Next Generation VLA Memo 67: Demonstration & Analysis of ngVLA core + Short Baseline Array Extended Structure Imaging

Brian Mason (NRAO)

November 17, 2021 (v3b)

Abstract

I have quantified the imaging performance of the ngVLA core, the Short Baseline Array (SBA), and the two used jointly together for use cases involving imaging spatially extended science targets. For the simplest simulation— a circular disk of uniform surface brightness at 100 GHz— I find that the Rev.C core recovers 90% of the input disk flux up to a disk diameter of $17''.3$ ($=0.85\lambda/b_{min}$), and the SBA up to $32''.8$ ($= 0.58\lambda/b_{min}$). I take these to define the Largest Angular Scale (LAS) accessible. Thus, the SBA interferometer component can accurately recover an LAS $1.9\times$ larger than the ngVLA core alone. I also evaluated the LAS of the RevC.01 array and find it to be modestly reduced ($15''.7$ at 100 GHz) due to the adjustments in the minimum antenna spacing.

I simulated observations of a more complex and spatially extended ($\sim 5'$) astronomical source including SBA total power data. For the 93 GHz center frequency of this simulated observation the expected core and SBA LAS values are $15''$ and $35''$, respectively. When SBA, total power, and core data are combined, the total flux densities obtained are correct to within 1% to 3%. The image fidelities range from 90% to 95%, depending on the precise combination and imaging strategy used, meeting ngVLA project requirements. The fidelity measurements use the metric adopted by the project, which in turn is based on the critical analysis and rationale also presented here. The RevC.01 SBA configuration gives almost identical results.

To obtain the high fidelity results presented it was necessary to carefully examine the simulated single-dish results and apply significant (nearly a factor of two) corrections to the simulated image calibration. These procedures are described in detail, as are the procedures needed to correctly image (non-ALMA) heterogeneous array data. Finally, I evaluate the impact of the Jorsater-VanMoorsel effect and find it is not a significant factor in this analysis. The analysis is scripted and repeatable, and makes use of core and SBA specific automasking parameters which I present here.

Version History:
v1 Oct 9, 2019: first released version.
v2 Jan 8, 2021: run correct joint SBA+core imaging and add appendix explaining how to do so. Add aperture flux analysis, LAS analysis (disk simulations), and rev.C/revC.01 comparison. Reorganize results section for clarity.
v3 Jun 24, 2021: add clarifying footnote about time ratios
v3b Nov 2021: move uvmodelfit explenation to the section where it belongs

Contents

1 Introduction

The Next Generation VLA (ngVLA) reference design calls for 18-meter diameter antennas with offset Gregorian optics (Selina et al. 2018a, 2018b). One consequence of this design is that the shortest baselines available to the ngVLA main array will be 38 meters due to mechanical clearance requirements driven primarily by the offset optics. The anticipated science targets of the ngVLA include numerous sources with spatial structures larger than what can be reliably recovered using this range of baselines (ngVLA SAC 2017; Selina, Murphy & Erickson 2017). To provide the capability of imaging these larger science targets,

a Short Baseline Array (SBA) has been designed and is included in the reference design (Mason et al. 2018). As currently envisaged the SBA comprises an array of 19 6-m diameter antennas operating as an interferometer, and four 18-m total power antennas. The 6-m array provides baselines between 60 meters and 11 meters, with the shortest spacing again driven mainly by feedarm clearance requirements. This memo presents a demonstration and quantitative analysis of the imaging capabilities of the SBA together with the ngVLA core.

Note that most of the work in this memo was done focusing on the current reference design, and in particular, I use the Rev.C configuration files publicly available on the ngVLA web page. Carilli et al. (2020, ngVLA memo 82) present a modestly improved version of the current (Rev. C) reference configuration[1], which they denote as RevC.01. The changes between Rev.C and RevC.01 include adjusting the core antenna spacing to respect mechanical clearance requirements (30m to 38m); moving the SBA to reduce shadowing of SBA antennas by nearby 18m antennas; and rotating the SBA by 30° to reduce self-shadowing. All results reported in this memo are for Rev.C except as noted.

The scripts used to produce most of the results in this memo can be found at

`https://github.com/teuben/dc2019/tree/master/scripts/ngVlaSbaSims`

and the modified "30 Dor" input FITS image is at

`https://astrocloud.nrao.edu/s/ZPbz95tkTi4sSPr`

Please contact me (at NRAO email "bmason") to report any issues.

2 Largest Angular Scale Analysis

In order to more precisely quantify the largest angular scale (LAS) recovered by the ngVLA core and SBA in the best possible case, I carried out simulations of disk of uniform surface brightness for a range of disk diameters. The surface brightness of the disk was the same in all cases and there was no thermal noise added to the simulation; the simulations were conducted at a frequency of 100 GHz. For the SBA both a snapshot ($b_{min} = 11m$) and a 5 hour track ($b_{min} = 9.2m$) were simulated[2]. For the core only a snapshot was simulated, resulting in $b_{min} = 30.5m$ — see below for more on this. Images were made using a multi-scale deconvolution in `tclean`, with an *a priori* clean region consisting of a circle equal in diameter to the simulated disk plus two synthesized beams. The fraction of integrated flux density recovered as a function of disk diameter is shown in Figure 1.

Using the ngVLA 90% fidelity requirement as a benchmark we can identify the largest angular scale recovered as the point at which the recovered flux is less than 0.9× the input disk flux. For the core I find $LAS_{core} = 17''.3 = 0.85\lambda/b_{min}$. The recovered flux falls below the 50% level at $21''.2$, very close to λ/b_{min} for the core. For the SBA I find

[1]Since the changes are modest and the configurations are under active review, a project wide change request was not deemed worthwhile. Consequently the RevC.01 files have not yet been publicly released.

[2]Because the mechanically allowed shortest spacings are larger than a dish diameter, considerably shorter spacings than 11m are possible, even after shadowed antennas are flagged (no visibilties in the 5h SBA dataset were in fact shadowed).

$LAS_{SBA} = 32''.8 = 0.58\,\lambda/b_{min}$ for both the snapshot and the long track. The recovered flux falls marginally less quickly for the long track (half flux at $45''.8$) compared to the snapshot (half flux at $43''.8$). Extrapolating these findings to other frequencies gives

$$LAS_{core} = 17''.3\frac{100\,\mathrm{GHz}}{\nu} \tag{1}$$

and

$$LAS_{SBA} = 32''.8\frac{100\,\mathrm{GHz}}{\nu}. \tag{2}$$

Additionally, the RevC.01 core configuration of Carilli et al. was analyzed. I find

$$LAS_{core,C.01} = 15''.7\frac{100\,\mathrm{GHz}}{\nu} \tag{3}$$

due to the revision in the minimum spacing from $30.5m$ to $38.2m$.

The performance as a function of angular scale seen here is considerably better than is found using aperture photometry of a more complex target (§ 6.4), presumably because the emission is simple, bright, and lacks faint diffuse structure which is very difficult for interferometers to recover.

I also attempted an independent measurement of LAS using `uvmodelfit` but encountered various problems. The fundamental limitation in the context of these simulations is that the current CASA UV-model fitting task does not support mosaic (multiple field ID) data.

3 Complex Extended Source Simulations

The ALMA image simulation library[3] image of 30 Doradus— an HII region in the Large Magellenic Cloud, also known as the Tarantual Nebula— was chosen as a target template for these simulations because the object shows structure over a wide range of spatial scales. For purposes of these simulations the target's Declination was changed to $+22:38$ in order to be observable from the Northern hemisphere. The image was also re-gridded onto a $10\times$ higher spatial resolution grid in order that the cell size adequately sample the ngVLA core beam at a $2''$ taper; the number of pixels in each orthogonal axis was also increased by a factor of 10 so that the overall image size was unchanged. The image is originally a SPITZER 8μm continuum image. Early ALMA observations of a small region of 30 Doradus are presented in Indebetouw et al. (2013).

The simulations were done with custom `simobserve` scripts configured to observe a $4'.5 \times 4'.5$ region at 93 GHz, resulting in a 187-pointing mosaic with the ngVLA core and 23 pointings with the SBA. The simulated single dish maps covered a larger region ($6'.75 \times 6'.75$) in order to provide a "buffer region" around the interferometer image; regions much smaller than this tend to introduce artifacts in the `feather` step. The ngVLA core simulation comprised 187 10-second integrations, for a total integration time of 1870sec,

[3] `https://casaguides.nrao.edu/index.php?title=Sim_Inputs`

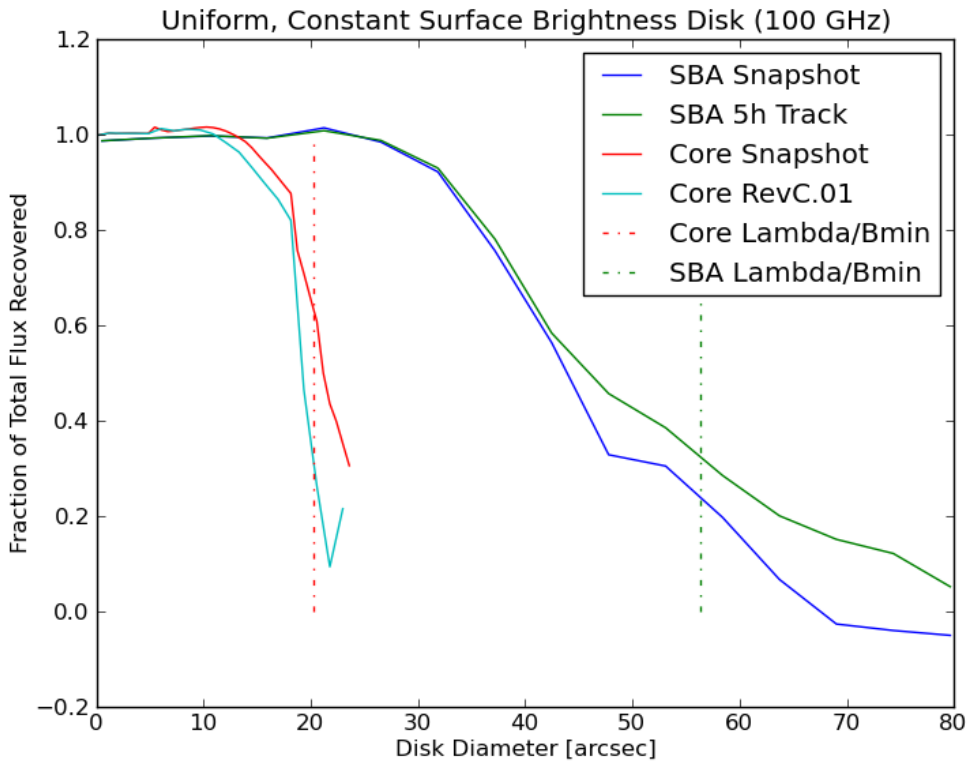

Figure 1: Fraction of flux recovered in a simulated observation of a uniform disk with the ngVLA core (snapshot; RevC and RevC.01) and the SBA interferometer (snapshot and 5h track).

and the SBA simulation comprised 23 180-second pointings (with 18× $10sec$ integrations for each pointing) for a total integration time of 4140sec. This integration time ratio[4] $t_{sba}/t_{core} = 2.2$ was used to match sensitivities in the overlapping range of baselines for the configurations, `ngVla-core-revC` and `ngVla-sba-revC`. None of the simulations included thermal noise or other systematic corruptions (*e.g.* tropospheric phase noise or antenna pointing errors). The single dish simulation total integration time was chosen so as to completely cover the required spatial area, but since thermal noise was not present in the simulations, the actual integration time was not consequential. In contrast the integration time (and integration durations) used in the interferometric simulations influences the *uv*-coverage of the data, which is an important variable in determining the characteristics of the resulting images. All simulations were run with zero sky opacity.

The antenna pointing patterns for the three arrays are shown in Figure 2. All simulations used the appropriate primary beams for each antenna: an Airy pattern for 6m or 18m diameter. The mosaic primary beams and the single dish weight map which results is also shown in Figure 2. For the wavelength and array configurations used in these simulations the largest spatial scales sampled by the ngVLA 18-m core array is $\lambda/b_{min} = 21''.8$ and for the SBA interferometer $60''.4$, although as shown in § 2 the actual largest scales accurately recovered at 93 GHz are at most $\sim 19''$ and $\sim 35''$. The natural beam sizes were $1''.9$ and $11''.3$ for the core and SBA, respectively.

The SBA (only) simulations described above were also repeated with the RevC.01 (rotated) configuration of memo 82.

4 Image Quality Metrics

It is crucially important for a telescope to be able to form accurate images of the sky brightness, with minimal (or known and well-understood) deviations. The quantitative measure of this capability is referred to as *image fidelity*. Because there is no general consensus on the best way to quantify image fidelity, I examined a variety of metrics that have been used. Defining the true sky brightness M (the input model to the simulation) and the synthesis image formed as I, one "classical" definition of image fidelity is:

$$F_1 = 1 - \frac{Max(|I - M|)}{Max(M)} \tag{4}$$

i.e., the image fidelity is taken to be the maximum deviation between the true sky and our (telescope-data based) model of it, divided by the peak true sky brightness. The evaluation is implicitly carried out over some predefined, potentially use-case dependent region of interest. This metric has a maximum value of 1.0, corresponding to perfect fidelity, but is unbounded from below. It has a simple interpretation— it is the maximum error

[4]The appropriate time ratio to use for the Rev.C configurations in question is 1.03; the value used here was erroneously carried forward from a previous simulation. Since thermal noise is not present in these simulations the impact is expected to be minimal; this will be verified in a future simulation. Note that the appropriate time ratio to match the Rev.D 4km core with the SBA is 0.59.

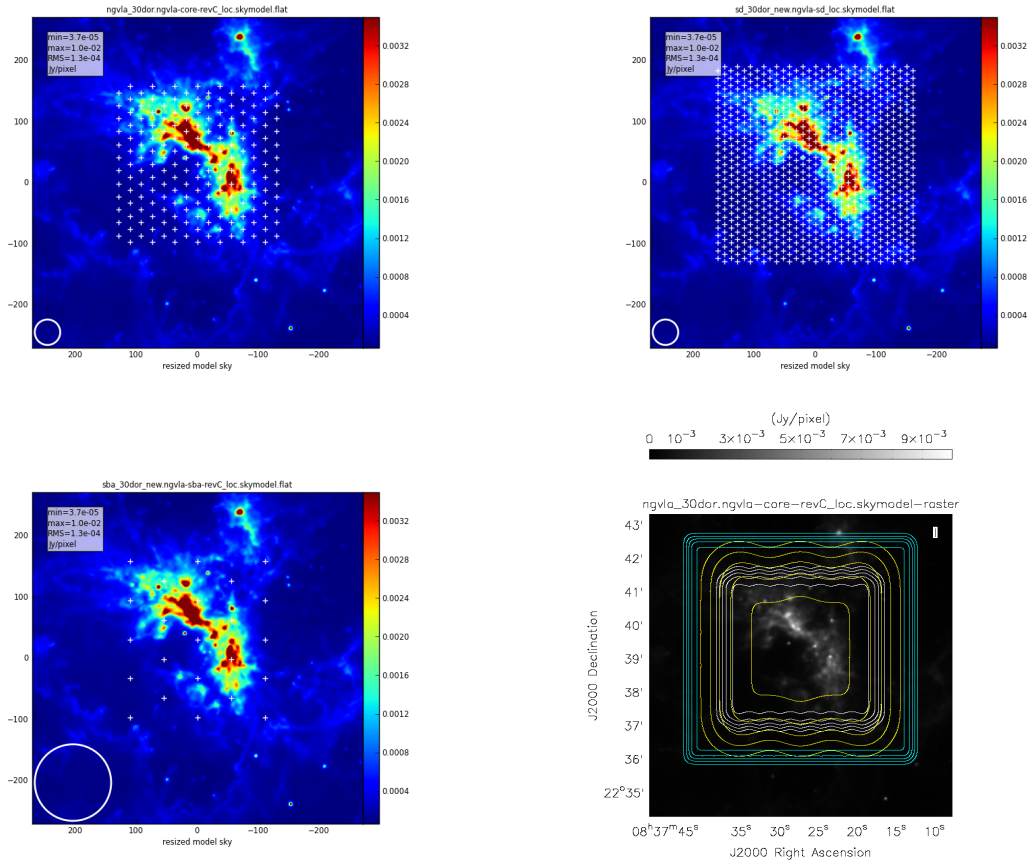

Figure 2: Simulation input with ngVLA core (top left), Total Power (top right), and Short Baseline Array (bottom left) antenna pointings. The lower right panel shows the total power weight map (cycan), the SBA mosaic primary beam (yellow), and the ngVLA core mosaic primary beam (white), along with the input model in the background. Primary beams are shown at levels of 20%, 40%, 60$, 80%, and 98% of the peak.

divided by the true peak sky intensity— but also has the significant drawback that it is by definition driven by the values of one or two pixels in the image, so does not quantify the typical imaging errors present in the image which in many cases are more scientifically important, and are more likely to correspond to intuitive impressions of image fidelity.

Most ALMA development studies (e.g., Tsutsumi et al. 2004) used a different metric of image fidelity[5]. Defining the fidelity of a given pixel i as

$$f_i = \frac{|M_i|}{|M_i - I_i|} \tag{5}$$

the overall image fidelity is then evaluated as the median of f_i values for pixels over some threshold of fractional peak intensity in the true sky image, for instance:

$$f_{alma,1\%} = Median(f_i|_{>1\%}) \tag{6}$$

Tsutsumi et al. (2004) used several thresholds (0.3% , 1%, 3%, 10%). This metric goes to infinity for a perfect image; unity when the M_i are $>> I_i$; and zero when the I_i are $>> M_i$. Its value can be sensitive to the chosen input image fractional intensity threshold. It also treats the true sky image and the model sky image asymmetrically in selecting the set of pixels to evaluate. This has the consequence that *image artifacts spatially removed from true sky structures do not affect the value of the image fidelity metric.* This will be discussed further below.

Versions of the ngVLA science requirements (ngVLA memo 65) prior to the writing of this document define image fidelity in two different ways. The immediately prior version (May 2019) posits

$$F_2 = 1 - \frac{\sum_i M_i |M_i - I_i|}{\sum_i M_i I_i} \tag{7}$$

which is equivalent to a weighted sum of the fractional error, with the weight being the product of the true and model sky images:

$$F_2 = 1 - \frac{\sum_i M_i I_i \frac{|M_i - I_i|}{I_i}}{\sum_i M_i I_i} \tag{8}$$

Here the fractional error is with respect to the formed image I_i. The original ngVLA definition (also adopted by Rosero et al. 2019) is

$$F_{2b} = 1 - \frac{\sum_i M_i |M_i - I_i|}{\sum_i M_i^2} \tag{9}$$

corresponding to a weighted sum of the fractional error with respect to the input (truth) image. These metrics are also unity for a perfect image and unbounded from below, and

[5]The formally adopted ALMA image fidelity requirement (2006-07-28-ALMA-90.00.00.00-001) is: "images shall be thermally limited at all points where the brightness is greater than 0.1% of the peak image brightness. This requirement applies to all sources visible to ALMA that transit at an elevation greater than 20 degrees at frequencies below 370 GHz."

they too have the undesirable characteristic that *pixels where either the true or model sky values are zero receive zero weight*. Therefore they are also insensitive to spurious structures in the reconstructed image in regions where the true sky pixels are close to zero; nor are they very sensitive to missing structures, if the reconstructed image values are close to zero. One result is that the dirty image of a point source— even when the Point Spread Function has dramatic sidelobes— can have an image fidelity quite close to unity. To quantitatively examine this behavior I calculated the fidelity metrics for an SBA point source dirty image and the corresponding cleaned image (Fig. 3 and Table 1).

In order to better quantify the effects of both missing and spurious structure I propose an alternate definition of image fidelity, F_3, defined as

$$F_3 = 1 - \frac{\sum_i \beta_i W_i |M_i - I_i|}{\sum_i \beta_i^2 W_i} \tag{10}$$

where $\beta_i = Max(|I_i|, |M_i|)$. It treats the images symmetrically, and in particular, it *provides significant weight to both missing flux and spurious features regardless of where they occur in the image*. I have also allowed a "window function" W_i which explicitly defines the region over which the fidelity is to be evaluated; within this region its value is 1, while it is 0 outside. This consideration is relevant in particular for mosaic imaging and feathering total power, since the spatial regions covered differ because of differing primary beams and the need for a total power "guard band" around the interferometric image, while the images themselves need to be co-registered for feathering. By default in this work I take $W_i = 1.0$ inside the 0.5 contour of the mosiac primary beam response. For other use cases, such as high angular resolution imaging of a debris disk, it will be much smaller. In these cases W_i could be implicitly defined simply the the size of the image chosen to reconstruct. The values of F_3 range from 1.0 for a perfect reconstruction to ~ 0 for M_i which are uncorrelated with the true sky intensity pixels P_i. It will in general only be negative if the reconstructed sky pixels are systematically anti-correlated with the true sky pixels values. The values of F_3 are also shown for the SBA point source case in Tabel 1. I define the fidelity error for an individual pixel j, following Eq. 10, as

$$\delta F_{3,j} = \frac{\beta_j |M_j - I_j|}{\sum_i W_i \beta_i^2} \tag{11}$$

Here I have assumed that this is evaluated for a pixel j within the region of interest ($W_j = 1$). This quantity is useful for determining what dominant features are responsible for deviations from perfect fidelity in the formed image. One minus the sum of $\delta F_{3,j}$ values over the region of interest in an image will equal F_3 for that image.

The Pearson correlation coefficient c was also calculated between the input model and the formed image. Like the fidelity comparisons the reference image is the model input smoothed with a Gaussian restoring beam of the size and shape given in the formed image header.

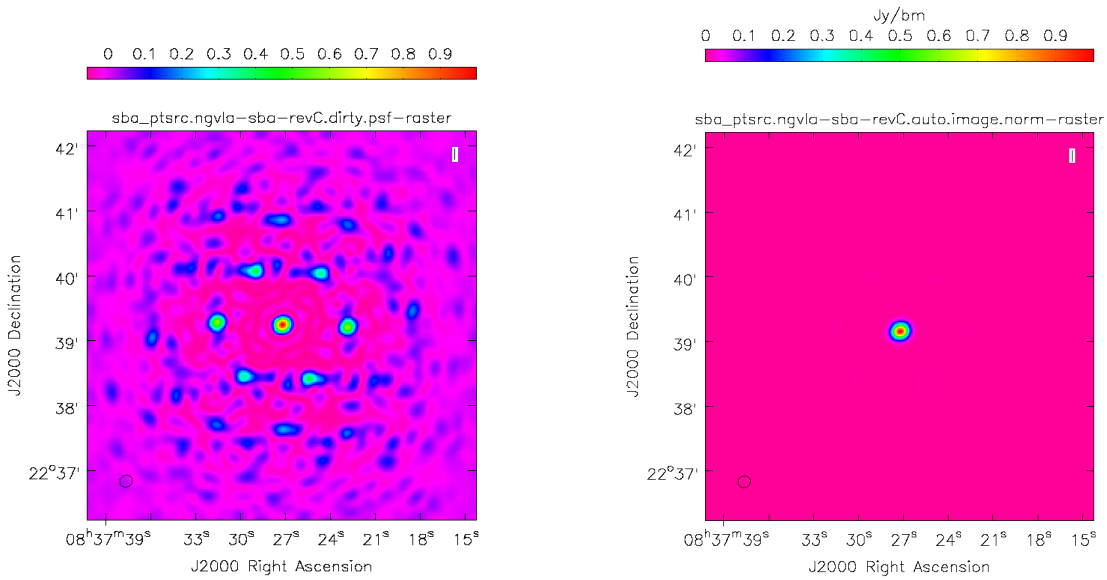

Figure 3: Dirty image (left) and auto-cleaned image of a point source observed with the SBA. By previous ngVLA image fidelity definitions (F_2, F_{2b}) both of these have fidelities within $\sim 5\%$ of unity (i.e. rather high fidelity).

Target	Array	Mask	F_1	F_2	F_{2b}	F_3	c
PtSrc	SBA	Dirty	0.615	0.946	0.949	0.507	0.71
PtSrc	SBA	Auto	0.996	0.999	0.999	0.999	0.999

Table 1: Fidelity metrics evaluated for an SBA dirty image of a point source and an auto-cleaned SBA image of a point source. c is the Pearson correlation coefficient and the image fidelities F_1, F_2, F_{2b} and F_3 are as defined in Eq. 4- 10.

5 Procedures

5.1 Total Power Imaging

I found that high-fidelity simulation and imaging of total power data for this use case required care and some special procedures. My initial attempt at simulating total power observations of 30 Dor straightforwardly comprised running `simobserve` with appropriate values (e.g. $D = 18m$ and the 30 Dor model already shown) and gridding the total power "MS" into an image following the guidance in the M100 CASAguide. Specifically, a `SF` gridding function was used with a support region[6] of $N = 7$ and 9 pixels per primary beam. Results were evaluated in comparison to the model imaged smoothed by a Gaussian with a width as specified in the total power image header. **This resulted in the simulated single dish image recovering only** 56% **of total flux of the input image**. A comparable image fidelity ~ 0.6 (with no noise corrupting the simulation) is obtained. The peak flux density per beam is also only 65% of what is obtained by smoothing the input model smoothed with the $47.7''$ (FWHM) Gaussian beam given in the image header.

Following this result, I simulated an observation of a known point source of 1 Jy. This resulted in a total flux low by exactly the same factor 0.56, and a peak surface brightness low by a factor of ~ 0.8. Inspecting the profile of the simulated point source, I find an actual, non-parametrically evaluated FWHM of the profile of $42.8''$, while a Gaussian fit to the profile gives a FWHM $40.6''$. The Gaussian fit was performed two different ways: first, with `au.fitFITSbeam`, and second, with completely independent code in IDL; results agreed to within $0''.01$. The actual primary beam for an 18m dish with 0dB edge taper is $37''.91$ (FWHM). The analysisUtil-ity `au.sfBeam` predicts a FWHM of the *gridded* primary beam of $42''.81$, in exact agreement with the observed PB FWHM; `au.sfBeam` gives a FWHM for the Gaussian fit to that of $42''.27$, somewhat larger than our observed Gaussian FWHM $40''.6$. *All* of these are *significantly* narrower than the $47.7''$ FWHM in the image header. For reference the relevant parameters for these calculations are: a central frequency of 93 GHz; 18m primary; 0dB edge taper; convsupport of $N = 7$ (width in pixels); and $4''.3$ pixels.

Correct peak surface brightnesses are obtained if the simulated single dish data is corrected so that a simulated point source gives the expected peak surface brightness of 1 Jy/bm. For our particular use case this implies scaling up the image by a factor of $1/0.78 = 1.28$. To get correct integrated total fluxes the beam size in the header also needs to be correct. I adopted the $40''.6$ FWHM Gaussian obtained by fitting the simulated point source. This procedure delivers $F_3 = 98.6\%$ fidelity for a simulated 30 Dor map and $F_3 = 97.0\%$ fidelity for a simulated point source map. The fidelity improves slightly if I use the actual simulated point source image as the PSF to compute the reference image in the fidelity calculation. In this case I find for 30 Dor $F_3 = 99.8\%$, and for the point source $F_3 = 99.6\%$. When using the exact PSF I also first rescaled the PSF to have unity peak, effectively applying the scale factor (1.28) previously mentioned; and put the Gaussian

[6]For most of this work $N = 7$ was used. The CASAguide recommends $N = 6$, which I also evaluated. The essential results were unchanged.

fitted value of 40″.6 for the beam FWHM in the image header for purposes of subsequent total flux calculations, since the brightness unit is Jy/bm. *When these physically-based corrections are applied, no arbitrary SDFACTOR is needed in the feather step to get correct total flux densities in the final, feathered image.*

Using the actual PSF and the input 30dor image, we expect a peak image surface brightness of 126 Jy/bm, in comparison to the observed 98.9 Jy/bm observed. One question is, how much of this is due to the simulation, and how much due to the imaging? To answer this I inspected the time-ordered raster data in the TP MS, finding a peak signal intensity of 106.8 Jy/bm (in scan 468, or fieldID 467). *This value is low by a factor of* 0.848, suggesting that whatever "de-gridding" procedure is used in the single dish simulation is responsible for the majority of the discrepancy[7]. However, the final SD simulation image is still low by $\sim 8\%$ from what would be expected in the time-ordered data (98.9 vs. 106.8 Jy/bm). This bias is occuring in the `sdimaging` step.

The M100 CASAguide does emphasize that it is necessary to directly calibrate the total power images using single dish maps of calibrators. One conclusion of this work is that the same sort of calibration step is needed for *simulated* total power observations as well. Preliminary inspection of an ALMA single dish pipeline cube suggests that they do *not* suffer from significantly erroneous beam sizes in the header. I did not check a manual TP reduction. Possible explanations of the observed differences include: 1) the task `sdimaging` does not correctly pick up the antenna illumination specified in the `vptable`, and instead uses defaults which are appropriate to ALMA (but not to the ngVLA) in deriving beam size from antenna diameter; 2) a different task is used for imaging (e.g., `tsdimaging`); 3) the beam size information in the header is separately corrected after imaging.

5.2 CLEAN Mask Definition

Deconvolving spatially extended structures is challenging, and the accuracy of the resulting images is often limited by the systematic errors and uncertainties inherent to the CLEAN process itself. These can to some extent be mitigated by careful CLEAN masking, which are therefore crucial in evaluating extended source imaging performance.

A general purpose, iterative auto-masking algorithm has been developed in support of the ALMA imaging pipeline (Kepley et al. 2020). This algorithm, called `auto-multithresh` in CASA, has been extensively validated against a range of real-world datasets. I chose to use `auto-multithresh` to facilitate reproducibility, and because the ngVLA operations plan calls for most data products to be automatically generated and delivered to PI's "Science Ready". I initially tried both ALMA 7m (for the SBA) and ALMA 12m-compact (for the ngVLA core) parameters. I also tuned the parameter specifically for the SBA and ngVLA core visually in an interactive clean session with somewhat better results. The main sense of the tuning was to make the box placement more conservative; to clean deeper

[7]Careful inspection of the input image at the RA and Dec of scan 468 shows that only about 2% of this can be ascribed to the fact that scan 468 is slightly offset spatially from the peak of predicted emission in the map.

Map	Total Flux [Jy]	Beam FWHM	Peak Brightness
30 Dor Input Model	**1432**	n/a	0.01 Jy/pix
Smoothed 30 Dor Model	1418	47″.7	154.3 Jy/bm
Smoothed, Regridded 30 Dor Model	1433	47″.7	154.3 Jy/bm
SD Simulation (SF, $N = 7$)	812	47″.7	98.9 Jy/bm
Point Source Model	**1**	n/a	1 Jy/pix
PtSrc, SD Simulation (SF, N=7)	0.56	47″.7	0.78 Jy/bm
PtSrc, SD Sim (SF, N=6)	0.60	46″.5	0.83 Jy/bm
PtSrc model smoothed by exact PSF	0.998	40″.6	1.0 Jy/bm
PtSrc Sim. (corrected beam and scaling)	1.000	40″.6	0.995 Jy/bm
30 Dor model smoothed by exact PSF	1436	40″.6	127 Jy/bm
SD Simulation (corrected beam and scaling)	1439	40″.6	126 Jy/bm

Table 2: Integrated flux density and peak surface brightness for a known model; smoothed and regridded versions of the known model; and images made from simulated single dish observations of these models. Two known models are considered: a point source and a 30-Dor-like science target. FWHM values are the values listed in the single dish image header after corrections discussed in the text. The final four lines show that simulated single dish observations+imaging of both a point source and the science target model give consistent results (total flux density and peak surface brightness) as what is obtained directly convolving the input model with the exact PSF (derived from the point source simulation) and applying corrections discussed in the text. It is also notable that the smoothing and regridding of the input model (top 3 lines of the table) do not introduce significant systematic errors.

Array	T_{side}	T_{noise}	$f_{minBeam}$	T_{low}	g	S	f_{cycle}
ALMA-7m	1.25	5.0	0.1	2.0	75	1.0	1.0
ngVLA-SBA	1.5	5.0	0.1	2.5	50	0.75	0.75
ALMA-12m	2.0	4.25	0.3	1.5	75	1.0	1.0
ngVLA-core	3.5	5.0	0.3	2.5	50	0.75	0.65

Table 3: Parameters used for `auto-multithresh` automated clean masking.

within each major cycle; and to make the boxes tighter. The `auto-multithresh` parameter values used are shown in Table 3.

Overall four strategies were employed for defining the CLEAN masks used:

- No mask (dirty image) – only evaluated for SBA-only case.
- TCLEAN automasking (`auto-multithresh` – Kepley et al. 2020.) using ALMA 7m (SBA) or 12m-compact (ngVLA-core) parameters.
- TCLEAN automasking (`auto-multithresh`) using parameters tuned for the SBA and ngVLA-core individually.
- Interactive, user-defined masking (ngVLA-core only).

Results from these different approaches are discussed in § 5.3. Note that the scripts on github only include those needed to reproduce the primary results of this investigation (the tuned automask method).

5.3 Interferometric Imaging

Several deconvolution and imaging strategies were used in order to characterize the impact of reconstruction method on the final result. The methods used to image the individual interferometric arrays' data, along with identifying names I assigned them, are as follows:

- For the SBA:
  - Sba-Dirty: form the dirty image.
  - Sba-AutoAlma: multi-scale deconvolution using automasking with ALMA 7m default parameters.
  - Sba-AutoTuned: multi-scale deconvolution using automasking (ngVLA tuned parameters).
  - Sba-SDmod: same as Sba-AutoTuned, but use the total power map as a starting model in `tclean`.
  - Sba-Feather: same as Sba-AutoTuned, but feather total power image in at the end.

  - Sba-SDmod-Feather: same as Sba-SDmod, but feather total power image in at the end.

- For the ngVLA core:
  - Core-User: interactive clean masking at each major cycle.
  - Core-AutoAlma: multi-scale deconvolution using automasking with ALMA 12m compact configuration default parameters.
  - Core-AutoTuned: multi-scale deconvolution using automasking (ngVLA tuned parameters).

All images were cleaned with a threshold of 1% the peak brightness seen in an initial simple clean of the brightest emission. In all SBA and joint cases the clean deconvolution was stable in the sense that there were no indications of divergence (e.g., the clean terminated by reaching the threshold). The ngVLA core-only cleans were less stable and typically did show some divergence; this isn't terribly surprising considering the extent of emission in the field in comparison to the spatial frequencies that these data sample well. For optimal surface brightness sensitivity natural weighting was used. To feather in the total power data I followed the procedures described in the M100 CASAguide.

Best practices suggest that the optimal approach to imaging multi-array data is to perform the deconvolution jointly, *i.e.*, including all of the interferometric data in a single deconvolution. This is the recommendation for ALMA 7m+12m data. For recent CASA versions prior to 5.6 it was not possible to correctly handle joint deconvolution of multi-array data (other than ALMA) in CASA using standard tools. The joint (SBA+core) image fidelities in Version 1 of this memo were affected by this limitation: the best-case image fidelities for TP+SBA+core images were in the 70% to 75% range and there were visually discernible, SBA-pointing related artifacts in the deconvolved images. The issues were more pronounced than those seen by Kundert et al. (2017), who imaged ALMA 7m+12m data using only the 12m primary beam (as a test of systematics). They found a limiting dynamic range ~ 70, corresponding to $F_3 \sim 0.98$ in ngVLA terms, which is more than adequate for purposes of the current study. The difference in the ngVLA and ALMA cases could be due to the fact that the ALMA 7m and 12m antenna diameters are more similar than the ngVLA SBA and main arrays.

With the release of CASA 5.6/6.1 it became possible to correctly handle heterogeneous, non-ALMA interferometric data using (mostly) standard tools[8]. Once it was appreciated that CASA 5.6/6.1 could correctly handle the SBA+core case, I repeated the joint imaging cases using procedures detailed in Appendix A of this memo. All of the simulations, imaging and analysis in this memo were done in CASA 5.4 except these SBA+core images, which were done in CASA 6.1.

[8]Some relevant CASA issues are `CAS-8592`, `CAS-11271`, `CAS-11464`, `CASR-301`, `CASR-470`, and `CAS-13010`. It was not generally expected that CASA 5.6/6.1 would handle this case correctly, a fact which was discovered in the course of testing `CAS-13010`

Two other procedures bear mentioning: it is necessary to manually adjust the relative weights of the 6m and 18m data by a factor of $(6/18)^4$; and, following the advice of the Data Combination 2019 working group, I concatenated the SBA and core MS's in *time order*. The methods used to image the data jointly were then as follows:

- Joint-AutoTuned: multi-scale deconvolution of SBA+Core using automasking (ngVLA core parameters).
- Joint-SDmod: same as Joint-AutoTuned, but using total power model as initial clean model.
- Joint-Feather: same as Joint-AutoTuned; feather total power in at the end.
- Joint-SDmod-Feather: same as Joint-SDmod; feather total power in at the end.

All images were cleaned with a threshold of 1% the peak brightness seen in an initial simple clean of the brightest emission. Unlike the previous attempt, the joint deconvolutions were stable and showed no particular signs of instability or divergence. In the process of repeating this analysis I identified a mistake in the SDmod procedure in analysis presented in v1: the initial single dish model should have been mutliplied by the interferometer mosaic primary beam prior to being passed into `tclean` as a startmodel and was not. Fixing this helped improved the SDmod results, though most of the improvement is from using the correct 6m primary beams (see discussion below).

6 Results

6.1 Imaging Results

The fidelity metrics for all of these images are presented in Table 4, along with the total flux density measured in the same large aperture in each image. The four best SBA and SBA+TP images are shown in Figure 4. When total power is included, the SBA robustly recovers accurate ($\pm 1\%$) total flux densities for the object and produces a high fidelity ($F_3 \sim 0.95$) image. In the absence of total power information the best fidelity achieved is 32%; in this case only about 20% of the total flux is recovered. When the TP is used as a starting model *only*, fair results are obtained: 86% fidelity and a total flux accurate to within $\sim$ 20%. Clearly superior results are obtained by including the TP data via feathering. Images of the fidelity error (Eq. 11) and the fractional flux error are presented in Figure 5. The ngVLA core-only results, not surprisingly, are poor: a few percent of the total flux is recovered and the image fidelity is correspondingly on the order of 10%. Tuned automasking gives the best results, though this is a question of the lesser of three evils. The best core-only image is presented in Figure 6.

The joint SBA+core images are a dramatic improvement on the ngVLA core-only images, and successfully reproduce higher-resolution structures that cannot be seen in the SBA images; see Figure 7. The total flux densities are similarly accurate when total power

Image	Flux [Jy]	F_1	F_2	F_{2b}	F_3	c
Sba-dirty	n/a	0.35	−10.8	0.08	0.08	0.28
Sba-AutoAlma	224.8 (−80%)	0.61	−1.15	0.32	0.32	0.83
Sba-AutoTuned	238.8 (−79%)	0.60	−1.16	0.32	0.32	0.84
Sba-SDmod	1436.8 (+17%)	0.74	0.87	0.85	0.86	0.98
Sba-Feather	1214.2 (−1%)	0.92	0.93	0.93	0.93	0.99
Sba-SDmod-Feather	1232.3 (+0.6%)	0.88	0.95	0.95	0.95	0.99
Core-User	17.3 (−99%)	−0.02	−9.1	0.07	0.09	0.20
Core-AutoAlma	35.2 (−97%)	−0.81	−5.63	−0.02	0.10	0.18
Core-AutoTuned	71.0(−96%)	0.38	−5.34	0.14	0.14	0.62
Joint-AutoTuned	251.0 (−80%)	0.72	−0.51	0.40	0.40	0.81
Joint-SDmod	1459.4 (+19%)	0.83	0.83	0.82	0.83	0.97
Joint-Feather	1214.2 (−1%)	0.89	0.90	0.90	0.90	0.97
Joint-SDmod-Feather	1259.0 (+3%)	0.86	0.92	0.92	0.92	0.98

Table 4: Image fidelity and total flux retrieval results for the range of datasets and reconstruction techniques discussed in the text. The total flux density of the input model over the region considered— defined to encapsulate the region within the 50% contour of the ngVLA core mosaic primary beam— was 1224.8 Jy. This region is slightly different from the one considered in Table 2. The case giving best overall fidelity is shown in **bold** for each array combination. *Note that our preferred metric is F_3, the next to last column in the table.*

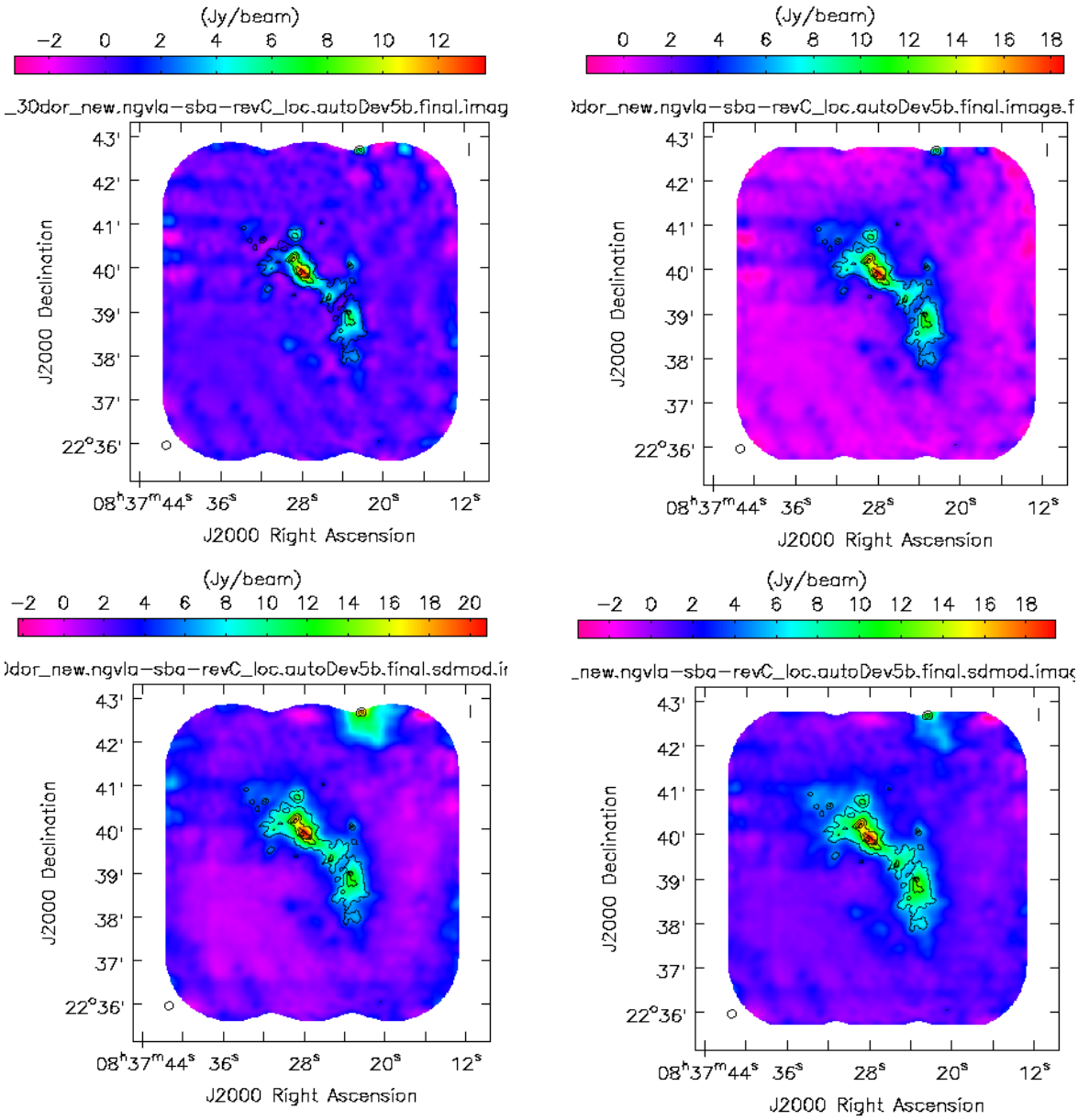

Figure 4: Recovered Short Baseline Array images. Top left: SBA alone; Top right: SBA+TP (feathered); Bottom left: SBA with TP starting model for CLEAN; Bottom right: SBA with TP start model, and TP feathered in afterwards. Black contours represent 20%, 40%, 60%, and 80% of the input model peak intensity (smoothed by the restoring beam).

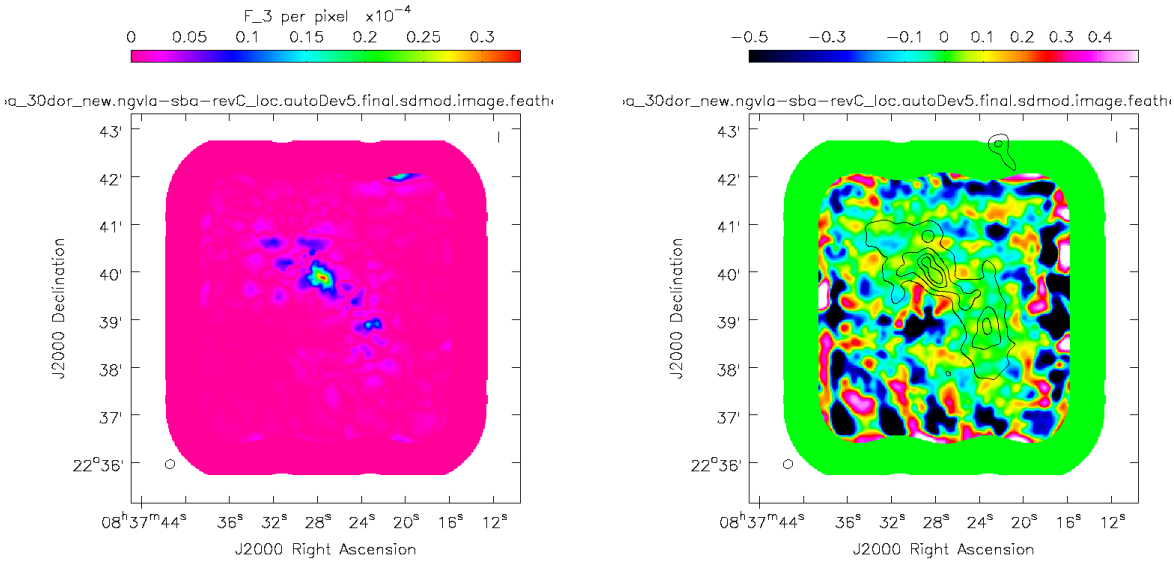

Figure 5: Fidelity error (left, following Eq. 11) and fractional error in reconstructed image (right) for the SbaSdmodFeather case. Black contours show the input model at the resolution of the restoring beam.

is feathered into the images (±1%). They meet the ngVLA project's quantitative fidelity goal, with fidelities of 90% and 92% for joint-feather and joint-SDmod-feather, respectively. The total flux accuracy of joint-feather is 1%, and that of joint-SDmod-feather is 3%. Overall the methods produce similar results. The joint-SDmod-feather case benefited notably from the total power startmodel issue noted previously (related to the mosaic primary beam), although the fact that joint-feather showed almost as much improvement demonstrates that most of the improvement is due to using the correct 6m primary beam (i.e., the fully correct heterogeneous array imaging capability described in Appendix A).

6.2 RevC.01 Configurations

The RevC.01 SBA simulations were imaged identically to the other SBA images discussed above. The fidelity results were identical to the previous/nominal Rev.C configuration to within 1%. The RevC.01 configuration could be expected to produce improved fidelity (compared to Rev.C) for tracks at lower elevation, particularly if the simulations include thermal noise, but these considerations were not within the scope of this investigation.

6.3 Comparison of Fidelity Metrics

In terms of the performance of the fidelity metrics, F_{2b} and F_3 track relatively well unlike the case shown in Figure 3. F_2 does as well except at low fidelity— $F_3 < 0.5$ say— where its

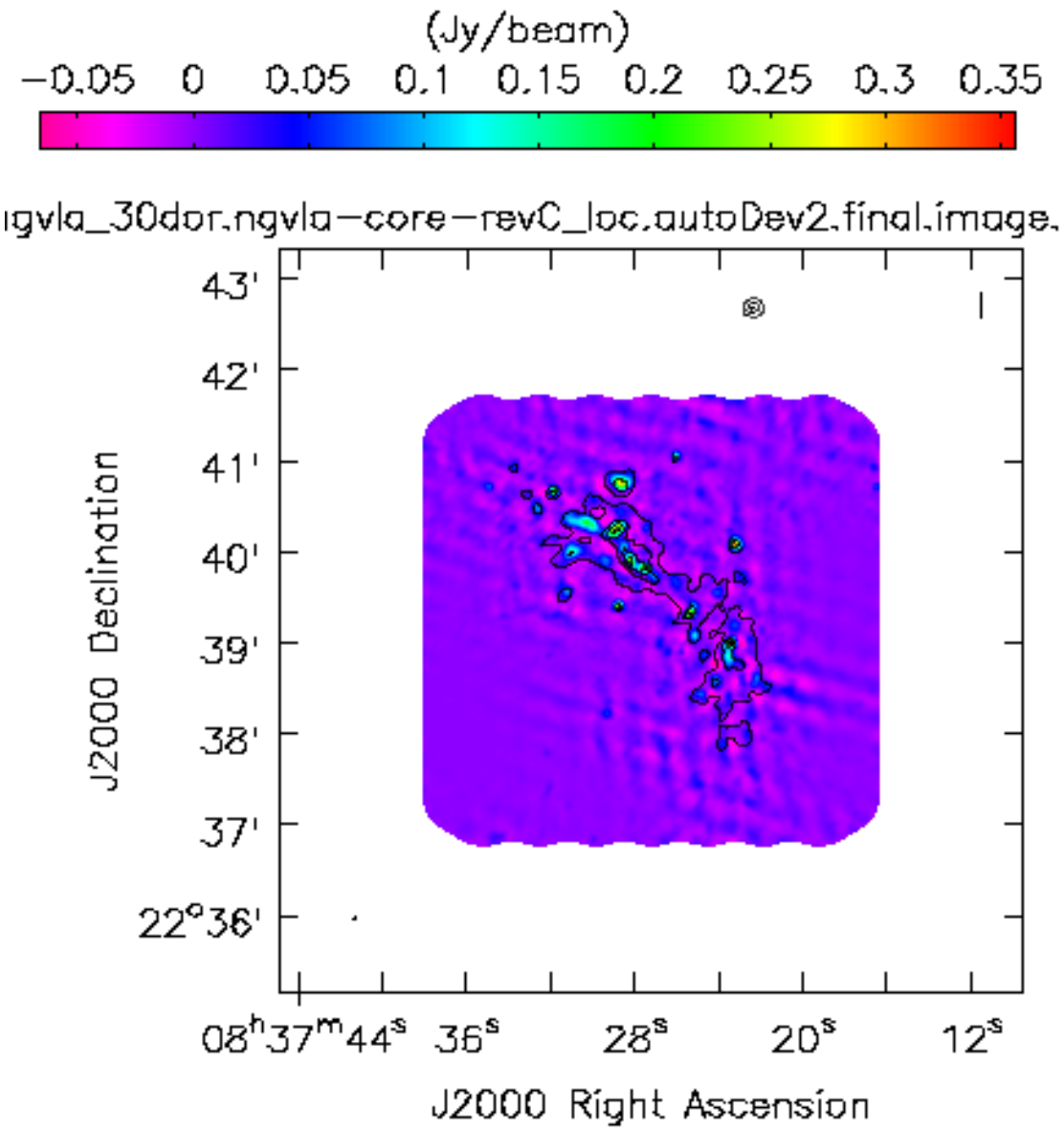

Figure 6: ngVLA core (only) recovered image. Black contours represent 20%, 40%, 60%, and 80% of the input model peak intensity (smoothed by the restoring beam).

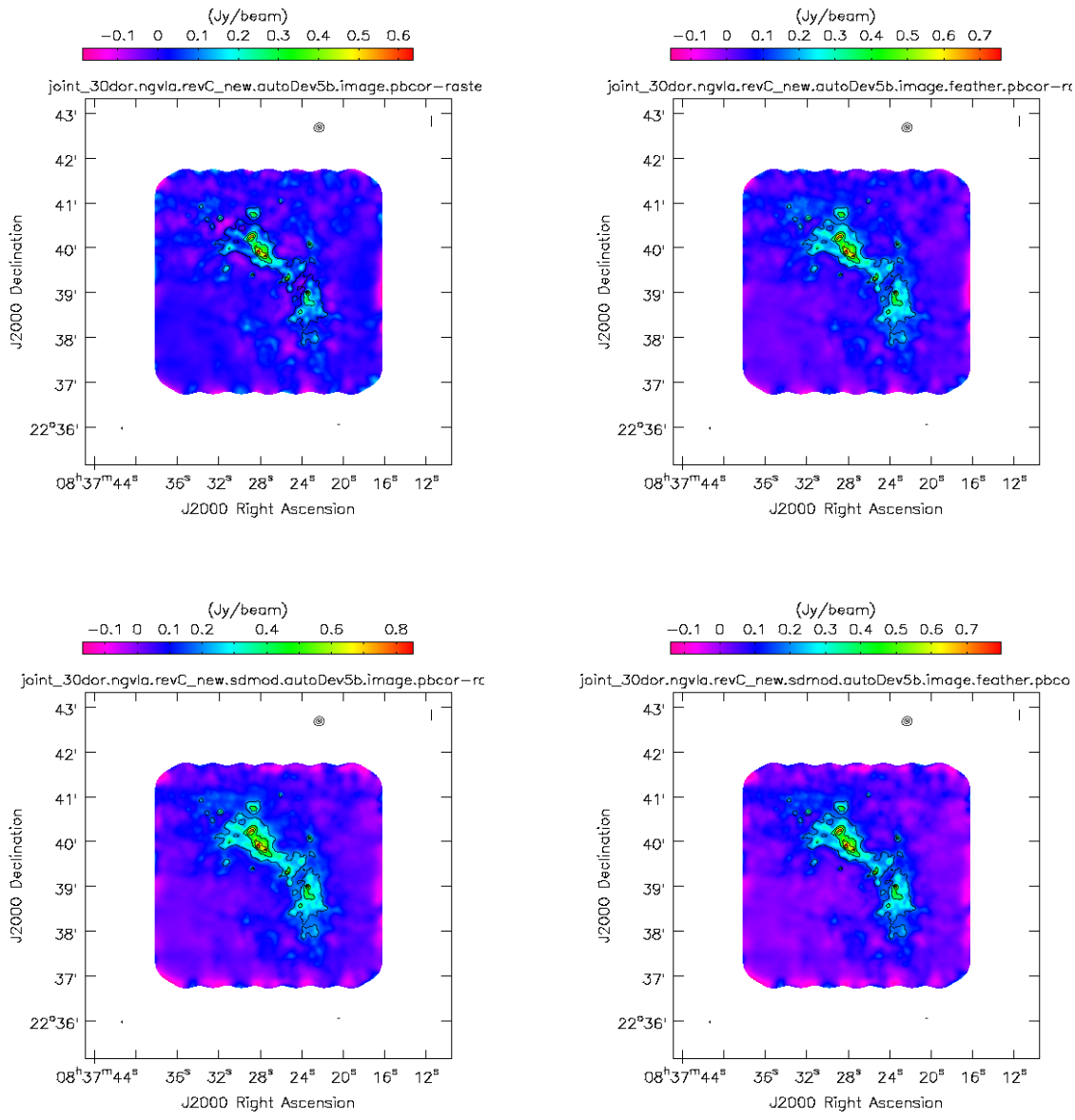

Figure 7: Recovered Joint SBA+Core images. Top left: Joint-AutoTuned; Top right: Joint-Feather; Bottom left: Joint-SDmod; Bottom right: Joint-SDmod-Feather. Black contours represent 20%, 40%, 60%, and 80% of the input model peak intensity (smoothed by the restoring beam).

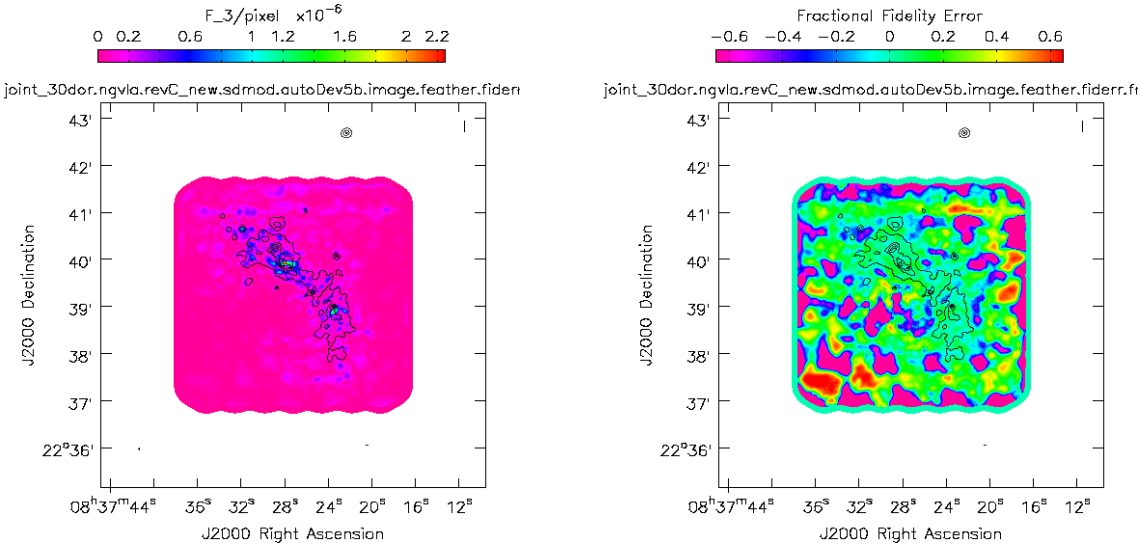

Figure 8: Fidelity error (left, following Eq. 11) and fractional error in reconstructed image (right) for the JointSdmodFeather case. Black contours show the input model at the resolution of the restoring beam.

results are generally nonsensical. F_1 sometimes tracks our preferred metric and sometimes doesn't, the most striking discrepancies being Sba-AutoTuned and Sba-AutoAlma ($F_3 = 0.32$ vs $F_1 \sim 0.6$). It is notable that F_1 is a fundamentally different metric in that it is in some sense not local: the numerator and denominator of Eq. 4 need not reference the same pixel, whereas the other metrics are all different forms of weighted fractional accuracies of individual pixels.

6.4 Aperture Flux Analysis

The accuracy of flux recovery as a function of angular scale was evaluated by first using a peak finding heuristic to identify peaks above 4.5σ in the reconstructed image, and measuring the ratio of integrated flux within an aperture of some diameter to the integrated flux in the same region of the simulation input ("truth") image. Figure 9 shows the median fraction of flux recovered as a function of aperture radius for each of the core (only), the SBA (interferometer only), and the Joint-SDmod-Feather image. The median fraction here is taken over individual apertures centered on the identified peaks. This analysis is similar to that performed for the NGVLA Cygnus-A simulation in ngVLA memo 85 (Rosero et al. 2020). Aperture diameters ranged from $2\times$ the beam FWHM of the array in question up to $90''$, about $1.5\times$ the SBA λ/b_{min}.

The NGVLA core and SBA interferometer alone recover at best $\sim 25\%$ to $\sim 60\%$ of the total flux, and considerably worse for large apertures ($D > 1'$ say). The performance

of either in this case is worse than what is seen in the LAS analysis of § 2, which considers flux retrieval of uniform disks of varying radii. Presumably this is due to the presence of varying levels of diffuse background emission in this case which is not present in the disk simulations. The results of Figure 9 show that at the nominal LAS of the SBA ($0.58\lambda/b_{min}$) $\sim 82\%$ of the maximum measured flux is recovered (compared to the maximum recovered by the SBA not the fraction of the total true flux, which is even less); at λ/b_{min} 51% of the peak flux is recovered. For the core alone these figures are worse: only 25% of the peak flux is recovered at the nominal LAS ($0.85\lambda/b_{min}$), and only 19% is recovered at λ/b_{min}.

The combination of the NGVLA core, the SBA interferometer, and total power data accurately recovers the input flux to within 2% for all aperture sizes measured *i.e.* up to $1'.5$ in diameter, at which point the mosaic size starts to become a limitation. This figure also shows, for the Joint-SDMmod-feather case, the 32 peaks identified by the peak finding heuristic which were the centers of the apertures for that measurement.

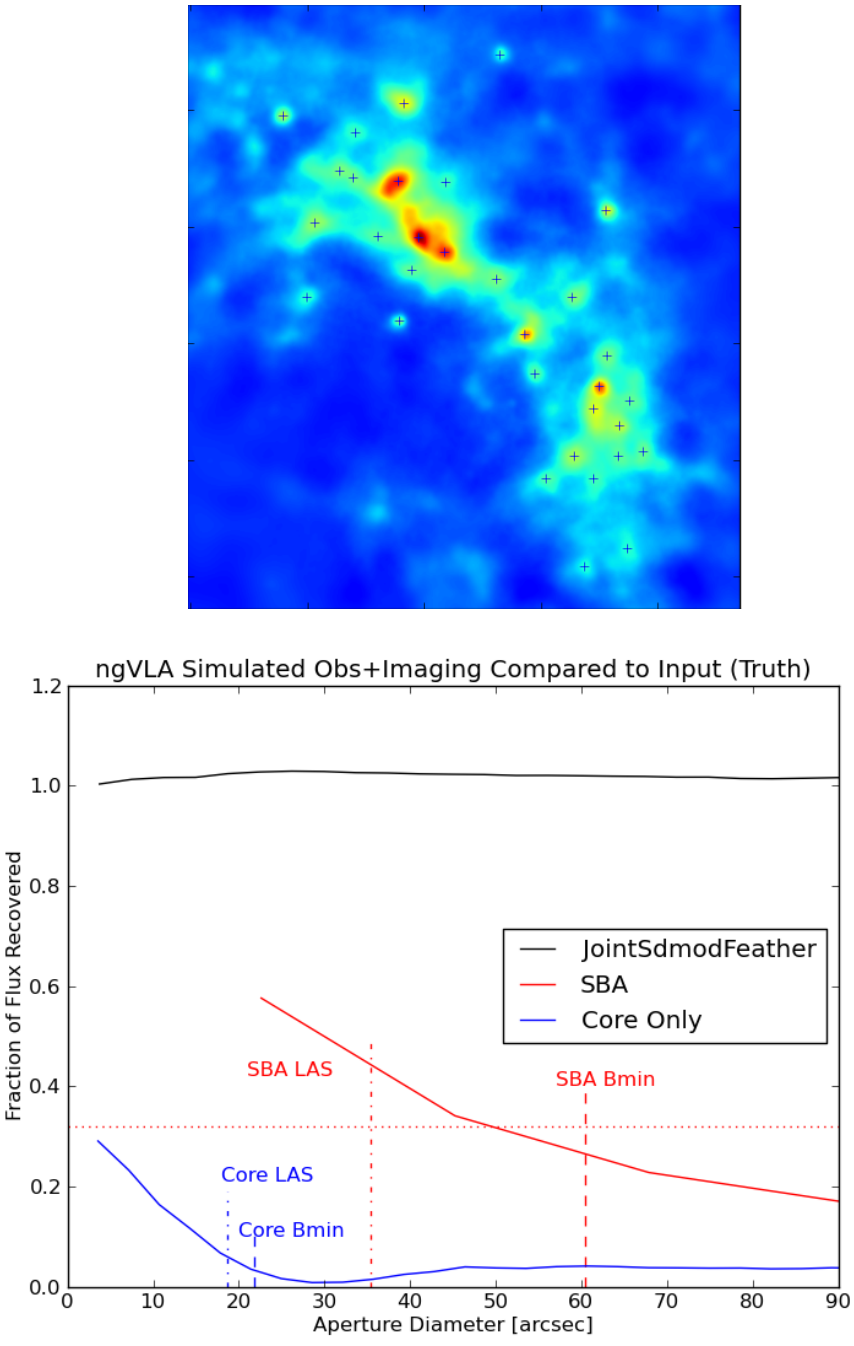

Figure 9: **Top:** The Joint-SDMOD-Feather reconstructed image. Black markers denote the local maxima which were the centers of the regions for aperture photometry. **Bottom:** Median fraction of integrated flux recovered as a function of aperture radius for the joint-SDmod-feather image, the SBA-only image, and the core only image. The dashed vertical lines show λ/b_{min} for the core and SBA, as well as the LAS values determined in § 2. The dotted horizontal line shows the measured fidelity (F_3) for the SBA, which should be comparable to the fraction of flux recovered.

6.5 Restoring Beam Flux Bias (Jorsater-VanMoorsel Effect)

When imaging extended objects the flux scale difference between the cleaned and residual maps can be a sigificant source of systematic error (Jorsater & VanMoorsel 1995, hereafter JvM). The effect can be particularly pronounced when multiple interferometer array configurations are combined, due to approximating a multi-resolution PSF with a single Gaussian PSF. Walter & Brinks (1999), for example, find that integrated flux densities from their multi-configuration VLA data (B+C+D) would be overestimated by a factor of two without the correction.

I calculated the magnitude of the restoring beam area bias correction using the expressions given in JvM and find it is a modest correction for the current use cases. Numerically, I find a correction factor (as defined in JvM) η of 1.14 for the SBA; 1.094 for the ngVLA core, tapered to $2''$; and 1.078 for the joint core+SBA. Applying the corrections typically resulted in a modest (1% - 2%) change in fidelity; it did not consistently improve fidelity.

7 Conclusion

The major conclusions of this work are as follows:

- For the simple case of imaging a uniform disk, the LAS of the ngVLA core and SBA individually are $\sim 17'' \times (\nu/100\,\mathrm{GHz})^{-1}$ and $\sim 33'' \times (\nu/100\,\mathrm{GHz})^{-1}$, respectively.
- When combined with 18m total power, the SBA robustly produces satisfactorily high-fidelity images ($F_3 \sim 94\%$) of structures extended over many fields of view and accurate tota flux densities to within 1%. In the absence of total power the image fidelity is $F_3 \sim 30\%$.
- The ngVLA core, when augmented by SBA and total power, produce images with $F_3 \sim 91\%$ and accurate total flux densities to within 3% or so. In the absence of total power only about 20% of the total flux is recovered.
- Feathering provides more accurate total power information than just using the total power data to initialize the interferometric clean model, although the latter is better than not using any total power information at all. Combining the two methods (SDmod+feather) method delivers slightly higer image fidelity than feathering alone.
- With the appropriate procedure, high -fidelity joint ngVLA core+SBA can be produced. The procedure is illustrated in the appendix to this memo.
- The RevC.01 configurations, as expected, give very similar results as the Rev.C configurations mostly examined here. The LAS of the RevC.01 core is $\sim 10\%$ smaller than that of Rev.C.
- Based on a quantitative evaluation, I recommend and use an improved fidelity metric F_3. For high-fidelity reconstructions, F_3 and the *original* ngVLA fidelity metric (F_{2b})

give very similar results for the main use cases I examined, though they differ for low-fidelity images. F_2 and F_1 are less well-behaved.

- Care is required to produce quantitatively correct single dish simulations. This being done there is no need for arbitrary scalings of the single-dish data in `feather()` (i.e., SDFACTOR can be left at unity). The most important issue to note here is that the beam size that `sdimaging` writes in the header is incorrect for our use case, possibly because of differences in the assumed antenna illumination profile (ngVLA: airy; ALMA: more heavily tapered). This underscores the importance of having well-calibrated, well-understood inputs to any combination method.

Interferometric imaging of objects of size $>> \lambda/b_{min}$ is intrinsically challenging so it is perhaps not surprising that less than perfect fidelity is achieved. One of the challenges is that automasking tended to work better if each major cycle goes deeper, avoiding premature termination of mask updates. Deeper major cycles are also, however, a source of systematic error in imaging since errors in the approximate PSF used in the minor cycles are corrected at each major cycle. Better results might be obtained in the future by varying the `minpercentchange` sub-parameter. Alternative deconvolution algorithms, such as SDINT (Rau et al. 2019) or TP2VIS (Koda et al. 2019), may also prove more stable and would be worth evaluating on an equal footing. One surprising result of this investigation was that the JvM corrections, while modest, did not consistently improve image fidelity. It would be worth exploring this in more detail, and in particular, if a different form of the correction is needed in the context of an image which also includes total power data, e.g., by feather. Finally, it's worth noting that while this is a continuum only simulation— important elements of the scripts probably won't work for cubes— it is in practical terms most indicative of one channel of a spectral line use case in that there is no guarantee that ngVLA TP will provide robust and sensitive continuum information, similar to the current situation with ALMA.

Acknowledgements: Thanks to Urvashi Rao and Kumar Golap for extensive assistance in getting ngVLA heterogeneous array imaging working; Amanda Kepley for helpful automasking advice; and Ryan Loomis for sharing his code to calculate the JvM correction.

References:

Carilli, C. et al. 2020, ngVLA memo 82: "Reference Design RevC.01 Description and Alterations"

Indebetouw, R. et al. 2013 ApJ 774, 73

Jorsater, S. & VanMoorsel, G. 1995, AJ 110 20137

Kepley, A. et al., "Auto-multithresh: A General Purpose Automasking Algorithm", PASP v132, 1008

Koda, J. et al. 2019, PASP v131, 999 p054505

Kundert, K. et al. 2017, IEEE Trans. Ant. & Prop. 65, v2 "Understanding Systematic Errors Through Modeling of ALMA Primary Beams"

Mason, B. et al. 2018, ngVLA memo 43: "The ngVLA Short Baseline Array"

ngVLA Science Advisory Council 2017, ngVLA memo 19: "Key Science Goals for the Next Generation Very Large Array (ngVLA): Report from the ngVLA Science Advisory Council"

Rau, U., Nikhil, N. & Braun, T. 2019, AJ 158, 3
Rosero, V. et al. 2019, ngVLA memo 65: "Sculpting of the Synthesized Beam and Image Fidelity Study of KSG 1: Imaging of Protoplanetary Disks"
Rosero, V. et al. 2020, ngVLA memo 85: "Comparison of Alternative Configurations for the ngVLA Plains Subarray"
Selina, Murphy & Erickson 2017, ngVLA memo 18: "Summary of the Science Use Case Analysis"
Selina, R. et al. 2018a, Proc. SPIE 10700E, 10S
Selina, R. et al. 2018b, ASPC 517, 15S
Tsutsumi, T. et al. 2004, ALMA memo 488: "Wide-Field Imaging of ALMA With the Atacama Compact Array: Imaging Simulations"
Walter, F. & Brinks, E. 1999, AJ 118 273

A Appendix: Heterogeneous Array Procedure in CASA

Special procedures are needed to get the correct primary beams in CASA for measurement sets that contain visibility data from *heterogeneous arrays* (i.e. where all antennas do not have the same illumination pattern or size). This section explains the procedures which are needed, and gives some information relevant to capabilities in recent CASA versions.

Several general considerations are useful to note. The first is that CASA's information about primary beams is managed by a software component known as the *voltage pattern manager*, or vp manager. The second is that voltage patterns are indexed or "looked up" mainly by *telescope name*. The third is that CASA uses the telescope name to track both geographical information (latitude, longitude) and antenna-specific characteristics like the voltage patterns. While some voltage patterns are indexed by telescope name, some have more sophisticated indexing available; in particular, explicit 2D illumination maps can be provided which are indexable by antenna name substring matching. For other voltage patterns, problems and ambiguities can arise for an observatory which has different types of antennas. CASA solves this problem for ALMA by hard-coding the cases at a low level. At present other observatories, like NGVLA, require specific procedures to achieve fully correct heterogeneous array imaging.

As described in NGVLA memo 43, SBA images can be made by overwriting the existing, default NGVLA entry in the VP manager with one suitable for a 6m antenna (instead of the default 18m). If you have an MS which contains simulated NGVLA (or other non-ALMA) visibility data from antennas with multiple diameters ("heterogeneous array" data), this will not work. The *recommended method* is to change the TELESCOPE_NAME field(s) in the OBSERVATION table of the simulated MS to something that does not correspond to an observatory (telescope) already in CASA's database (e.g., NGVLA1); and then to define VP entries with voltage patterns corresponding to the appropriate antenna diameters. These entries should also use the same telescope/observatory name, in this example, NGVLA1. This procedure is illustrated below; it is known to work in CASA 5.6 and 6.1, and is expected to work in future CASA versions. It does *not* work in CASA 5.4.1, and likely

does not work in other versions over the preceding several years.[9]

For CASA 5.7/6.2, another fallback method is expected to be available, which is to set the telescope name in the MS to something not known to CASA and leave the VP's at their defaults– the so-called "unknown observatory trick". In this case CASA will assume a uniform illumination pattern (Airy pattern primary beam) and get the antenna diameter(s) from the MS on a per-visibility basis. Tests on `CAS-13010` indicate that this works as expected for heterogeneous array data.

```
# Take as a given the existence of an MS containing visibilities
#  from a heterogeneous array ("HA")
ha_ms = 'myvisibilities.ms'

# Change TELESCOPE_NAME to a value CASA does not already have
#  in its observatory database (here we choose NGVLA1):
tb.open(ha_ms+'.ms/OBSERVATION',nomodify=False)
telcol=tb.getcol("TELESCOPE_NAME")
new_telcol = np.array(["NGVLA1" for i in range(len(telcol))],dtype=telcol.dtype)
tb.putcol("TELESCOPE_NAME",new_telcol)
tb.flush()
tb.done()

# Define voltage pattern entries which have the desired information
#  the telescope name should match what is in the MS
myvp='ngvlaCustom.tab'
vp.reset()
try:
    rmtables(myvp)
except:
    print("Table wasnt there alls good")
vp.setpbairy(telescope='NGVLA1',dishdiam=6.0,blockagediam=0.0,maxrad='3.5deg',
    reffreq='1.0GHz',dopb=True)
vp.setpbairy(telescope='NGVLA1',dishdiam=18.0,blockagediam=0.0,maxrad='1.78deg',
    reffreq='1.0GHz',dopb=True)
vp.saveastable(myvp)
# provide 'ngvlaCustom.tab' to imaging tasks such as TCLEAN
```

[9]At present there is not a known and documented procedure to correctly image heterogeneous array interferometric data *other than ALMA* in CASA 5.4 and recent, previous versions using standard tasks such as CLEAN or TCLEAN. The issues involved relate to the primary beam so for some use cases may not be important.