TITLE

MARAT MUSIN 1 2, HAOJING YAN 2, Draft version September 12, 2018

ABSTRACT

Subject headings: infrared: galaxies — submillimeter: galaxies — galaxies: starburst — methods: data analysis

1. INTRODUCTION

1.1. Lilly-Madau formalism

1.2. SED fitting as a standard technique of mass and redshift estimation

SED fitting is now a standard technique of deriving stellar mass and photometric redshifts for a large set of galaxies. In this method multi-band photometry for a given galaxy is fitted to a series of a templates predicted by a certain stellar population synthesis (SPS) model. The best-fit template gives the parameters of the galaxy, including its redshift and mass. Historically, SPS models were using restframe optical photometry. One caveat is the degeneracy between the dust extinction and age of the stellar population, as both make the color of galaxy red, i.e. galaxy can be red because it is intrinsically red with no young massive star and ongoing star-formation, or it can be very dusty, or it can be metal-rich and metals effectively absorb light in the bluer bands. Solution to this is to implement restframe near-IR where light suffers much less extinction (comparing to restframe UV and optical) and thus the degeneracy can be broken. We aim to build the largest sample of galaxies with optical and near-IR photometry over a large sky area. The natural choice for us then is to use optical Sloan Digital Sky Survey (SDSS) and IR all-sky data from Wide-Field Infrared Survey Explorer (WISE).

1.3. problems associated with construction of the catalog

Blending, poor spatial resolution in IR our method template fitting

1.4. Goal of this paper

In this paper we present our technique for construction a catalog of galaxies with reliable SED data in optical and near-IR in Stripe 82 field. We discuss data selection, sources identification in different bands and problems associated with it.

2. DATA DESCRIPTION

2.1. SDSS and Stripe 82

The imaging component of the SDSS, which was done in five broad bands (u'g'r'i'z'), has covered 14,555 deg². In most area, the SDSS only scanned for one pass at

¹ Chinese Academy of Sciences South America Center for Astronomy (CASSACA), National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China
² Department of Physics & Astronomy, University of Missouri,

Columbia, MO 65211, USA

an exposure time of 53.9 seconds per band, and thus is rather shallow (for example, the r'-band 5 σ limiting magnitude is 22.2 mag). For this reason, in most cases the SDSS can only probe the normal galaxy population up to $z \approx 0.4$. However, the Stripe 82 region, which is a long stripe along the equator that spans $20^h < RA < 4^h \text{ and } -1.26^o < Dec < 1.26^o$, is the exception. It was repeatedly scanned (~ 70 -90 times, depending on RA) for calibration purpose during the survey (Adelman-McCarthy et al. 2007), and thus the combined scans can reach much better sensitivities.

A number of teams have created deep Stripe 82 stacks and made them available to public. The first such stacks were produced by Annis et al. (2014) based on the data obtained up to December 2005 (20-35 runs), which achieved 1-2 magnitude deeper limits than the singlepass SDSS images. Several other teams (e.g., Jiang et al. 2009; Huff et al. 2014) produced different stacks using different procedures to optimize the image qualities.

Jiang et al. (2014, hereafter J14) released a new version of stacks using only the images that were taken under the best weather conditions. These stacks are ~ 0.2 mag deeper than those produced by Annis et al. (2014), reaching 5 σ limits of 23.9, 25.1, 24.6, 24.1, 22.8 mag in u'q'r'i'z', respectively, and also have better PSF characteristics. We adopt these stacks in our work.

2.2. Structure of SDSS Stripe 82 files

We use description from J14 to present the structure of optical data. An SDSS run (strip) consists of six parallel scanlines, identified by camera columns (Figure??). The scanlines are 13.5 arcmin wide, with gaps of roughly the same width, so two interleaving strips make a stripe that consists of total 12 scanlines (columns). Their names in the direction of increasing declination are: $col01 \rightarrow$ $col07 \rightarrow col02 \rightarrow col08 \rightarrow col03 \rightarrow col09 \rightarrow col04 \rightarrow$ $col10 \rightarrow col05 \rightarrow col11 \rightarrow col06 \rightarrow col12.$

The size of each co-added SDSS image is 2854 x 2048 pixels, or roughly 18.8' x 13.5' (RA x Dec), with a pixel size of $0.396^{''}$ and an average full width at half maximum (FWHM) of $\sim 1.5^{"}$ in u-band, $\sim 1.3^{"}$ in g-band, and $\sim 1^{''}$ in r-, i-, and z-bands. In total there are 401 SDSS images in each column and overall $12 \cdot 401 \cdot 5 = 24,060$ SDSS images in all 5 bands. Each SDSS image has a corresponding weight.fits image, that records relative weights at individual pixels.

2.3. WISE and unWISE

WISE is a near-to-mid IR space telescope launched in 2009 and has performed an all-sky imaging survey in four bands at 3.4, 4.6, 12, and 22 μ m (denoted as W1, W2, W3, and W4, respectively). Complete all-sky coverage in two epochs was achieved after three phases of observa tion, namely WISE Cryogenic Survey (120% of the sky is covered), WISE 3-band Survey (W4 band excluded, 30% of the sky is covered) and NEOWISE Post-Cryo Survey (only bands W1 and W2 were used, 70% of the sky is covered). The nominal 5 σ limits in four bands are 0.08, 0.11, 1.0, and 6.0 mJy, respectively (see Wright et al. 2010, for details).

The AllWISE Data Release 1 was announced 2013 Nov 13, and includes images known as the "Atlas Images' Co-added matched-filtered Atlas Images from all three phases are intentionally convolved by the point-spread functions (PSFs) for better detection of isolated sources

However, this operation has the drawback that it reduces the image resolutions, which is not desirable in many applications. To deal with this problem, Lang (2014, hereafter L14) reprocessed all the WISE images independently without the PSF convolutions, and produced the stacks that preserve the original WISE spatial resolutions. These image products of L14, dubbed as the "unWISE" images, have the PSF full-width at half maximum (FWHM) values of 6" in W1, W2 and W3 and 12" in W4. We use these unWISE images for this work.

2.4. Structure of unWISE files - once again maybe I need to omit it in the paper

The unWISE coadds are on the same tile centers as the WISE tiles with 18,240 images per band, 1.56 x 1.56 degrees each. The tiles are named by their RA, Dec center: tile "0591p530" is at RA = 59.1, Dec = +53.0 degrees; i.e., the first four digits of the tile name is $int(RA \cdot 10)$, then "p" for +Dec and "m" for -Dec, then three digits of $int(abs(Dec)\cdot 10)$. For each tile and band w1-w4, several images are produced, we shall list only the ones that we make use of:

unwise-0000p000-w1-img-m.fits - "Masked" image, 2048 x 2048 pixels, TAN projected at 2.75"/pixel. Background-subtracted, in units of "Vega nanomaggies" per pixel: $mag = -2.5 \cdot (log_{10}(flux) - 9)$. This is the science image, the word "masked" means that some pixels have no unmasked pixels and no measurement at all: pixel value 0 and infinite uncertainty.

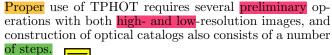
- unwise-0000p000-w1-std-m.fits - Sample standard deviation (scatter) of the individual-exposure pixels contributing to this coadd pixel.

Three unWISE images centered at the same RA cover the whole width of Stripe 82 in Dec (-1.26 δ ; +1.26). We shall call three such unWISE images a frame. There may be up to 72 SDSS images within one frame.

3. OVERVIEW OF METHODS FOR ANALYSIS

The most critical factor in the success of the SED fitting and estimation of the galaxies' properties is consistent photometry in available bands. This is challenging because the spatial resolutions of WISE is $\sim 6x$ worse than that of SDSS. For this reason, the objects detected in WISE often suffer blending. Even for relatively isolated WISE sources, the photometric apertures appropriate for the (low resolution) WISE images cannot guarantee the same fraction of light being included as what is done in the (high resolution) SDSS images. Such a systematic offset, which is different for every galaxy, severely skews the SED fitting.

To best address this problem, we opt to use the TPHOT software, which recently emerged as a robust and flexible tool to perform template fitting. The basic idea is to use a high-resolution image (here SDSS) as the prior to build the morphological template of the source under question, convolve this template with the PSF of the low-resolution image (here unWISE), and fit this degraded template to the low-resolution image to obtain the total flux that is within the aperture as defined by the high-resolution image. In this way, we get reliable color information (i.e., flux ratio) in the most consistent manner. It is important to note that high-resolution source does not have to be point-like - its morphological features will be preserved and fitted to the low-resolution source. This implies the biggest assumption for this technique - that morphology of the source is wavelength-independent. While generally it is not true, we anticipate that this will not create any significant bias. Firstly, because most of the galaxies have small angular sizes and such variation is negligible (galaxy at z=0.7 that is 40 kpc in diameter has an angular size of only 6 pixels), and secondly, because we chose SDSS r-band (6202.46 Å), as a high-resolution image there should not be much morphological difference between r-band and W1 or W2 bands.





3.1. matching optical catalogs

TPHOT input files must satisfy certain requirements: high- and low-resolution images should have the same type of projection, reference coordinate and orientation written in their FITS headers [Pence et al., 2010]. We verify that SDSS and WISE images indeed have the same tangential projection and orientation $(CD1_2)$ $0, CD2_{-1} = 0$). Meanwhile each WISE frame covers $\approx 3.95 \, deg^2$ and there may be up to 72 SDSS images that cover the same area, so we use coordinates of the center of WISE images (CRVAL1 and CRVAL2) as the anchor values and run SWarp [Bertin et al., 2002] to change reference pixels in all SDSS images. The reference pixel (CRPIX1 and CRPIX2) for SDSS images can now be outside of the image itself to as far as 0.7 deg.

There are SDSS files that lie in the overlapping region of two adjacent unWISE images. In such a case we duplicate SDSS image and run SWARP twice, assigning new SDSS image reference coordinates from each of the adjacent unWISE images. This operation reduced the blind zone but increased the number of SDSS files from 4812 to 5556 images per band.

* construction of the PSF and kernels

Optical part of the catalog consists of 5 SDSS bands, each of which has different sensitivity and FWHM, so each field has different number of sources in different bands. We shall choose the band that we use for detection, i.e., the source is added to our catalog if only it is detected in this band with sufficient SNR. Also position of the source in this band is used as the true position in all other bands. Another issue is the flux measurement -



















reliable SED fitting requires consistent multiwavelength photometry while different photometric conditions and FWHM could result in a non-consistent colors even in optical filters alone. Thus the measurement that are made even with flexible elliptical apertures known as the Kron apertures [Kron, 1980] that are denoted in SExtractor as FLUX AUTO [Bertin and Arnouts, 1996] will not be measuring the same fraction of the total flux from the given source in different filters. For that reason we convolve griz bands to the one with the worst FWHM, i.e., u-band. In such case the difference in sizes will be due to the intrinsic difference in morphology at different wavelength rather than due to inconsistent photometry. For the price of loosing some sources due to blending we shall achieve more robust optical colors. We use IRAF/psfmatch task that requires supplying the image with the corresponding kernel - PSF matching function. Construction of a kernel requires knowledge of the PSFs of the input image and a matched image. That means we need to construct PSF for each image in each band $(5, 556 \quad 5 = 27, 780 \text{ PSFs})$ and this is the most timeconsuming part of the project.

One of the widely-used ways to construct PSF is by using PSFEx software [Bertin, 2011]. We tested it with TPHOT and found that the quality of residual images is not satisfactory, so we decided to use more conservative algorithm in which IRAF/psf task creates a PSF function based on a set of selected point-like sources (so called PSF stars). Usually that strategy involves consecutive run of IRAF/find, IRAF/phot, IRAF/pstselect and IRAF/psf tasks that find stars, determine its magnitude, select the ones that do not have any morphological features and finally construct a PSF function. This algorithm, though very reliable, is also not ideal - in each image there are $\approx 10\%$ of sources that do not satisfy the criteria of PSF stars (are elongated, blended or have noisy background) and have to be rejected in the manual regime (selection in interactive mode). This is extremely time-consuming and inefficient given the total number of images in 5 bands. We decide to use a variation of this method to create PSF functions in a more automatic fashion.

We run SExtractor on each SDSS image and only mark sources from the output catalog as a point-like, if their stellarity index (determined by a CLASS STAR parameter) is close to 1, sources are bright, but not saturated, they do not lay within 2. PSF radius to the edge of the FITS image, and have all their flux contained within some small aperture. The later was calculated as the difference MAG diff between MAG APER that returns all flux within some fixed aperture and MAG BEST that returns MAG AUTO value in the absence of contamination from the nearby sources and MAG ISOCOR otherwise. Exact parameters depend on the band and also were adjusted during test runs so that each SDSS image has no less than 20 PSF stars (stars were sorted by their magnitude, from bright to dim), with the maximum number of PSF stars limited to 160 (larger values significantly slow down IRAF).

* matching all 5 bands to the PSF of the u-band

When all PSF stars are selected PSF was constructed using standard IRAF/psf. Then we run IRAF/seepsf, a task that takes the input PSF computed by the IRAF/psf task, consisting of the parameters of a 2D analytic func-

tion stored in the image header, and computes 21x21 pixel output PSF FITS image consisting of the sum of the analytic function and the residuals. Visual inspection of all PSF images revealed a certain fraction of defect PSFs. The reason for it can be either a blended source within the PSF radius, high background values due to the nearby bright star (within up to several arcmin), or noticeable elongation of the source (may be due to inclusion of a galaxy in our sample or due to intrinsic problems with SDSS image). The fraction of such images varied in different bands but generally was $\approx 4\%$ and for such images we re-run IRAF/psf in interactive mode (manually selecting PSF stars from a preselected catalog). All PSF images were normalized to unity total counts using wcstools/sumpix [Mink, 1998] and IRAF/imarith tasks. Several PSFs for all 5 bands and 6 different columns centered at RA=21:56:46 are shown in Figure 3.2. We used IRAF/lucy, a task that uses algorithm developed independently by Lucy [Lucy, 1974] and Richardson [Richardson, 1972], to create kernels for pairs of images in x- and u-bands, where x- stands for g-, r-, i-, or z-band

With kernels in hand it is now possible to run IRAF/psfmatch that convolves input image with the kernel to produce a psf matched output image (Figure 3.3). From when u-band, g-band, r-band etc images mean "SDSS Stripe 82 images stacked by J14 with new CR-VAL and new CRPIX values", while g matched u, etc. means images in the g-band matched to the PSF of the u-band image.

- * catalog construction with SExtractor in dual mode, using r_matched_u band for detection SDSS co-adds from J14 already have associated catalogs on the whole Stripe 82 region. We decide not to use it and create our own for several reasons:
- J14 catalogs are not matched between the bands, resulting in different number of objects in different bands.
 SNR at which the source is rejected is too high and there are a lot of sources that are identified in several bands when we perform our own photometry that were not included in J14 catalog.
- Bright and saturated objects are detected as multiple sources, none of which is at the center of such object making correct SED fitting impossible.

On Figure ?? we show a comparison of the original catalog from J14 (sources within green regions) and our catalog (sources within red regions). In the next 2 subsections we show the strategy for catalog construction with consistent optical colors in all 5 SDSS bands.

To construct a consistent optical catalog we run SExtractor in the dual mode - the first image is used for detection and astrometry information, while the second is used solely for photometry. We take r mathced u image as detection image and supply consequently all 5 bands to extract photometry of the sources and reject sources with SNR; 5 based on the r matched u flux. R matched u catalogs as well as corresponding segmentation maps also serve as an input to the TPHOT to derive consistent photometry on near-IR bands. Our parent catalog now consists of 26,585,000 sources. We notice that very few sources were rejected at this stage - objects that appear almost undetectable had FLUX AUTO / FLUX-ERR AUTO larger than 5 or even 7.

* correction of the error magnitudes in g ,r ,i and z

matched u bands

Photometry comparison between original x-band and x matched u-band catalogs shows a systematic underestimation of the errors associated with the source's flux (Figure 3.4). We believe that the reason for this was that we supplied to SExtractor science images matched to the u-band while the weight images were not matched to the u-band. This was the reason for high SNR ratios and low rejection rate for the faint sources. To correct for that we use STILTS code [Taylor, 2006] to match x-band and x matched u catalogs in 5 bands and for each separate image we calculate the mean ratio of the flux error before and after PSF matching. All magnitude errors in a given image are then multiplied by this coefficient. Figure 3.5 shows coefficients for all 5556 images for 4 bands matched to the u-band. We did not apply IRAF/psfmatch task to the u-band, but still calculated that correcting coefficient because r matched u FITS file was used as detection image in SExtractor. All coefficients for this band are less than 1. (CHECK THIS!)

We anticipate that while corrected values account for statistical error, there should also be a systematic error in magnitude that needs to be taken care of. We performed a series of tests where different constant errors were added in quadrature to the reported magnitude error and then the goodness of fit on the graph z spec vs z phot was checked. We found that correlation is the tightest when 0.04 magnitude error is added in quadrature (so final error cannot be smaller than 0.04). Each source in each band was assigned by a error in magnitude that was calculated using the following equation:

$$magerr_corrected = \sqrt{0.04^2 + (\frac{1.0857}{SNR} \cdot correction_coefficient)^2} \quad \text{4. REDSHIFT AND MASS ESTIMATION}$$
The availability of near-IR filters helps improving

- ** masking area around bright stars
- ** removing double objects
- ** star-galaxy separation
 - 3.2. preparatory work with input files to TPHOT
 - * SWARPing of unWISE files

We use SWarp to change the pixel scale of all unWISE images from original 2.75 to 2.772 to match the integer pixel scale ratio with SDSS images = 7. Now we construct unWISE PSF functions that is to be convolved with SDSS PSF to produce kernels - one of the most important set of the input files to TPHOT. There are 240 unWISE images within Stripe 82 footprint per band. Bands w3 and w4 are too shallow and no reasonable flux can be extracted for the vast majority of optical sources in these bands, so for the purpose of this project we only use w1 and w2 bands. We followed the strategy from the previous section (i.e., running SExtractor, selecting potential PSF stars using STILTS) to construct PSF for all 480 images with two major differences in the procedure: 1) Center pixels of saturated sources in unWISE standard deviation images (unwise- 0000p000-w1-std-m.fits) always have zero value and that is an invalid input for TPHOT. We use IRAF/imcalc to detect such pixels and change its value to 9999. 2) We construct each PSF function using IRAF/psf interactive mode. This was done to perform more robust selection of stars as 3 PSFs from one unWISE frame are convolved with up to 72 SDSS PSFs and thus its quality is crucial - incorrect PSF profile leads to wrong flux estimation associated with such sources and also characteristic positive and negative ringshaped patterns in the residuals.

After running the IRAF/seepsf task we get PSF FITS images, 19x19 pixels each. All unWISE PSFs are then normalized to unity total counts using IRAF/imarith and sub-sampled in size by the factor of 7 using IRAF/imlintran (to 133x133 pixels) to match the pixel scale ratio between unWISE and SDSS images. Result is presented on Figure 3.6 * kernel construction We use r matched u band as a detection band for the optical sources and also as a high-resolution image for the template fitting. We run the same algorithm to create 5556 r matched u PSFs (not to be mixed up with the r matched u PSF that we use to change FWHM of the original r-band). IRAF/lucy is used to convolve normalized r matched u PSF with normalized and sub-sampled un-WISE PSF to create kernels. These kernels serve as one of the inputs for TPHOT.

* naming convention

Both final and intermediate results are a combination of SDSS and unWISE data and we introduce our new naming convention that is used throughout the project. E.g. file kernel.w1.0000p015 12r u 111.fits is a kernel made by convolution of the PSF function of the file unwise-0000p015-w1-img-m.fits and PSF function of the file S82 12r 111.fits in r-band, that has been previously matched to the PSF of the u-band.

The availability of near-IR filters helps improving the z phot accuracy beyond z 1.3, where the 4000 break goes out of the z 0 filter and the Lyman break is not yet

detectable in the u band. (from 1809.03373.pdf)

4.1. TPHOT

parameters, limitations, computation on the cluster

4.2. Final catalog with photometric redshifts

matching by the source number comparison output fluxes from TPHOT to the unWISE catalogs

4.3. EAZY

redshift determination, matching to the spectroscopic redshifts comparison of the redshift with and without unWISE data (similar results) star-galaxy separation final cleaning of the catalog removing of bright stars and duplicated sources

5. RESULTS

6. DISCUSSION

7. SUMMARY

REFERENCES

J.K. Adelman-McCarthy, AGUEROS M.A., ALLAM S.S., ANDERSON K.S.J., ANDERSON S.F., ANNIS J., BAHCALL N.A., BAILER-JONES C.A.L., BALDRY I.K., BARENTINE J.C., BEERS T.C., BELOKUROV V., BERLIND A., BERNARDI M., BLANTON M.R., BOCHANSKI J.J., BOROSKI W.N., BRAMICH D.M., BREWINGTON H.J., BRINCHMANN J., BRINKMANN J., BRUNNER R.J., BUDAVARI T., CAREY L.N., CARLILES S., CARR M.A., CASTANDER F.J., CONNOLLY A.J., COOL R.J., CUNHA C.E., CSABAI I., DALCANTON J.J., DOI M., EISENSTEIN D.J., EVANS M.L., EVANS N.W., FAN X., FINKBEINER D.P., FRIEDMAN S.D., FRIEMAN J.A., FUKUGITA M., GILLESPIE B., GILMORE G., GLAZEBROOK K., GRAY J., GREBEL E.K., GUNN J.E., DE HAAS E., HALL P.B.,

J. Annis, M. Soares-Santos, M. A. Strauss, A. C. Becker, S. Dodelson, X. Fan, J. E. Gunn, J. Hao, Ž. Ivezić, S. Jester, L. Jiang, D. E. Johnston, J. M. Kubo, H. Lampeitl, H. Lin, R. H. Lupton, G. Miknaitis, H.-J. Seo, M. Simet, and B. Yanny. The sloan digital sky survey coadd: 275 deg² of deep sloan digital sky survey imaging on stripe 82. ApJ, 794:120, October 2014. doi: 10.1088/0004-637X/794/2/120. URL http://adsabs.harvard.edu/abs/2014ApJ...794..120A. E. M. Huff, C. M. Hirata, R. Mandelbaum, D. Schlegel, U. Seljak, and R. H. Lupton. Seeing in the dark - i. multi-epoch alchemy. MNRAS, 440:1296–1321, May 2014. doi: 10.1093/mnras/stu144. URL

http://adsabs.harvard.edu/abs/2014MNRAS.440.1296H.
L. Jiang, X. Fan, F. Bian, J. Annis, K. Chiu, S. Jester, H. Lin, R. H. Lupton, G. T. Richards, M. A. Strauss, V. Malanushenko, E. Malanushenko, and D. P. Schneider. A survey of z ~ 6 quasars in the sloan digital sky survey deep stripe. ii. AJ, 138: 305–311, July 2009. doi: 10.1088/0004-6256/138/1/305. URL

http://adsabs.harvard.edu/abs/2009AJ....138..305J.
L. Jiang, X. Fan, F. Bian, I. D. McGreer, M. A. Strauss, J. Annis, Z. Buck, R. Green, J. A. Hodge, A. D. Myers, A. Rafiee, and G. Richards. The sloan digital sky survey stripe 82 imaging data: Depth-optimized co-adds over 300 \deg^2 in five filters. ApJS, 213:12, July 2014. doi: 10.1088/0067-0049/213/1/12.

URL http://adsabs.harvard.edu/abs/2014ApJS..213..12J. D. Lang. unwise: Unblurred coadds of the wise imaging. AJ, 147:108, May 2014. doi: 10.1088/0004-6256/147/5/108. URL http://adsabs.harvard.edu/abs/2014AJ....147..108L.

E. L. Wright, P. R. M. Eisenhardt, A. K. Mainzer, M. E. Ressler, R. M. Cutri, T. Jarrett, J. D. Kirkpatrick, D. Padgett, R. S. McMillan, M. Skrutskie, S. A. Stanford, M. Cohen, R. G. Walker, J. C. Mather, D. Leisawitz, T. N. Gautier, III, I. McLean, D. Benford, C. J. Lonsdale, A. Blain, B. Mendez, W. R. Irace, V. Duval, F. Liu, D. Royer, I. Heinrichsen, J. Howard, M. Shannon, M. Kendall, A. L. Walsh, M. Larsen, J. G. Cardon, S. Schick, M. Schwalm, M. Abid, B. Fabinsky, L. Naes, and C.-W. Tsai. The wide-field infrared survey explorer (wise): Mission description and initial on-orbit performance. AJ, 140:1868-1881, December 2010. doi: 10.1088/0004-6256/140/6/1868. URL http://adsabs.harvard.edu/abs/2010AJ....140.1868W. http://adsabs.harvard.edu/abs/2010AJ....140.1868W.

APPENDIX

APPENDIX

A. PROSPECTIVE AUTOMATION OF THE DECOMPOSITION PROCESS