULTRADEEP IRAC IMAGING OVER THE HUDF AND GOODS-SOUTH¹: SURVEY DESIGN AND IMAGING DATA RELEASE

I. LABBÉ², P. A. OESCH^{3,4}, G. D. ILLINGWORTH⁵, P. G. VAN DOKKUM⁴, R. J. BOUWENS², M. FRANX², C. M. CAROLLO⁶,
M. TRENTI⁷, B. HOLDEN⁵, R. SMIT⁸, V. GONZÁLEZ⁹, D. MAGEE⁵, M. STIAVELLI¹⁰, M. STEFANON²

Draft version July 31, 2015

ABSTRACT

The IRAC ultra-deep field (IUDF) and IRAC Legacy over GOODS (IGOODS) programs are two ultra-deep imaging surveys at $3.6\mu\text{m}$ and $4.5\mu\text{m}$ with the *Spitzer* Infrared Array Camera (IRAC). The primary aim is to directly detect the infrared light of reionization epoch galaxies at $z > 7$ and to constrain their stellar populations. The observations cover the Hubble Ultra Deep Field (HUDF), including the two HUDF parallel fields, and the CANDELS/GOODS-South, and are combined with archival data from all previous deep programs into one ultra-deep dataset. The resulting imaging reaches unprecedented coverage in IRAC $3.6\mu\text{m}$ and $4.5\mu\text{m}$ ranging from > 50 hour over 150 arcmin^2 , > 100 hour over 60 sq arcmin^2 , to ~ 200 hour over $5 - 10 \text{ arcmin}^2$. This paper presents the survey description, data reduction, and public release of reduced mosaics on the same astrometric system as the CANDELS/GOODS-South WFC3 data. To facilitate prior-based WFC3+IRAC photometry, we introduce a new method to create high signal-to-noise PSFs from the IRAC data and reconstruct the complex spatial variation due to survey geometry. The PSF maps are included in the release, as are registered maps of subsets of the data to enable reliability and variability studies. Simulations show that the noise in the ultra-deep IRAC images decreases approximately as the square root of integration time over the range $20 - 200$ hours, well below the classical confusion limit, reaching 1σ point source sensitivities as faint as of 15 nJy (28.5 AB) at $3.6\mu\text{m}$ and 18 nJy (28.3 AB) at $4.5\mu\text{m}$. The value of such ultra-deep IRAC data is illustrated by direct detections of $z = 7 - 8$ galaxies as faint as $H_{AB} = 28$. *Subject headings:* galaxies: evolution — galaxies: high-redshift

1. INTRODUCTION

Recent years have seen dramatic progress in studies of the early universe, in large part due to sensitive observations with the Wide Field Camera 3 (WFC3) on *HST* which detects the rest-frame UV light of distant galaxies. Studies now routinely identify large numbers of Lyman Break Galaxies (LBGs) in the first billion years of the universe (redshifts $6 < z < 8$) at the edge of the reionization epoch (e.g., Oesch et al. 2012; McLure et

al. 2013; Finkelstein et al. 2012; Grazian et al. 2012, Schmidt et al. 14). Recently, Hubble pushed the frontier even further, finding several galaxies at higher redshifts $z > 9$ (around 500 million years after the Big Bang, e.g., Bouwens et al. 2011, Zheng et al. 2012, Ellis et al. 2013, Oesch et al. 2014).

While HST is crucial for selecting the galaxies and determining the redshifts, *Spitzer*/IRAC (IRAC; Fazio et al. 2004) excels at detecting the infrared emission of high redshift galaxies. IRAC is currently the only instrument capable of measuring the rest-frame optical light of sources at $4 < z < 10$. The combination of Hubble and Spitzer has proven extremely powerful and provided estimates of the build up of the stellar mass density (e.g., Labbé et al. 2010, Gonzalez et al. 2011, Stark et al. 2013, Oesch et al. 2014, Duncan et al. 2014; Grazian et al. 2014) and the average specific SFR at $3 < z < 7$ (Gonzalez et al. 2010,2014, Stark et al. 2013, Steinhardt et al. 2014; Salmon et al. 2014). Comparing average IRAC colors of redshift $z \sim 4 - 8$ galaxies subsequently showed that star forming galaxies must exhibit very strong nebular emission lines, boosting the Spitzer/IRAC photometry (e.g., Schaerer & de Barros 2010, Labbé et al. 2010a,2010b,2013, Shim et al. 2011, Stark et al. 2013, Gonzalez et al. 2014, Smit et al. 2014). This realization has led to the first estimates of nebular emission line equivalent width at $z > 4$ and improved estimates of the stellar masses (e.g., Shim et al. 2011, Labbé et al. 2013, Stark et al. 2013), which is of vital importance for understanding the mass build up, feedback, and metal production in the earliest stages of galaxy formation. The current-best example of joint

¹ Based on observations made with the NASA/ESA *Hubble Space Telescope*, obtained from the data archive at the Space Telescope Science Institute. STScI is operated by the Association of Universities for Research in Astronomy, Inc. under NASA contract NAS 5-26555. Based on observations made with the *Spitzer Space Telescope*, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. Support for this work was provided by NASA through an award issued by JPL/Caltech.

² Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands

³ Yale Center for Astronomy and Astrophysics, Physics Department, New Haven, CT 06520, USA

⁴ Department of Astronomy, Yale University, New Haven, CT 06520

⁵ UCO/Lick Observatory, University of California, Santa Cruz, CA 95064

⁶ Institute for Astronomy, ETH Zurich, 8092 Zurich, Switzerland

⁷ School of Physics, University of Melbourne, Parkville, Victoria, Australia

⁸ Centre for Extragalactic Astronomy, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK

⁹ University of California, Riverside, 900 University Ave, Riverside, CA 92507, USA

¹⁰ Space Telescope Science Institute, Baltimore, MD 21218, United States

Hubble+Spitzer studies was the robust detection of a small sample of very bright $z \sim 10$ candidate galaxies and a first estimate of the galaxy stellar mass density at only 500 Myr after the Big Bang (Oesch et al. 2014). The joint *HST*+*Spitzer* Frontier Fields campaigns provided other examples of bright, lensed high redshift galaxies (e.g., Atek et al. 2014; Laporte et al. 2014; Zheng et al. 2014; Bradač et al. 2014)

Nevertheless, Spitzer/IRAC observations of earlier programs such as GOODS (PID 194; PI Dickinson) were only deep enough to individually detect a small fraction of the $z > 6$ sources. For example, Labbé et al. 2010b reported only 2/13 detected at $> 5\sigma$ from a sample of $H_{AB} < 27.5$ galaxies at $z \sim 7$ over the HUDF. Stacking was necessary to access typical $< L^*$ galaxies (e.g., Labbé et al. 2010a) as the $3.6\mu\text{m} - 4.5\mu\text{m}$ fluxes of individual sources were too low signal-to-noise (SNR) to be useful. In general, to extract meaningful information from the rest-frame optical SEDs, it is necessary to obtain SNR ratios of > 5 in each of the 3.6 and $4.5\mu\text{m}$ band for typical sources at $z > 7$.

To achieve this we initiated two ultradeep surveys in areas with existing ultradeep ACS+WFC3 data. The first was the cycle 7 IRAC Ultradeep Field (IUDF) program (PI Labbé; PID 70145) covering the HUDF/XDF and the two HUDF parallels to $\sim 50-100$ hours. The second was the IRAC Legacy over GOODS (IGOODS) program in cycle 10 (PI Oesch; PID 10076), which was aimed at filling out half of the GOODS-South and GOODS-North areas to ~ 200 hours depth, but which was only 10% completed before being terminated.

This paper described the survey design, data reduction, image quality analysis, and presents the public data release of the IUDF and IGOODS programs, after combining the two ultradeep programs with all archival data over GOODS-South. The paper is structured as follows: §2 describes the observations, section §3 summarizes the data reduction and introduces a new technique for creating PSF maps, §4 describes the resulting ultradeep IRAC mosaics, their properties, and simulations to test prior-based photometry, §5 discusses the role of IRAC photometry for high redshift galaxies, while a summary is provided in §6.

2. OBSERVATIONS

The IRAC surveys were all conducted in a single area of the sky, approximately centered on the HUDF in the GOODS-South field around $\alpha = 03 : 33$, $\delta = -27 : 48$. This field is very well suited for IRAC surveys as it has low infrared background and excellent visibility for Spitzer. GOODS-South and the HUDF enjoy the highest quality optical+NIR observations from Hubble (e.g., Giavalisco et al. 2004, Beckwith et al. 2006, Grogin et al. 2011, Koekemoer et al. 2011, Illingworth et al. 2013, Ellis et al. 2013). The high resolution imaging data at shorter wavelengths are necessary for detecting high redshift galaxies and determining their redshift from the location of the redshifted Lyman break. These HUDF data have resulted in some of the largest known samples of high-redshift $z > 7$ galaxies. As we shall see, the knowledge of the prior position and size of all sources in the field enables accurate modeling and extraction of the IRAC fluxes.

The GOODS-South field enables the maximum efficiency of any IRAC survey. The existing contiguous WFC3+ACS mosaic over scales of $10-15$ arcmin fills the full IRAC footprint. It also enables parallel $3.6\mu\text{m}$ and $4.5\mu\text{m}$ observations, which is relevant as high redshift studies require equally deep observations in both IRAC bands. Finally, very substantial investments in IRAC imaging have already been made in the GOODS fields (amounting to $\gtrsim 500$ hour per band) so it is more efficient to continue to build upon previous programs rather than starting from scratch.

Here we combine all programs to create single, contiguous ultra-deep images in the $3.6\mu\text{m}$ and $4.5\mu\text{m}$ bands. Below we discuss the individual programs that contributed to the data that were used to construct the field (dubbed "IRAC Ultra Deep Field", IUDF).

2.1. IRAC Ultradeep Field (IUDF)

The IUDF cycle 7 program integrated for 210 hours in both IRAC filters, covering the HUDF/XDF WFC3 field of the HUDF09 survey (PI Illingworth), including its two flanking fields HUDF09-1 and HUDF09-2. These fields are unique due to the concentrated investment of HST time and the large existing samples of ~ 190 $z > 7$ galaxies available immediately for study (Bouwens et al. 2014).

While the HUDF was previously covered with IRAC with 46 hours of cryogenic observations from GOODS (PI Dickinson), the parallel HUDF1 and HUDF2 had received limited and uneven coverage. The IUDF solves this by observing both HUDF parallels to $50 - 100$ hour at $3.6\mu\text{m}$ and $4.5\mu\text{m}$, while using roll angle constraints to obtain deeper imaging on the HUDF/XDF, increasing the exposure time to $100 - 120$ hour at $3.6\mu\text{m}$ and $4.5\mu\text{m}$. The HUDF + parallels are the deepest-ever ACS+WFC3+IRAC of any field on the sky.

2.2. IRAC Legacy over GOODS (IGOODS)

The completion of the IUDF and the success of the first joint ultradeep WFC3+IRAC analyses in the HUDF/XDF (e.g., Oesch et al. 2012, 2013, Labbé et al. 2013) demonstrated the scientific value of deep IRAC data as well as the feasibility of ultradeep studies. However, much larger samples to even deeper limits are needed for a proper characterization of the $z > 7$ universe.

The IGOODS cycle 10 aimed to achieve this by increasing the IRAC depth to a homogenous 200 hours per sky position, while covering much larger areas ~ 200 arcmin² in GOODS-South and GOODS-North. These depths and areas are a sweet spot: sensitive enough to provide direct detections of sub- L^* star forming galaxies at $z \sim 8$, while providing enough area for large samples and good statistics (> 200 galaxies at $z > 7$ with $> 5\sigma$ IRAC photometry).

Of the approved 800 hours, 200 were earmarked as higher priority to demonstrate the feasibility and usefulness of IRAC data to these limits over the HUDF and GOODS-S. Even though less than 10% (< 70 hour) of the program was executed before the program was terminated due to scheduling conflicts, the program was

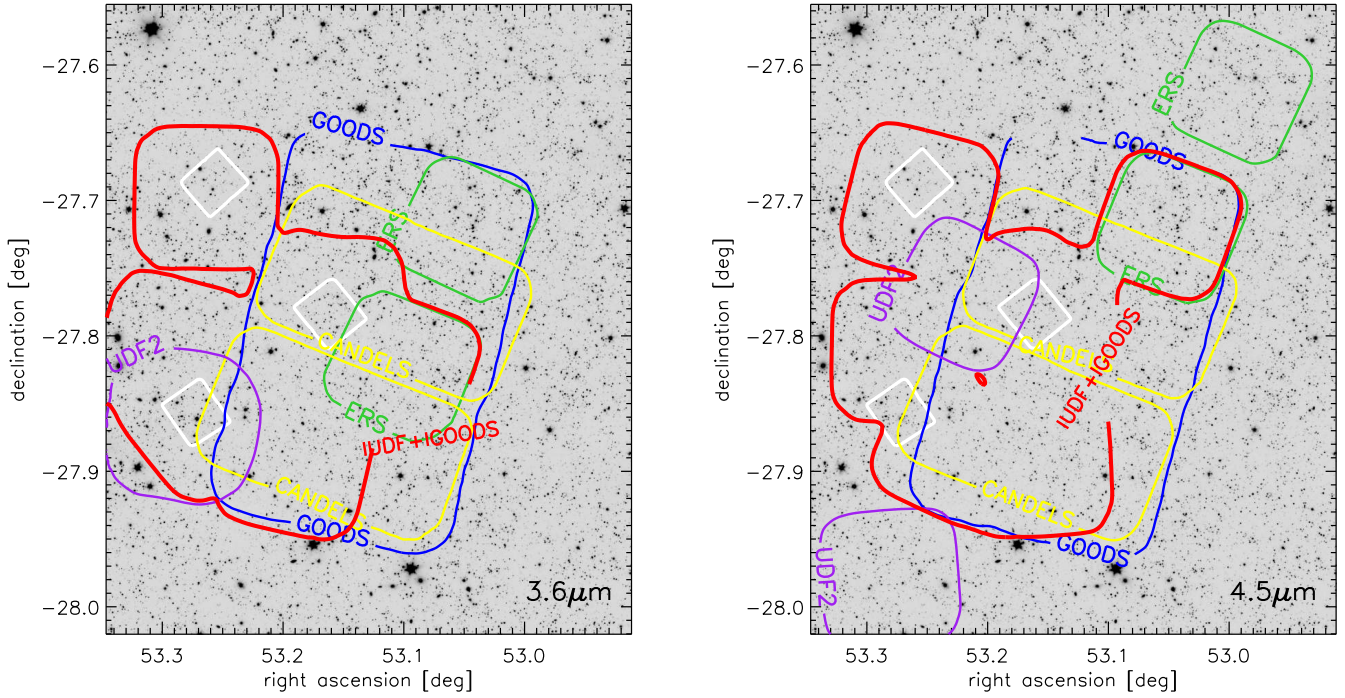


Figure 1. Layout of the IUDF and IGOODS observations (*red*) on top of the IRAC imaging at $3.6\mu\text{m}$ (*left*) and $4.5\mu\text{m}$ (*right*) from SIMPLE (Damen et al. 2011). Also shown are all other ultradeep IRAC observations used in this paper, including warm mission data from ERS (*green*), S-CANDELS (*yellow*), and cryogenic data from GOODS (*blue*) and UDF2 (*purple*). Table 1 lists the all programs and PIs. The IUDF observations cover the HUDF/XDF and the two parallel fields (*white*), while IGOODS fills out part of the GOODS-South area.

Table 1
Summary of IRAC observations

program	PID	PI	max exp.(h) ^a	# pointings	total exp.(h)	# frames	SSC pipeline version ^d
IUDF	70145	Labbé	100	3	215.3	8280	S19.0.0/S18.18.0
IGOODS	10076	Oesch	46	2	65.5	2520	S19.1.0
GOODS	194 ^c	Dickinson	46	8	180.4	3356	S18.25.0
ERS	70204	Fazio	75	2	162.9	6264	S18.18.0
S-CANDELS	80217	Fazio	25	4	101.1	3888	S19.0.0/S19.1.0
SEDS	60022	Fazio	12	20 ^b	209.3 ^b	8051	S19.0.0/S18.18.0
UDF2	30866 ^c	Bouwens	28.1	1	28.1	1080	S18.25.0
total					962.6	33439	

Note. — Program PID 20708 was omitted because the exposure time is negligible over the central parts of the GOODS-S region.

^a Maximum exposure time per position on the sky per channel.

^b Only the central $\sim 60\%$ of the full SEDS data are used.

^c Cryogenic mission observations; all other programs are warm mission.

^d The calibration pipelines used were the most recent available from the Spitzer heritage archive at the time of writing. No significant changes since S18.18.0 have been reported for 3.6 and $4.5\mu\text{m}$ observations.

Table 2
Summary of individual AORs

PID	AOR key	MJD ^a	area ₅₀ ^b	<exptime> ₅₀ ^c
70145	40849920	55487.9259899	41.5	1.41
70145	40850176	55493.6466019	36.5	1.55
70145	40850432	55493.5162173	36.1	1.57
70145	40850688	55611.5557390	34.4	1.61
70145	40850944	55603.7885426	36.5	1.56

Note. — Table 2 is published in its entirety in the electronic edition of ApJS. A portion is shown here for guidance regarding its form and content.

^a Modified Julian Day (JD-2400000.5) in UTC at start of observation

^b Total area in arcmin² with > 50% of the maximum exposure time on sky.

^c Mean exposure time in hour over area₅₀.

successful in one aspect. By placing the observations on areas with the deepest overlapping coverage from archival data, it produced the first > 150 hour deep data in two separate 25 arcmin² fields in the central part of GOODS-S.

2.3. Archival data

Apart from the IGOODS and IUDF programs, there exists a wealth of ultradeep IRAC archival data from various programs (most of which are discussed in, e.g., Ashby et al. 2013, Ashby et al. 2015¹¹). Table 1 provides an overview of the programs, the respective PIs, the number of exposures and total integration time. We downloaded all data from the Spitzer Heritage Archive and combined them with our data sets, reducing all in a consistent manner, and coadding them into one ultradeep mosaic. The 7 programs are divided up in 353 Astronomical Observation Requests (AORs), consisting of 33439 exposures, and totaling 3.47 Ms (962.6 hours) in each of the 3.6 μ m and 4.5 μ m filters for a total of 1925 hours of IRAC data. At the deepest location the coverage reaches \sim 220 hours at 3.6 μ m and \sim 190 hours at 4.5 μ m over an area of \sim 5 arcmin².

3. REDUCTION

The reduction of the IRAC data was carried out starting with the corrected Basic Calibrated Data (cBCD) generated by the Spitzer Science Center (SSC) calibration pipeline. A custom pipeline written by IL was used to post-processes and mosaic the cBCD frames. The reduction pipeline was also used for reducing the SIMPLE IRAC Legacy Survey (PI van Dokkum) and described in detail in Damen et al. 2011.

3.1. IRAC Reduction Process

¹¹ We note that Ashby et al. (2015) present different reductions of very similar observations as described here. We note several key differences: 1) we do not include the shallow and wide field PID 81 and PID 20708 data, but we do include the deep IGOODS PID 10076 observations, 2) reduction and interpolation method are a weighted sum on 0''.6 pixel scale in Ashby et al. (2015) versus Drizzling on 0''.3 here, and 3) the release in this paper of PSF maps corresponding to the reduced mosaics.

The reduction uses a two pass procedure. The first pass comprises background structure removal, artifact correction, persistence masking, and a first-pass coaddition. First, a median image is constructed from all frames in the AOR, to remove background or bias structure and artifacts, and it is subtracted from each frame¹². Then the cBCDs are inspected and additional artifacts are corrected. The most important effect is residual column-pulldown and pull-up. The pull-up/down, caused by bright stars or cosmic rays at levels > 10 – 20 MJy/Sr in 3.6 μ m and 4.5 μ m, shifts the intensities of the column above and below in slightly different ways. We correct for it by subtracting a median above and below the affected pixels after excluding any sources. Persistence from very bright stars, leaving positive residuals on subsequent readouts of the array, is masked by rejecting all highly exposed pixels in the subsequent 4 frames (\approx 400s). A constant background pedestal is determined and subtracted from each frame, by iteratively masking pixels associated with sources and determining the mode of the remaining background pixels.

Finally the post-processed cBCD frames of each AOR are registered and median combined and a Median Absolute Deviation (MAD) map is calculated (reflecting the uncertainty in the combined output pixels). The data are very well dithered, hence the images are free from deviant pixels and can be used to create an object mask.

The second pass comprises cosmic ray rejection, astrometric calibration, background structure removal, and a final coaddition. First, the first-pass median image is de-registered and subtracted from each frame. The difference images are divided by the MAD uncertainty image and used as detection maps for cosmic rays and hot/cold pixels. Pixels are flagged if they deviate more than $4.5\sigma_{MAD}$, while pixels adjacent to outliers are iteratively clipped at a more aggressive $> 2.5\sigma_{MAD}$ threshold. The first-pass image is also used to calibrate the astrometry. The frames in an AORs are corrected for a simple shift in RA and Dec using sources in common with the deep WFC3 maps of 3D-HST (Skelton et al. 2014). These maps are used as they convenient as they include the WFC3 observations of the CANDELS/GOODS-South, the WFC3 ERS, and the HUDF + parallel fields. The rms residuals of individual IRAC source positions is 0''.05 – 0''.07 rms with systematic differences on scales of a few arcmin of \lesssim 0''.02. The Skelton et al. (2014) astrometry was calibrated to the CANDELS/GOODS-South (Koekemoer et al. 2011) mosaics and to the GEMS (Rix et al. 2004) mosaic for the HUDF parallel fields.

A new median background structure map is created from all frames in the AOR, this time masking objects and outlier pixels. The frames are then drizzled (Fruchter & Hook 2002) per AOR using a pixfrac=0.2 on reference grid defined by the CANDELS tangent point and a fine 0''.3 pixel scale. A final background was subtracted by iteratively clipping pixels belonging to objects and subtracting the mode of the background pixels. Finally, the AORS are weighted by the exposure

¹² This procedure works well for the IUDF, IGOODS, GOODS, and UDF, which take one frame per dither position, but not for SEDS, S-CANDELS, and ERS, which make use of in-place repeats. This leads to different bias patterns in the “first frame” and the “repeat frame” of each dither. We subtract these by creating two median images, one for all first frames and one for all repeat frames.

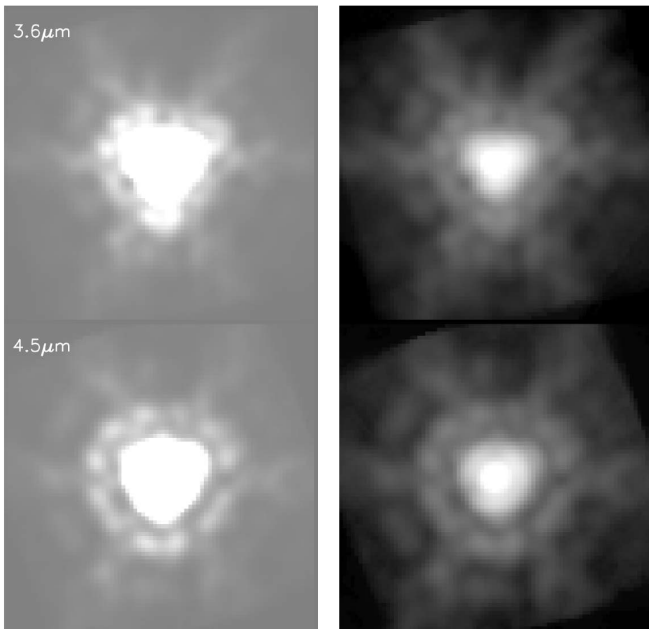


Figure 2. The empirical template PSF at $3.6\mu m$ and $4.5\mu m$ created from stacked images of stars spread across the field from all 353 AORs. The left column shows the PSFs with linear scaling, the right column with a logarithmic scaling to capture the entire dynamic range and highlight the core structure as well as the PSF wings. The images are $24''.4 \times 24''.4$ and the PSF is sampled on a $0.3''$ grid ($\sim 1/4$ th native IRAC pixel).

time per pixel and combined into the ultradeep mosaic. The cryogenic observations (GOODS and UDF2) data sets at $3.6\mu m$ are $\sim 30\%$ more sensitive than those of the warm mission hence we increase their contribution to the final mosaic and exposure time maps by a factor 1.7. There are no significant differences in sensitivity in $4.5\mu m$. The data release includes the full-depth mosaics in both $3.6\mu m$ and $4.5\mu m$, as well as mosaics for each AOR in both filters (353 total, on the same grid and final mosaic position angle).

3.2. Point-Spread Function (PSF) Construction

Accurate point spread functions are needed to facilitate IRAC photometry using PSF fitting techniques or using the high resolution HST imaging as a prior. Empirical PSFs created from the reduced mosaics are preferable, as the observation and reduction processes change the PSFs in subtle ways. However, extracting clean PSFs to large radii and high dynamic range is challenging due to crowding of neighboring sources and the small number of stars usually available in deep blank fields. To complicate matters, the layout and different rotation angles of the AORs cause the effective PSF of the combined mosaic to change rapidly on small spatial scales.

To solve this we generate a spatially varying IRAC PSF. First we take advantage of the optical stability and the fine sampling to generate one template “super PSF” at $3.6\mu m$ and $4.5\mu m$. Two hundred stars were identified in deep HST imaging based on their FWHM and magnitude (e.g., Skelton et al. 2014) and requiring an axis ratio of $b/a > 0.85$. At corresponding locations in each of the 353 AOR mosaics (which are on the same

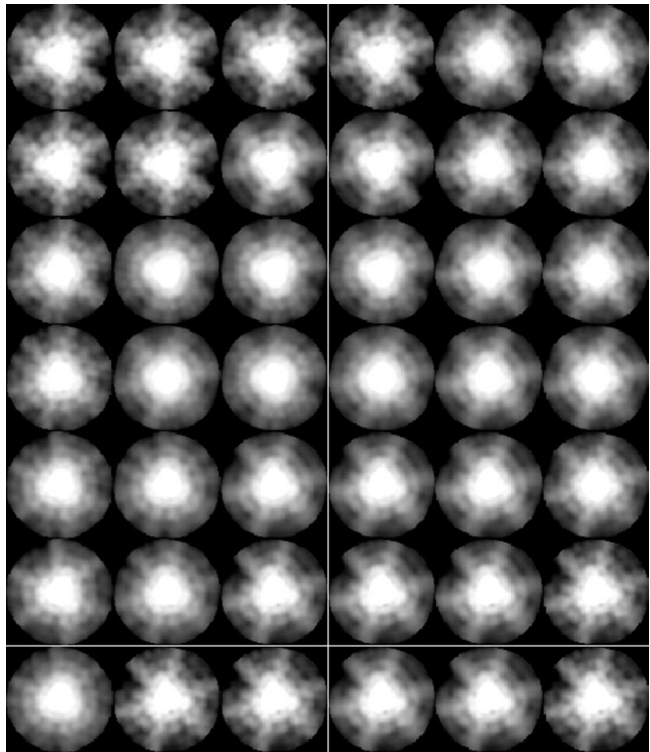


Figure 3. The reconstructed $3.6\mu m$ PSF mapped on a coarse grid in steps of $2.5'$, highlighting the spatial variation over the $12.5' \times 15'$ central area. The PSFs map is created by rotating and combining the template PSFs in the same way as the science data.

grid and PA as the full-depth mosaic), image stamps of the stars were extracted to $R = 20''$ radius. Saturated star images and those with $SNR < 300$ were rejected. The remaining 2050 star images were then rotated to the native orientation of the IRAC frames to align the PSF features. Subsequently the images were normalized and median stacked, sigma clipping outlier pixels due to neighboring objects. The stacking was iterated three times while growing the outlier masks by 1 pixel in each iteration. Note that some stars are imaged in more than 100 distinct AORs. Therefore the distribution of position angles causes objects close to the stars to fall on different locations on the IRAC frames. This makes it easier to separate between true PSF structure and faint signal from neighboring sources, turning the complex nature of the observations into an asset.

The resulting template PSFs are shown in Figure 2 and are of much higher quality and SNR than usual for deep extragalactic fields. The drizzling on a fine pixel scale of $0''.3$ helps to recover high frequency features of the PSF, while the large number of high SNR images results in a dynamic range of $> 10,000$.

The second step is to combine the template PSF in such a way that simulates the combination of the AOR into the full-depth mosaic. We map the exposure time and rotation angles of each AOR on a fine grid ($12''$) covering the output image. Then we reconstruct the effective full-depth PSF, by rotating¹³ and weighting the

¹³ Rotation and bicubic interpolation of the template PSF intro-

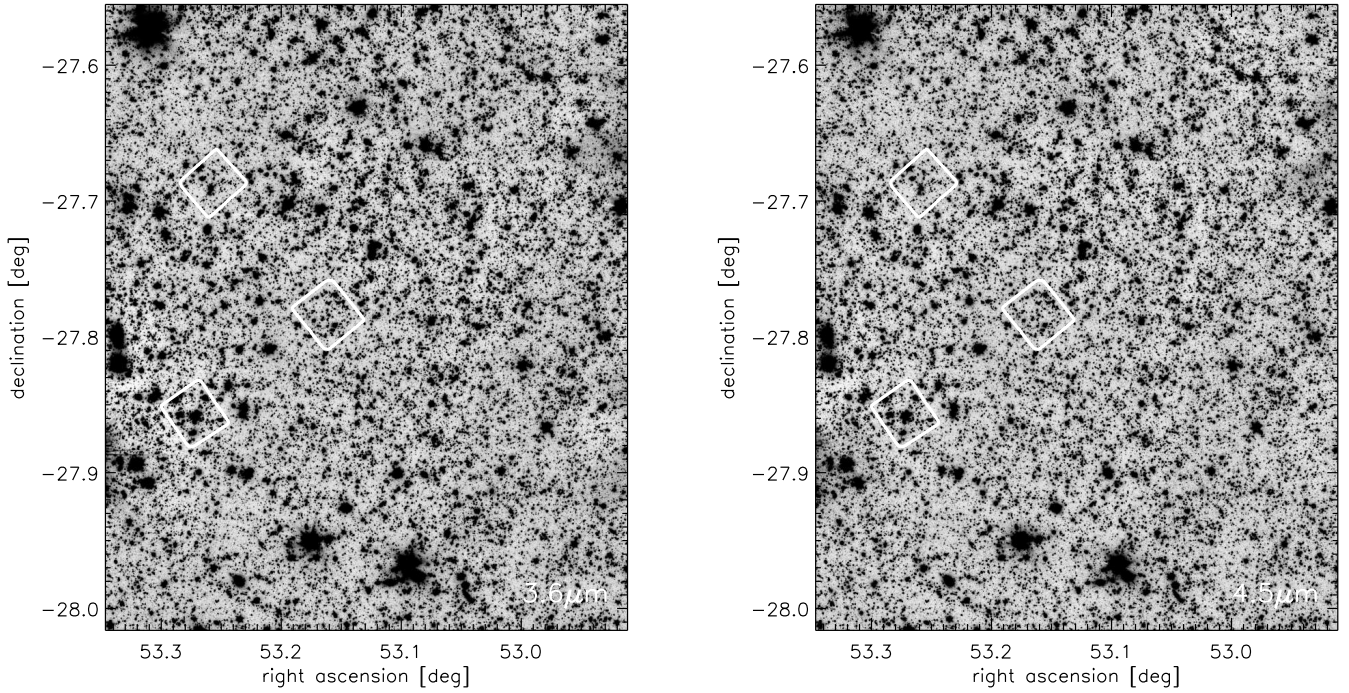


Figure 4. The full IRAC mosaics over GOODS-South and the HUDFs at $3.6\mu\text{m}$ (*left*) and $4.5\mu\text{m}$ (*right*), shown in inverted linear grayscale from -7 to 7 nJy / pixel (-0.003 to 0.003 MJy sr $^{-1}$). Each mosaic consists of 33439 exposures totaling 962.6 hours of observations. Shown in white are the locations of the HUDF/XDF and the two parallel fields.

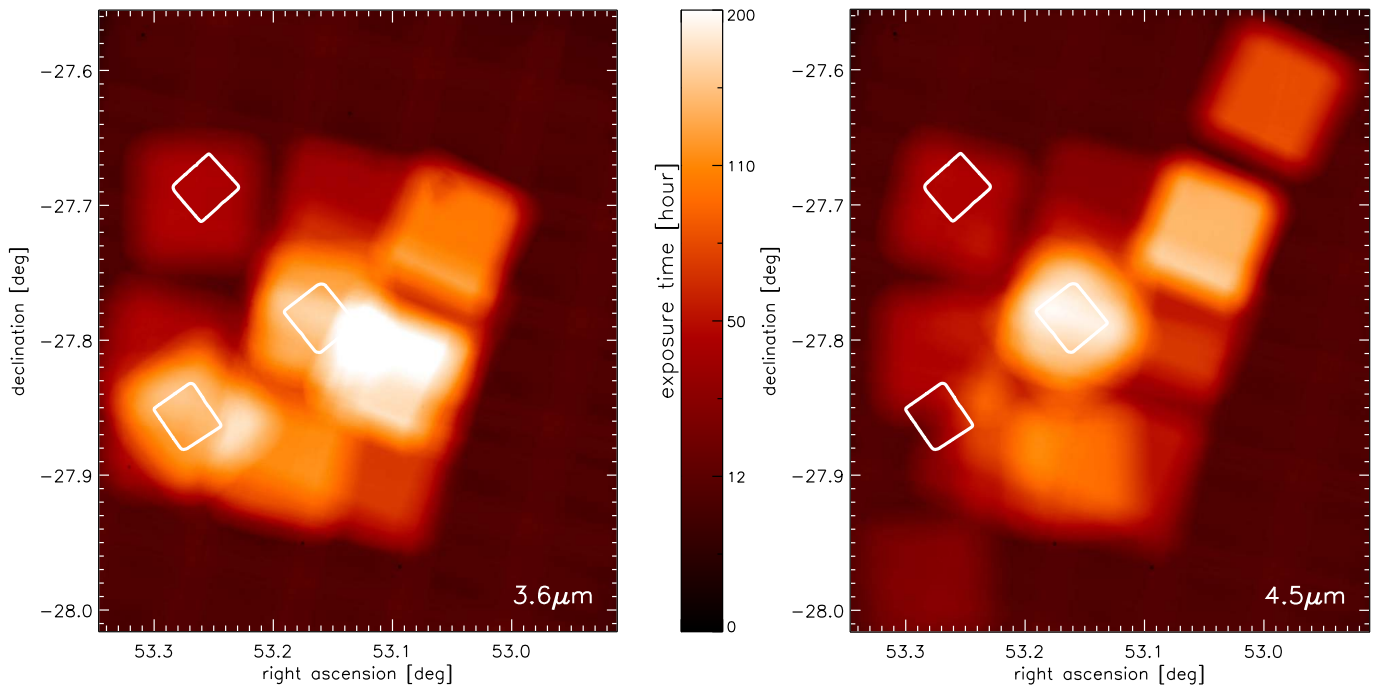


Figure 5. The IRAC coverage maps in GOODS-South and the HUDF fields, shown in heatmap scaling from 0 to 200 hours using a square root stretch. Targeted observations from IUDF and IGOODS and additional fortuitous overlap from many previous IRAC surveys yield total integration time exceeding > 100 hours over 60 arcmin 2 and $> 180 - 200$ hours over $\sim 5 - 10$ arcmin 2 .

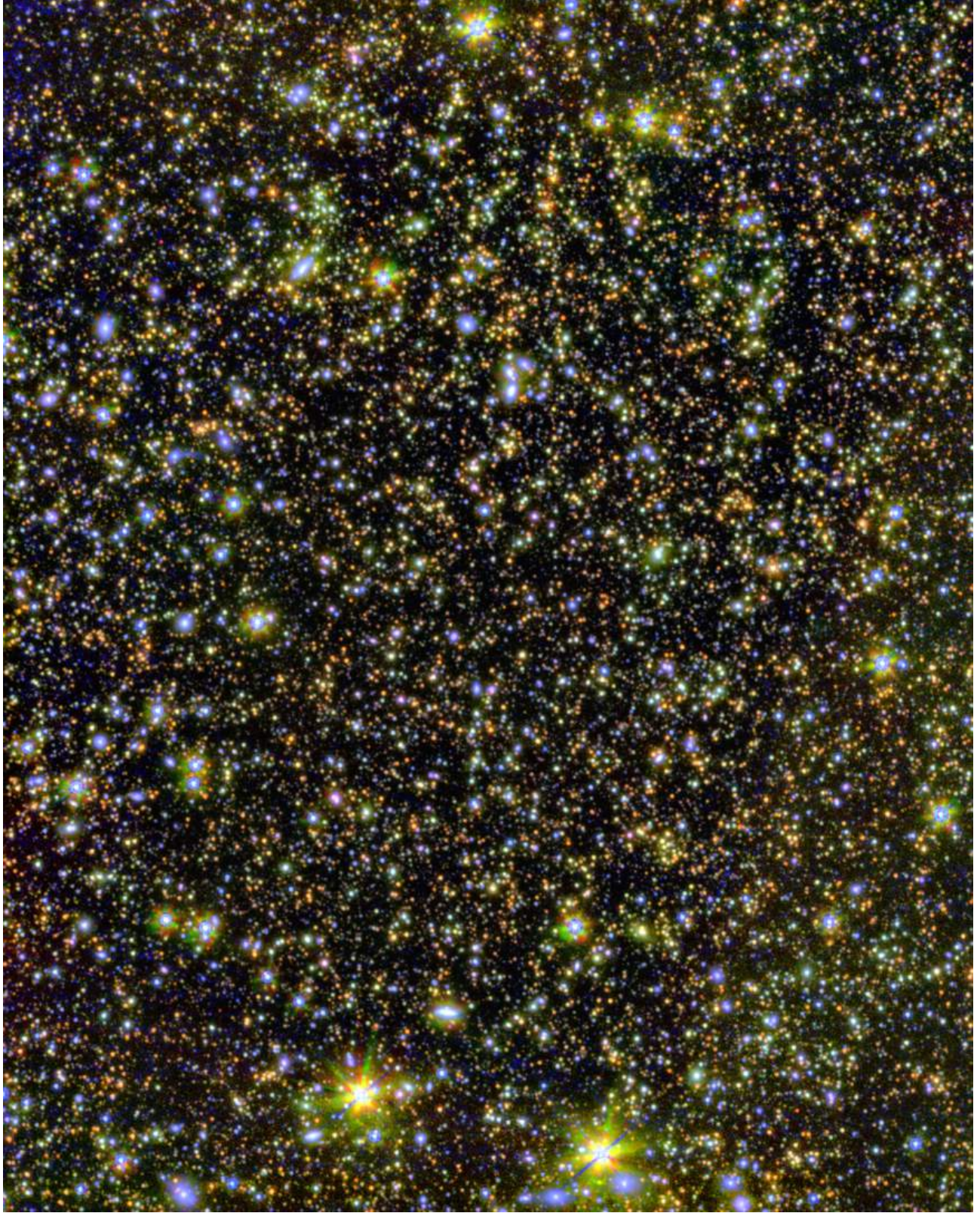


Figure 6. A color composite image of the central deepest region of the GOODS-S field. Deep K_s -band data from the TENIS (Hsieh et al. 2012) and HUGS (Fontana et al. 2014) programs are shown as blue, IUDF $3.6\mu\text{m}$ is green, and $4.5\mu\text{m}$ is red. The field size is $18' \times 22'$ and North up is up.

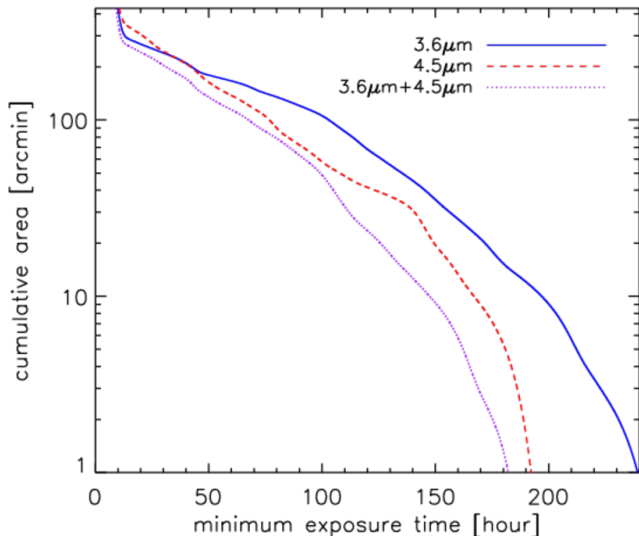


Figure 7. Area covered versus exposure time for all data over GOODS-South and HUDF fields at $3.6\mu\text{m}$ (blue solid), $4.5\mu\text{m}$ (red dashed), and joint $3.6\mu\text{m}$ and $4.5\mu\text{m}$ (purple dotted). The uncoordinated nature of the various programs contributing to the ultra-deep mosaics causes the area covered in both bands to be much smaller than the area covered at $3.6\mu\text{m}$ or $4.5\mu\text{m}$.

template PSF for each AOR contributing to that grid location.

Figure 3 shows the reconstructed PSFs in steps of 2.5 arcmin, illustrating the strong spatial variation. Bootstrap resampling the star list and repeating the process results in uncertainties much smaller than the spatial variation in constructed PSF. This indicates that survey geometry has a much larger impact on the effective IRAC PSF than the intrinsic variation of the PSF over a single IRAC pointing. Both the super PSFs and the maps are made available in the data release.

4. RESULTS

4.1. Reduced Image Properties

The reduced IRAC mosaics are shown in Figure 4 and the corresponding coverage maps are shown Figure 5. A color composite using K_s -band, $3.6\mu\text{m}$ and $4.5\mu\text{m}$ is shown in Figure 6. The combined observations of all previous programs results in extremely deep coverage, due in part to targeted observations over the HUDF/XDF from the IUDF and IGOODS programs, and in part from fortuitous overlap from archival data. The uncoordinated nature of the programs is revealed by the much smaller area covered in both filters simultaneously: the area is smaller by a factor of > 2 at > 100 hr and factors of > 5 at > 150 hours). Simultaneous coverage is crucial for placing constraints on emission line strengths and stellar masses at $z > 7$ (e.g., Labbé et al. 2013). Presently, two small ultra-deep (180 – 200 hr) areas in GOODS-S exist (9 arcmin² in $3.6\mu\text{m}$ and $4.5\mu\text{m}$ each).

The final mosaics are cosmetically clean and the background is flat to $5 \times 10^{-5} \text{MJy sr}^{-1}$ (~ 31 mag/arcsec² AB) on scales of 1 arcmin. The small area that reaches to 180 – 200 hours allows us to evaluate the improvement in background noise relative to the existing deep 25 hour integrations. As illustrated in Figure 8 the improvement is obvious in both IRAC bands, with large increases in the number of detected ultrafaint sources

and in the SNRs of brighter objects.

The image quality of the full depth mosaics is excellent and constant over the field. The 1-D gaussian full-width at half-maximum (FWHM) over the field is $1''.49 \pm 0.015$ at $3.6\mu\text{m}$ and $1''.48 \pm 0.025$ at $4.5\mu\text{m}$. These values are identical to those of the cryogenic GOODS v0.3 public data release, and 20% smaller than those of the SEDS (PID 60022; Ashby et al. 2013) and SIMPLE mosaics (PID 20708; Damen et al. 2011). The difference with the latter two programs is due to the native IRAC pixels undersampling the PSF and using drizzling instead of interpolation when resampling the IRAC frames.

We verify the photometric calibration by comparing the fluxes of bright sources (< 20 mag AB) in $5''$ diameter aperture to earlier measurements. The agreement with the IRAC $3.6\mu\text{m}$ and $4.5\mu\text{m}$ imaging of the Spitzer Extended Deep Survey (SEDS; Ashby et al. 2013) is excellent ($< 1\%$ offset). Comparing to cryogenic GOODS-S imaging (PID 194, PI Dickinson, data release DR3) reveals that the GOODS fluxes are brighter by 8% and 2% in $3.6\mu\text{m}$ and $4.5\mu\text{m}$ respectively. This is due to a change in BCD pipeline calibration: the FLUXCONV values reported in the PID 194 headers (GOODS DR3, v0.30/v0.31, BCD pipeline S10.5.0) are 7% and 1% brighter than the FLUXCONV values in the most recent calibrations of the same data (BCD pipeline version S18.25.0). Comparisons to our own reduction of the recalibrated GOODS data shows no offset.

4.2. Photometry and Confusion

The total integration times of the mosaics (50 – 200 hours) run well into the classical “source confusion” regime for low background extragalactic observations, where crowding by nearby sources affects the reliability of photometry. The classical confusion limit predicted by Franceschini et al. (1991) is $0.6\mu\text{Jy}$ (24.5 AB mag), but in reality confusion is not a hard limit. For example, the classical limit is strictly speaking not relevant when the positions of the sources are known a priori. In GOODS-South and the HUDFs deep ($H_{AB} = 27 - 30$), high-resolution (FWHM = $0''.16$) HST/WFC3 imaging is available and the IRAC images are registered to the WFC3 images to very high accuracy ($\lesssim 0''.02$ systematic). Using the source positions and sizes in the high resolution image, combined with knowledge of the PSFs of WFC3 and IRAC, it is possible to extract the source flux by modeling the IRAC surface brightness distribution. Although surface brightness distribution can vary with wavelength, such procedures already greatly reduce the effect of confusion and open up the possibility of extracting fluxes well beyond the classical limit.

Prior based photometric techniques on blended sources and multi-resolution data sets have been used by many groups in the past with good results (e.g., Fernandez-Soto et al. 1999, Papovich et al. 2001, Shapley et al. 2005, Labbé et al. 2005, 2006, 2010, 2013, Grazian et al. 2006, Wuyts et al. 2007, DeSantis et al. 2007, Laidler et al. 2007). As demonstrated in Figure 9 these techniques can work extremely well. Note that the photon noise for most sources is negligible compared to the background noise. Therefore, when sources can be modeled and subtracted perfectly, most of the field can be considered empty sky from the perspective of faint source detection.

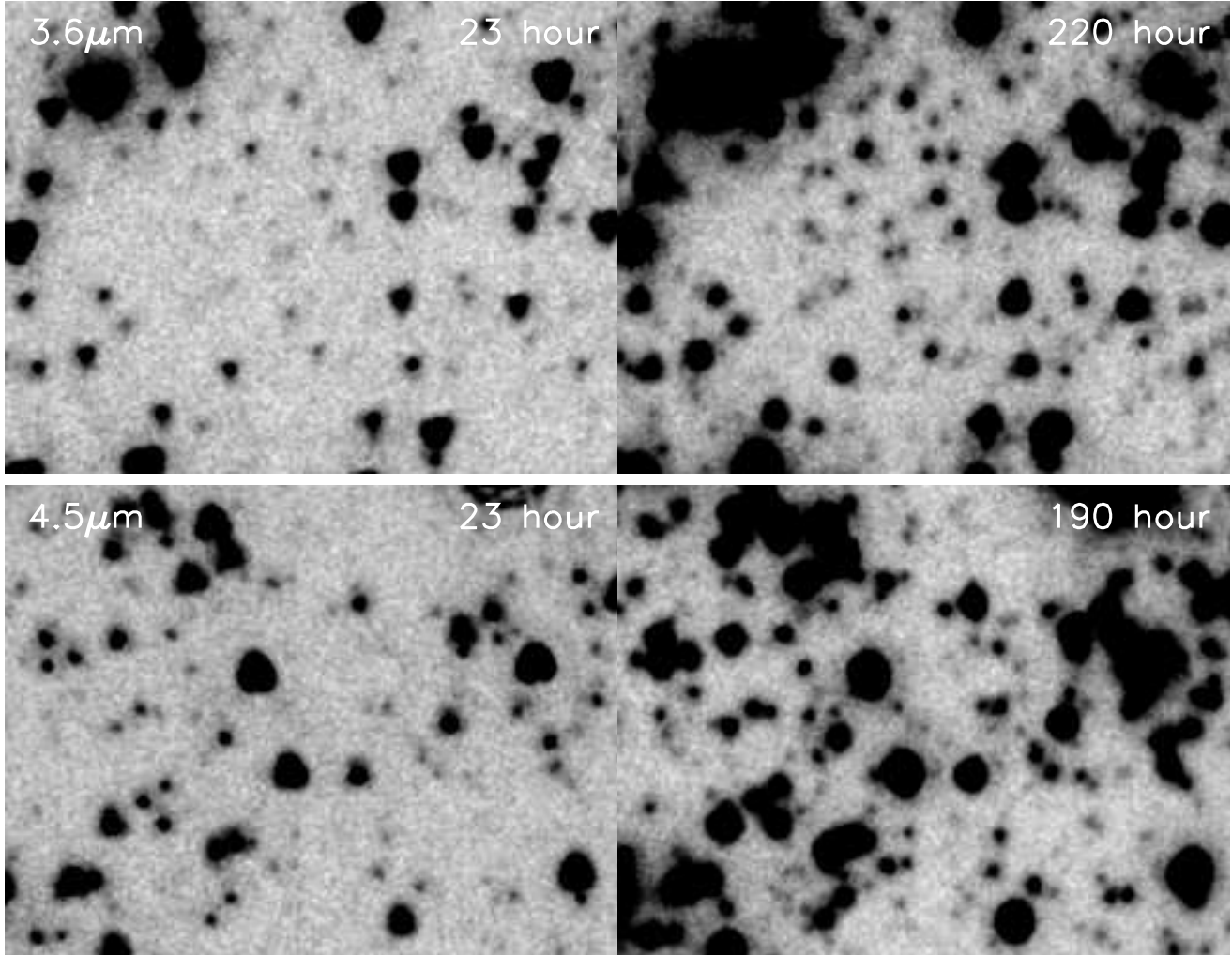


Figure 8. A comparison of Spitzer/IRAC band images of 23 hour exposure time (GOODS program single epoch; *left*) and the new ultradeep imaging at ~ 200 hours of this paper (*right*). Different $1.5' \times 1.0'$ locations are shown for $3.6\mu m$ (*top*) and $4.5\mu m$ (*bottom*). Image panels are shown in inverted linear grayscale keeping the background noise at a constant level. The stretch used is -9 to 9 nJy / pixel (-0.0042 to 0.0042 MJy sr^{-1}) at 23 hours and -3 to 3 nJy / pixel (-0.0014 to 0.0014 MJy sr^{-1}) at ~ 200 hours. A large improvement in signal-to-noise ratio with increased exposure time is visible and a larger number of faint detected sources.

While good results can already be obtained by simple PSF fitting (i.e., assuming point sources and a negligible size of the high resolution WFC3 PSF), for the best results and smallest residuals near the cores of bright sources, it is necessary to account for both the source size and the detailed shape of the WFC3 and IRAC PSF. This can be done by convolving the isolated high resolution object by a kernel, constructed by deconvolving the low resolution PSF by the high resolution PSF (e.g., Labbé et al. 2003, Labbé et al. 2005).

4.3. Depth

The large variation in integration time makes it possible to study the relation between sensitivity and integration time using prior based photometry. We measure the sensitivity limits of the IRAC images by placing artificial sources of zero flux on 15,000 random locations in the mosaic and extracting their flux using the WFC3 image

as a prior, as previously described and shown in Fig. 9. To enable straightforward comparisons with other noise measurements, we do not use the best-fit flux directly but subtract the best-fit model of all neighbors to give a “cleaned” image of the source. Then we measure the unweighted flux in $D = 2''0$ diameter circular apertures (without further corrections for light outside the aperture).

The histograms of extracted fluxes are shown in Figure 10), grouped in bins of integration time. As expected, the scatter histogram becomes progressively narrower with increasing integration time, with no evidence for bias even at the largest integration times. To compare to the scatter expected from pure background noise, we compute for each fake source the local background RMS in empty regions of the residual image (away from bright sources). We bin by 6×6 pixels ($1''.8 \times 1''.8$) to approximate the area of a $D = 2''0$ aperture. The local empty

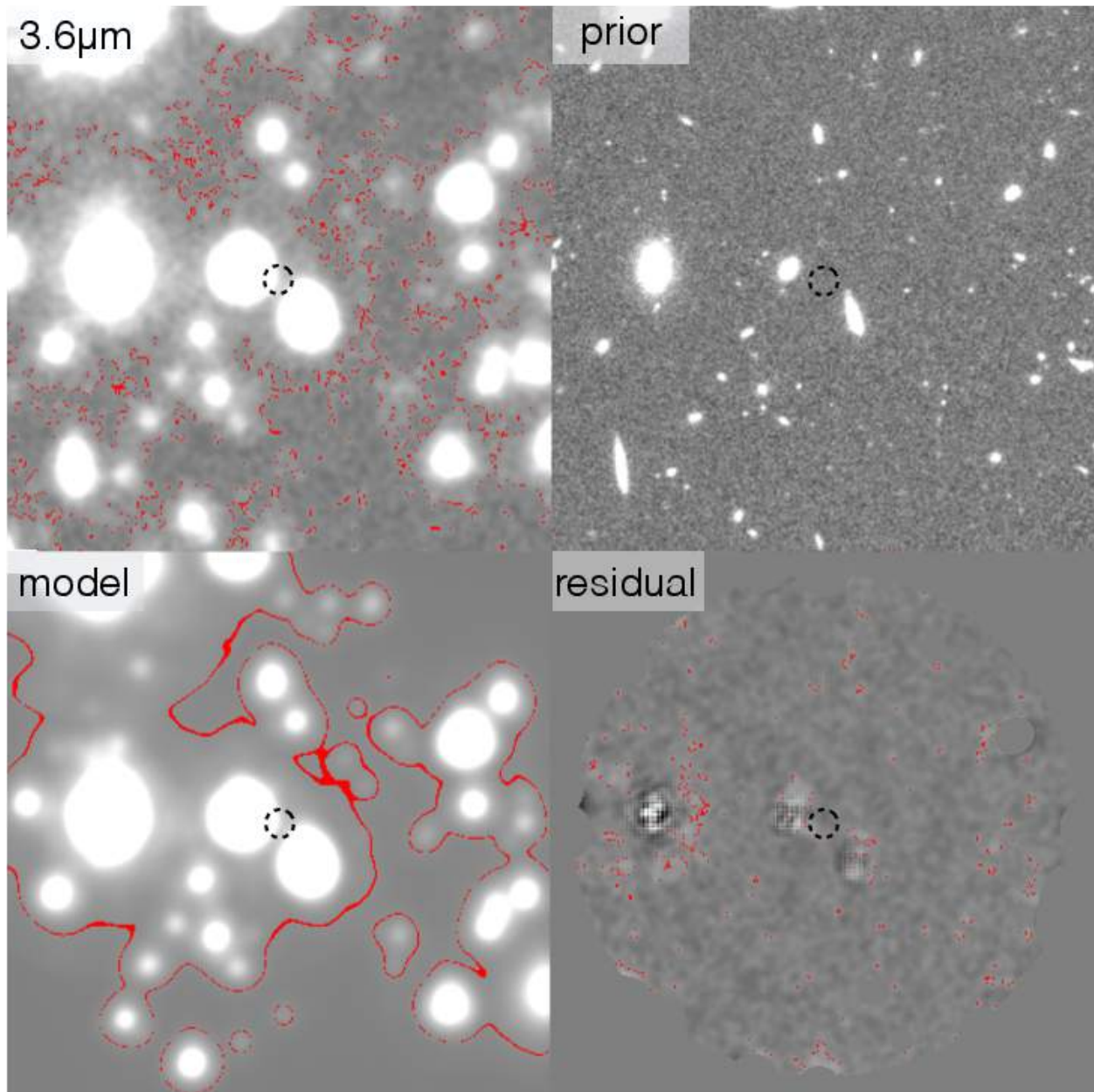


Figure 9. A demonstration how prior-based IRAC photometry can recover the full depth of the IRAC data. (*top left*) An ultradeep (146 hour exposure time) $40'' \times 40''$ section of the IRAC $3.6\mu\text{m}$ mosaic. The red contours shows the 2.5σ isophote above the background, indicating that $\sim 70\%$ of the background is contaminated by the PSF wings of sources. The black dashed aperture shows the location where a flux measurement is desired. (*top Right*) Deep HST/WFC3 imaging of the same location on the sky, which accurately determines the positions and sizes of the sources. (*bottom left*) A model is constructed by first convolving each WFC3 detected source by a kernel to approximate the IRAC PSF, and then fitting the flux for each individual source simultaneously. A high quality IRAC PSF model is needed to account for the PSF wings. (*bottom right*) The residual image shows that the sources are modeled and subtracted very well and that source confusion is greatly reduced. Small residuals remain around bright sources due to intrinsic color gradients and small imperfections in the PSF. The flux measurement in the central aperture in the residual image is within 1σ of the background.

background RMS is optimistic and only representative of the uncertainty in absence of confusion. As shown in Figure 10) (*right*) the two estimates agree very well for 90% of the sources: the histogram of the ratio of aperture flux to local background error resembles a standard normal $\mathcal{N}(0, 1)$ distribution. There is a slight skew towards positive flux levels, indicated by excess positive residuals for $\sim 5\%$ of the sources in the $2 - 3\sigma$ range. About 12% of the fluxes deviate by more than 5σ (10% high, 2%

low), nearly all due to strong residuals near the centers of very bright IRAC sources. About 3% deviate because of confusion in the high resolution WFC3 prior image.

We further investigate the relationship between contamination fraction and integration time, defining “strongly contaminated” as $> 5\sigma$ deviations from the local empty background RMS. Using simple aperture photometry (e.g., SExtractor) on the full-depth mosaics we find high contamination fractions: $\sim 80\%$ at $3.6\mu\text{m}$ and

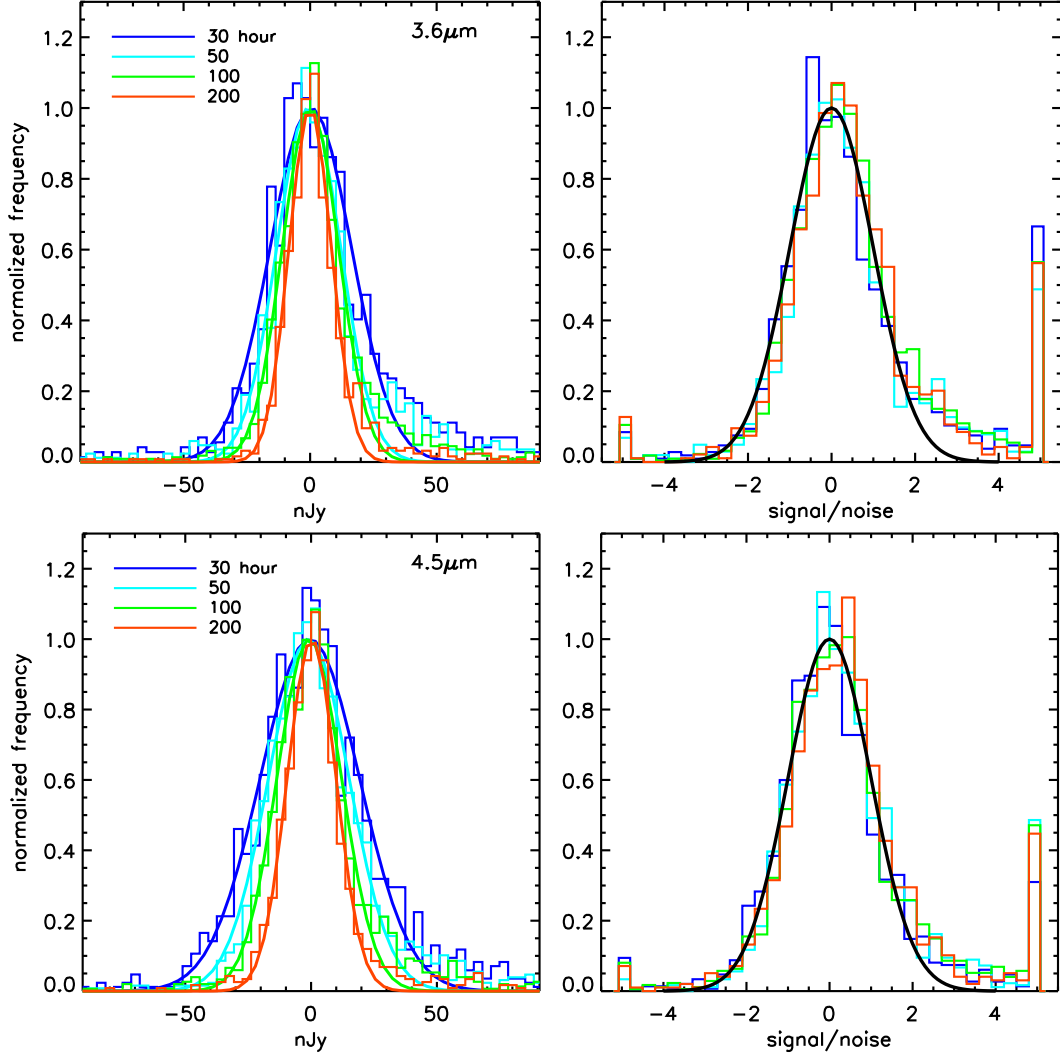


Figure 10. (*left*) Histograms of measured fluxes of artificial sources of zero flux, placed on 15,000 random locations in the full-depth mosaic, and grouped by integration time. The fluxes were measured in circular apertures of $D = 2''.0$ after modeling and subtracting neighboring sources following the procedure in Fig. 9. The solid lines show gaussian fits to the histograms. (*right*) The histogram of extracted fluxes divided by the local background rms in $1''.8 \times 1''.8$ binned pixels. The black curves show a standard normal $\mathcal{N}(0, 1)$ which would be expected in the absence of confusion, indicating that any residual confusion is not severe for most of the sources, even at 200 hours depth. There is a slight skewness towards positive flux levels, indicated by excess positive residuals for $\sim 5\%$ of the sources. About 12% of the fluxes deviate by more than 5σ .

$\sim 70\%$ in $4.5\mu\text{m}$. There is only a weak trend of contamination with integration time, likely because most flux comes from moderately bright sources and the PSF surface brightness profile is steep at small radii $R < 10''$ (e.g., Spitzer Observer Manual, SOM, section 6.2.4.1.5). For the cleaned photometry there is no trend with integration time over 20 – 200 hour (and a constant $\sim 12\%$ contamination). Hence prior based cleaning reduces the contamination fraction for these data sets by a constant factor $6\times$.

Figure 11 shows the relation between sensitivity and integration time based on the simulated sources. The noise decreases with a power-law slope of $t_{exp}^{-0.45 \pm 0.01}$ in both IRAC bands. The decrease is only slightly slower (at 2.5σ significance in each filter) than the $\sqrt{t_{exp}}$ expected for poisson noise. Following the definition of the IRAC integration time calculator (SENS-PET), we

convert aperture scatter to point source sensitivity by square root scaling the noise to an equivalent area of 10.5 arcsec^2 . This area represents the number of “noise pixels” (see SOM Table 6.1), which would effectively contribute to the uncertainty of linear least-squares fit of a point source. This amounts to optimal weighting by the PSF and improves the SNR by $\sim 30\%$ compared to unweighted apertures.

The best fit in magnitudes is:

$$\text{mag}(3.6\mu\text{m}, 1\sigma, AB) = 25.81 + 1.132 \log_{10} t_{exp} \quad (1)$$

$$\text{mag}(4.5\mu\text{m}, 1\sigma, AB) = 25.66 + 1.141 \log_{10} t_{exp} \quad (2)$$

or equivalently in flux densities:

$$\sigma(3.6\mu\text{m}, nJy) = 172 * t_{exp}^{-0.453} \quad (3)$$

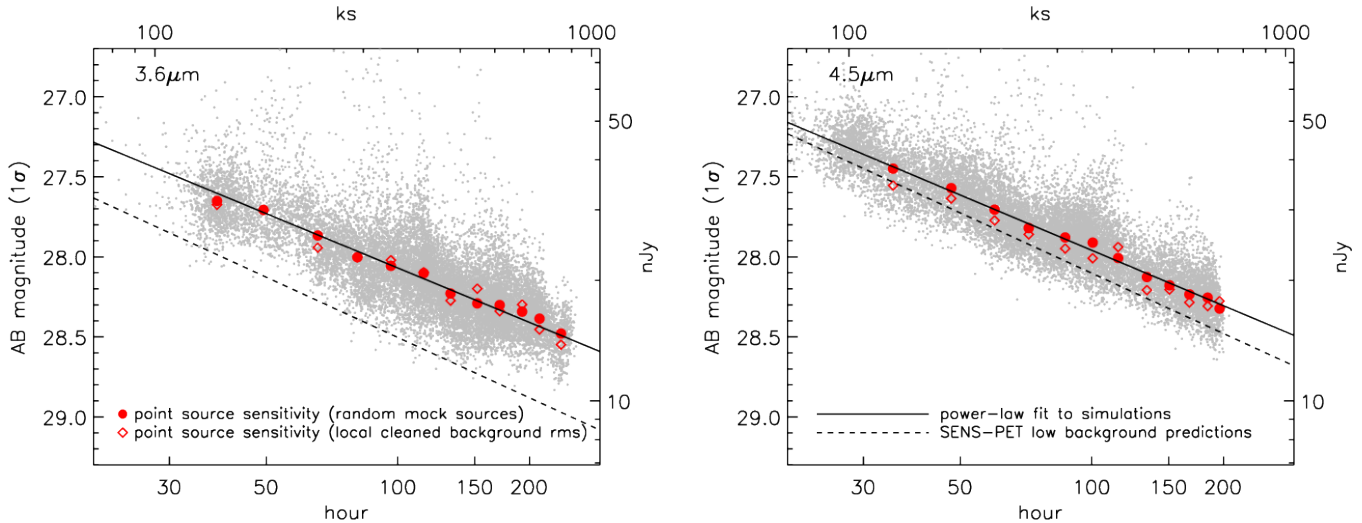


Figure 11. The relation between median point source sensitivity as a function of integration time based on simulations. Gray points show the point source fluxes extracted at a large number random locations, after fitting and subtracting neighboring sources using the WFC3 images as a prior. Red solid points show their medians in bins of exposure time. Open diamonds show the local background rms away from bright sources. The solid line is a power-law fit to the red solid points, with a best fit slope of $t_{exp}^{-0.45 \pm 0.01}$ in both IRAC bands. The decrease in noise with exposure time is only slightly slower (at 2σ significance in each filter) than the $1/\sqrt{t_{exp}}$ expected for Poisson noise, without evidence for a confusion limit or noise floor. The dashed line show predictions from the SENS-PET exposure time calculator.

$$\sigma(4.5\mu m, nJy) = 197 * t_{exp}^{-0.456} \quad (4)$$

which gives the median point source sensitivity as function of integration time in hours. No evidence is found for a confusion limit or noise floor, although the relation is consistently 10 – 30% less deep than predicted by SENS-PET for low background conditions. A possible explanation for the lower sensitivity is residual confusion by, e.g., sources below our detection limit or a background of faint overlapping PSF wings at larger radii than our PSF model. Note that the true uncertainty for individual sources can be much higher than the median if the source is located close to a bright neighbor.

4.4. Public Data Release

The data release consists of reduced images of all ultra-deep IRAC observations in the GOODS-South. The images are available from the IUDF website¹⁴ and the Infrared Science Archive¹⁵ (IRSA).

The data release contains the following:

- Science images and exposure time maps in both $3.6\mu m$ and $4.5\mu m$. Our reduction uses the same tangent point as CANDELS on pixel scales of $0''.3$, so the IRAC maps can be easily rebinned and registered to HST/WFC3 data.
- Reduced images of all individual 353 AORs, drizzled onto the same grid, which may be useful to study the reliability or variability of sources.
- Template PSFs and spatial maps of the weights and position angles of each AOR, allowing the reconstruction of the PSF at arbitrary locations. Example IDL code is provided.

The units of the science images are $cMJy/sr$, where constant $c=16.54$ represents the change from the native IRAC pixel scale to $0.3''/pixel$ due to flux conservation during the reduction process. Equivalently, flux densities can be obtained by multiplying the image pixel values by $34.994 \mu Jy/pixel$, corresponding to an image AB zeropoint of 20.04.

5. EXAMPLES

One of the main goals of the IUDF program is to obtain high SNR ($> 5\sigma$) at 3.6 and $4.5\mu m$ for normal $\lesssim L^*$ galaxies in the epoch of reionization. Comparing the detection rates of $H < 27.5$ galaxies at $z > 6$ to previous deep IRAC observations from the GOODS program (PID 194), we find that ~ 46 hour GOODS data yields $SNR > 5\sigma$ measurements for 25-30% of the sources, compared to 75-80% for 150 – 200 hour in the IUDF images.

Here we provide several examples of objects detected in the IUDF images. In Figure 12 we show 4 ultrafaint sub- L^* galaxies at $z \sim 7 - 8$. The galaxies are clearly detected at high significance in the new images, compared to the earlier 50 hour deep images. In the deeper images a clear difference in observed IRAC color is seen between the $z \sim 7$ and $z \sim 8$ galaxies, likely due to strong $[O III]+H\beta$ line emission moving from $3.6\mu m$ to $4.5\mu m$ with increasing redshift. These differences were recently demonstrated in stacked SEDs (e.g., Labbé et al. 2013) and in small samples of brighter and lensed galaxies (Smit et al. 2014, Smit et al. 2015), but are now apparent even in individual sub- L^* galaxies. This shows the potential of $\sim 150 - 200$ hour data for placing improved constraints on the emission line strengths of individual galaxies ($H\alpha+[N II]$ at $z = 4-5$ and $[O III]+H\beta$ at $z = 7-8$).

Furthermore, ultra-deep IRAC data may be the only way to detect potentially important overlooked constituents of the high redshift universe until the arrival

¹⁴ <http://www.strw.leidenuniv.nl/iudf/>

¹⁵ <http://irsa.ipac.caltech.edu/data/SPITZER/IUDF/> <http://irsa.ipac.caltech.edu/data/SPITZER/IUDF/>

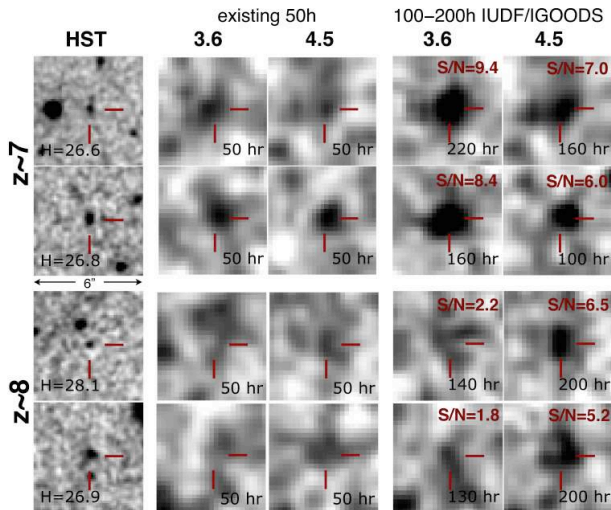


Figure 12. Inverted grayscale image stamps of two $z \sim 7$ and two $z \sim 8$ galaxy candidates in GOODS-South, after modeling and subtracting flux of neighboring sources based on the high-resolution HST image. The panels compare the 50 hour IRAC existing data to the full 100 – 200 hour dataset including IUDF + IGOODS (right columns). The stamps are $6 \times 6''$. Existing 50 hour data refer to a combination of GOODS-S (PID 194) and SCANDELS (PID 80217) data. The observed IRAC color changes between $z \sim 7$ and $z \sim 8$ galaxies (bright at $3.6\mu\text{m}$ vs bright at $4.5\mu\text{m}$) as strong [O III]+ $H\beta$ line emission moves from $3.6\mu\text{m}$ to $4.5\mu\text{m}$ (e.g., Labbé et al. 2013). Sources as faint as ($H_{AB} \sim 28$ mag) are detected.

of JWST. Massive $M \gtrsim 10^{10} M_{\odot}$ passive galaxies at $z > 4$ can be too faint to be detected by Hubble and even actively star forming, dusty galaxies with SFR $50 - 100 M_{\odot}/\text{yr}$ could have escaped detection by both Hubble and existing FIR/sub-mm surveys at these redshifts. Enigmatic IRAC-selected “HST-dropouts” have been identified on the basis of their very red $H - 4.5$ colors (e.g., Huang et al. 2011, Caputi et al. 2013). The origin of these objects is unknown as it is difficult to determine their redshifts, but the observed SEDs of some galaxies can be fit with quiescent galaxy models at high redshift $z > 4$. If this interpretation is correct, then these objects are the quenched remnants of massive starbursts at earlier times, and they provide compelling targets for early JWST spectroscopic follow up. Such a population likely places powerful constraints on models for star formation quenching, and may inform us indirectly about high mass star formation during the epoch of reionization.

6. SUMMARY

The IUDF and IGOODS programs are the deepest and most recent probes of the infrared emission at $3.6\mu\text{m}$ and $4.5\mu\text{m}$ with Spitzer/IRAC, ideally suited for faint studies of high redshift galaxies. Combining with all ultradeep archival data from all previous programs, and using consistent reduction procedures, we present reduced image mosaics reaching extremely deep coverage of 50 – 200 hours and covering all of GOODS-S, the HUDF/XDF, and the two HUDF parallel fields.

In summary:

- We release the full-depth reduced science mosaics

at $3.6\mu\text{m}$ and $4.5\mu\text{m}$ and the corresponding exposure time maps. The IRAC mosaics are placed on the same astrometric system and reference grid as the CANDELS WFC3 mosaics.

- The combined mosaics are the deepest ever taken at $3.6\mu\text{m}$ and $4.5\mu\text{m}$ with the integration times ranging from > 50 hour over 150 arcmin^2 , > 100 hour over 60 sq arcmin^2 , to $\sim 180 - 200$ hour over $5 - 10 \text{ arcmin}^2$. The image quality is $\text{FWHM} = 1''.49$ in both bands with $< 1.5\%$ spatial variation.
- The release also includes the separate reduced mosaics of all individual 353 AORs of the 7 programs involved in this release, registered and drizzled onto the same grid, to study the reliability or variability of sources.
- We present a new procedure to construct IRAC PSF maps from the data, well suited to deep fields with relatively few bright stars and complicated survey geometry with repeat observations under varying roll angles. The PSF maps are included in the release to facilitate PSF-fitting or joint IRAC+WFC3 photometry.
- Simulations are performed to quantify the confusion due to crowding by neighboring sources. We demonstrated using the new ultradeep 200 hour data that IRAC observations are not significantly impacted by confusion when using deep high resolution priors from HST/WFC3. In the reduced mosaics 70 – 80% of the area is originally contaminated by flux of neighboring sources. Using HST-based priors reduces this to a constant $\sim 12\%$, with no dependence on exposure time over the range 20 – 200 hours. The remaining catastrophic outliers are nearly all very close to the centers of bright IRAC sources and in 3 – 4% are even confused in the high resolution HST image. In general, prior based photometry works very well, reducing the contamination fraction by $6\times$.
- The simulations further demonstrate that the rms noise in the ultradeep IRAC images decreases nearly as the square root of integration time over the range 20 – 200 hours, without any evidence for a hard confusion limit. The maximum 1σ point source sensitivities reaches as faint as of 15 nJy (28.5 AB) at $3.6\mu\text{m}$ and 19 nJy (28.2 AB) at $4.5\mu\text{m}$. These sensitivities are systematically 10% – 30% less deep than predicted by the IRAC ETC (SENS-PET), likely due to residual effects of confusion. We provide fitting formulas in §4.3 to estimate the effective depth as a function of exposure time.

The value of ultradeep IRAC data is illustrated by direct detections of sub-L* $z > 7$ galaxies, where the joint

measurement at $3.6\mu\text{m}$ and $4.5\mu\text{m}$ places constraints on the $[\text{O III}]+H\beta$ emission line strengths of individual galaxies to very faint limits $H_{AB} \sim 28$. Future observations of larger samples over wider areas will become available as part of Exploration Science program GREATS (GOODS Reionization Era wideArea Treasury from Spitzer, PI Labbé), which will map part of GOODS-S and GOODS-N to 200 hours depth. These data offer the prospect of studying the distribution of inferred EWs and comparisons to the entire rest-frame SEDs, from HST to ALMA, will enable studies of the dust attenuation, ionization processes, and star formation histories. The combined HST+Spitzer ultradeep imaging legacy will be useful for planning efficient imaging and spectroscopic follow-up surveys with JWST and provide interesting targets for the first cycles of JWST NIRSPEC observations. Spitzer's heritage will extend well into the JWST era.

We thank the anonymous referee for careful comments which improved the clarity of the paper. We thank Justin Howell for IPAC assistance with the data release. Support for this work was provided ERC grant HIGHZ no. 227749 and NL-NWO Spinoza. We acknowledge support from NASA grants NAG5-7697 and HST-GO-11563. This research has made use of the NASA/IPAC Infrared Science Archive, which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

REFERENCES

- Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2013, *ApJ*, 769, 80
- Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2015, *ApJS*, 218, 33
- Atek, H., Richard, J., Kneib, J.-P., et al. 2014, *ApJ*, 786, 60
- Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, *AJ*, 132, 1729
- Bouwens, R. J., Illingworth, G. D., Franx, M., & Ford, H. 2008, *ApJ*, 686, 230
- . 2011, *ApJ*, 737, 90
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2014, [arXiv:1403.4295](https://arxiv.org/abs/1403.4295)
- Bradač, M., Ryan, R., Casertano, S., et al. 2014, *ApJ*, 785, 108
- Bruzual, G. & Charlot, S. 2003, *MNRAS*, 344, 1000 (BC03)
- Caputi, K. I., Dunlop, J. S., McLure, R. J., et al. 2012, *ApJ*, 750, LL20
- Damen, M., Labbé, I., van Dokkum, P. G., et al. 2011, *ApJ*, 727, 1
- Duncan, K., Conselice, C. J., Mortlock, A., et al. 2014, *MNRAS*, 444, 2960
- Ellis, R. S., McLure, R. J., Dunlop, J. S., et al. 2013, *ApJ*, 763, LL7
- Fazio, G. G., et al. 2004, *ApJS*, 154, 10
- Fernández-Soto, A., Lanzetta, K. M., & Yahil, A. 1999, *ApJ*, 513, 34
- Finkelstein, S. L., Papovich, C., Salmon, B., et al. 2012, *ApJ*, 756, 164
- Franceschini, A., Toffolatti, L., Mazzei, P., Danese, L., & de Zotti, G. 1991, *A&AS*, 89, 285
- Fruchter, A. S., & Hook, R. N. 2002, *PASP*, 114, 144
- Gialvalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, *ApJ*, 600, L93
- González, V., Labbé, I., Bouwens, R. J., Illingworth, G., Franx, M., Kriek, M., & Brammer, G. B. 2010, *ApJ*, 713, 115
- González, V., Bouwens, R., Illingworth, G., et al. 2014, *ApJ*, 781, 34
- Grazian, A., Castellano, M., Fontana, A., et al. 2012, *A&A*, 547, AA51
- Grazian, A., Fontana, A., Santini, P., et al. 2014, [arXiv:1412.0532](https://arxiv.org/abs/1412.0532)
- Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, 197, 35
- Huang, J.-S., Zheng, X. Z., Rigopoulou, D., et al. 2011, *ApJ*, 742, LL13
- Illingworth, G. D., Magee, D., Oesch, P. A., et al. 2013, *ApJS*, 209, 6
- Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, *ApJS*, 197, 36
- Labbé, I., et al. 2005, *ApJ*, 624, L81
- Labbé, I., Bouwens, R., Illingworth, G. D., & Franx, M. 2006, *ApJ*, 649, L67
- Labbé, I., et al. 2010, *ApJ*, 708, L26
- Labbé, I., et al. 2010, *ApJ*, 716, L103
- Labbé, I., Oesch, P. A., Bouwens, R. J., et al. 2013, *ApJ*, 777, LL19
- Laidler, V. G., Papovich, C., Grogin, N. A., et al. 2007, *PASP*, 119, 1325
- Laporte, N., Streblyanska, A., Clement, B., et al. 2014, *A&A*, 562, LL8
- McLure, R. J., Dunlop, J. S., Bowler, R. A. A., et al. 2013, *MNRAS*, 432, 2696
- Oke, J. B., & Gunn, J. E. 1983, *ApJ*, 266, 713
- Oesch, P. A., et al. 2009a, *ApJ*, 690, 1350
- Oesch, P. A., et al. 2010, *ApJ*, 709, L16
- Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2012, *ApJ*, 759, 135
- Oesch, P. A., Labbé, I., Bouwens, R. J., et al. 2013, *ApJ*, 772, 136
- Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2014, *ApJ*, 786, 108
- Reach, W. T., Megeath, S. T., Cohen, M., et al. 2005, *PASP*, 117, 978
- Rix, H.-W., Barden, M., Beckwith, S. V. W., et al. 2004, *ApJS*, 152, 163
- Salmon, B., Papovich, C., Finkelstein, S. L., et al. 2014, [arXiv:1407.6012](https://arxiv.org/abs/1407.6012)
- Salpeter, E. E. 1955, *ApJ*, 121, 161
- Schaerer, D., & de Barros, S. 2010, *A&A*, 515, A73
- Shim, H., Chary, R.-R., Dickinson, M., et al. 2011, *ApJ*, 738, 69
- Schmidt, K. B., Treu, T., Trenti, M., et al. 2014, *ApJ*, 786, 57
- Shapley, A. E., Steidel, C. C., Erb, D. K., et al. 2005, *ApJ*, 626, 698
- Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, *ApJS*, 214, 24
- Smit, R., Bouwens, R. J., Labbé, I., et al. 2014, *ApJ*, 784, 58
- Smit, R., Bouwens, R. J., Franx, M., et al. 2015, *ApJ*, 801, 122
- Stark, D. P., Ellis, R. S., Bunker, A., Bundy, K., Targett, T., Benson, A., & Lacy, M. 2009, *ApJ*, 697, 1493
- Stark, D. P., Schenker, M. A., Ellis, R., et al. 2013, *ApJ*, 763, 129
- Steinhardt, C. L., Speagle, J. S., Capak, P., et al. 2014, *ApJ*, 791, LL25
- Zheng, W., Postman, M., Zitrin, A., et al. 2012, *Nature*, 489, 406
- Zheng, W., Shu, X., Moustakas, J., et al. 2014, *ApJ*, 795, 93