Direct Optimal Mapping Memo

Kyle Miller

Pober Group, Department of Physics, Brown University

ABSTRACT

Context. To help quantify the numerical and statistical properties of a novel imaging process to be used specifically in the search for the highly redshifted ($z > 5$) Epoch of Reionization (EoR) 21 cm neutral hydrogen signal.
Aims. This new imaging process—Direct Optimal Mapping (DOM)—requires more computational resources than typical; so, showing how the results of this imaging process scales with resolution and stacks up against the competition is the ultimate goal.
Methods. DOM is compared and contrasted with the Common Astronomy Software Applications (CASA) imaging package using visibilities generated with `pyuvsim` and an "expectation" image generated to use as a proxy for "ground truth".
Results. The computational headroom of DOM is clear and memory requirements grow quickly as a function of field of view. DOM provides access to a full covariance matrix and has roughly the same image fidelity as CASA, further studies will have to show how this can be harnessed to extract the EoR signal from high dynamic range maps with bright foregrounds. A slightly trivial example also shows the potential for future algorithms to potentially achieve a linear deconvolution pipeline.

Key words. radio astronomy – epoch of reionization – synthesis imaging – precision cosmology

1. Introduction

In the search for the Epoch of Reionization (EoR) signal using radio telescopes there are a number of large obstacles; see Morales (2010) and Zaroubi (2012). There is local Radio Frequency Interference (RFI), ionospheric phase fluctuations limiting observations to above 10 MHz and baselines to less than 10 km, and extremely bright galactic and extra-galactic foregrounds. The dynamic range of the sky brightness distribution (∼300K-10mK) requires a precise measurement of the radio interferometric visibility function to better than one part in 10^4. The HERA collaboration is using highly redundant baselines to improve self-calibration techniques, extensive RFI mitigation, and a radio-quiet location to accomplish this DeBoer (2017). Further improvement of the data processing pipeline naturally follows with a *direct* and *optimal* mapping process that has well-defined statistical properties (i.e., a full pixel covariance matrix); see Dillon (2015) and Xu (2022).[1].

Typically, for radio interferometers, the visibilities generated by correlation of various antennae pairs are placed on a grid amendable to a 2D Fast Fourier Transform (FFT). The process of gridding can often be more computationally intensive than the optimized FFT, which reduces an order n^2 computational problem to $n \log(n)$. So-called "optimal" mapping methods have been implemented before in precision cosmology applications, where Tegmark (1997) showed they can be used to generate lossless images with minimum errors by keeping track of error propagation with a restricted set of traceable operations; *optimal* is basically a stand-in for *linear*. Direct Optimal Mapping (DOM) builds on this framework and extends it to wide-field radio interferometric imaging. The final result is a spherical (healpix) map with a corresponding pixel covariance matrix, which quantifies the correlation between all pixels in the map; while the process of precision map making is computationally intensive it can also be used for data compression.

[1] github.com/HERA-Team/direct_optimal_mapping

2. Direct Optimal Mapping, simulated visibilities, and an expectation image

The field of precision cosmology has requires an enhanced approach to imaging. In order to produce reliable statistical measures of the large spatial structure contained in maps (instead of modeling and fitting specific localized objects) an understanding of the imaging process and its effect on the output map is important. This was first spearheaded by the optimal imaging methods built for studying the Cosmic Microwave Background (CMB) and is being extended into 21 cm cosmology. Most projects working in this field are utilizing aperture synthesis radio telescopes, which adds a considerable amount of complexity to the imaging process. Many of the industry standard imaging techniques for these interferometers rely on simplifying assumptions to speed up the process—for precision cosmology these types of trade-offs are not acceptable.

While the other imaging techniques can produce accurate maps they do not preserve the inputs. This is especially true in non-linear deconvolution algorithms. DOM, in theory, not only produces an accurate map but also a precise mapping from the inputs (i.e., samples of the visibility of the sky brightness distribution) to the map without squeezing and streching the inputs to fit a grid; *direct* is basically a stand-in for *ungridded*. While this technique is more computationally expensive $\leq O(n^4)$ the immediate benefits are three fold: 1) the imaging process effective compresses the data; 2) a w-term approximation is not needed for any non-coplanarity; and 3) the maps can be decomposed using sufficiently well-established models of foregrounds and any sources of noise.[2] The other upshot of this technique would be a linear deconvolution algorithm, which could be useful in other areas of radio astronomy.

[2] It is hoped this computational complexity can be eventually reduced to $\sim n \log(n)$; see Dillon (2015). The run times for this work were considerably longer than CASA. Calculating the matrix $\mathbf{P}$ (see Equation 3) was especially time consuming.

A brief review of DOM hinges upon the equation (and its assumption of a linear relation between time series visibility data and the resulting map):

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{1}$$

where the instrument $\mathbf{A}$ operates on the (true) sky brightness distribution $\mathbf{x}$ to produce the measurements of the visibility function $\mathbf{y}$ with any other noise $\mathbf{n}$. While seemingly innocuous the matrix $\mathbf{A}$ grows with observation time, bandwidth, and baselines; it has the shape $m \times n$ with m being the number of visibility samples and n the discretization of the sky in terms of pixels and frequency. The imaging process then is essentially solving for $\mathbf{x}$, however, while noise can usually be subtracted (using techniques like frequency switching) the matrix $\mathbf{A}$ is most often not invertible—this would require a square matrix, which would necessitate a large oversampling of the visibility function to match a discretization of the sky commensurate with a given telescope's resolution.[3] So, instead of solving for the solution, a map estimator is calculated from the following equation:

$$\hat{\mathbf{x}} = \mathbf{D}\mathbf{A}^{\dagger}\mathbf{N}^{-1}\mathbf{y}, \tag{2}$$

where $\mathbf{N}^{-1}$ is the inverse noise matrix and $\mathbf{D}$ can be any invertible matrix used to normalize.[4] Expanding the right side and averaging over time (to remove the noise term in $\mathbf{y}$) produces a matrix of point spread functions $\mathbf{P}$, e.g.,

$$\langle\hat{\mathbf{x}}\rangle = \mathbf{D}\mathbf{A}^{\dagger}\mathbf{N}^{-1}\mathbf{A}\mathbf{x} = \mathbf{P}\mathbf{x}, \tag{3}$$

that precisely maps pixels from the true sky to the estimator; and, if it could be inverted, this would allow a linear deconvolution, i.e., solving for $\mathbf{x}$.

In order to test and validate DOM it makes sense to use simulated visibilities both with and without noise. This allows for a quantitative comparison between standard imaging techniques and an exploration of how DOM could allow for extraction of foregrounds and noise from the output maps. A catalog of $^1/_2$ Jy point sources that spells out HERA in a patch of sky roughly $\sim 20^\circ \times 10^\circ$ was used to generate visibility data with `pyuvsim`.[5] A high-precision beam file based on the HERA dipole feed was used in conjunction with an array with similar *uv*-plane coverage to the Murchison wide-field Array (MWA; specifically an 89 antennae version without the core cluster). The visibility data were converted to a measurement set to image with CASA and run through the DOM code.

The *uvw*-plane coverage is shown in Figure 1 where the array is visibly shown to be on an inclined plane in the E-W direction. The $N = 89$ elements produced $N(N-1)/2 = 3916$ baselines. These point sources should be expected to be unresolved and this was confirmed by plotting the visibility amplitude against baseline length in Figure 2. A Savitzky–Golay filter of order 3 was applied with two window lengths of 101 and 301, which recovers the zero spacing amplitude and shows a flat line confirming unresolved sources. These noiseless visibility amplitudes could then be perturbed with noise; see Section 7.

[3] This could possibly be realized using multi-element feeds.

[4] This is taken to be $\sim 1/beam^2$ in this memo, although there is an ideal alternative which explained in further detail by Dillon (2015)

[5] The simulation was run from an exact text-based point source catalog and a healpix map with "smeared" point sources. All images presented are based on the latter, although errors are provided for both. github.com/RadioAstronomySoftwareGroup/pyuvsim

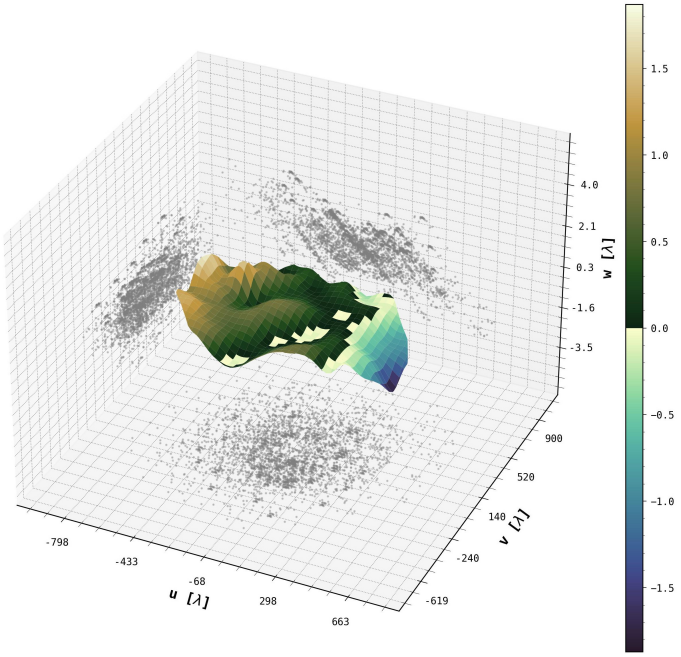

Fig. 1. *uvw* coverage in units of wavelength $\lambda \approx 3$ meters for one time. Contour plot shows an interpolated representation of the array topology.

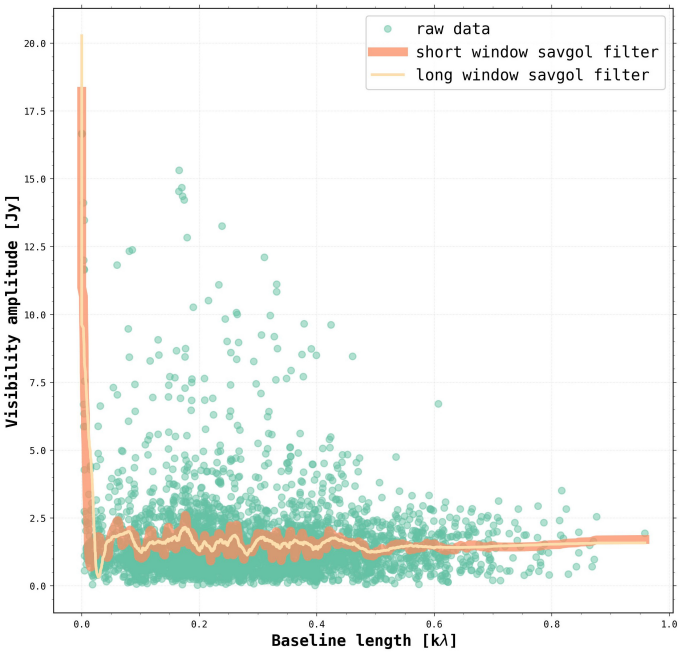

Fig. 2. Baseline length plotted against visibility amplitude.

Usually, in radio astronomy, the exact sky brightness distribution is not known. There is no "ground truth" to compare the output of a given imaging process with. For the purposes of testing DOM the simulated visibilities can be circumvented with a direct point source image. This map with zero flux everywhere besides the point source locations could then be multiplied by the matrix $\mathbf{P}$ above to produce what would be expected from DOM; this would be the sky brightness distribution "convolved" with the instrument response, i.e., the dirty beam, which is like an optical point spread function. Any differences should be due to the visibilities relative to the exact point sources in (α, δ), which had to be "smeared" onto the healpix map.

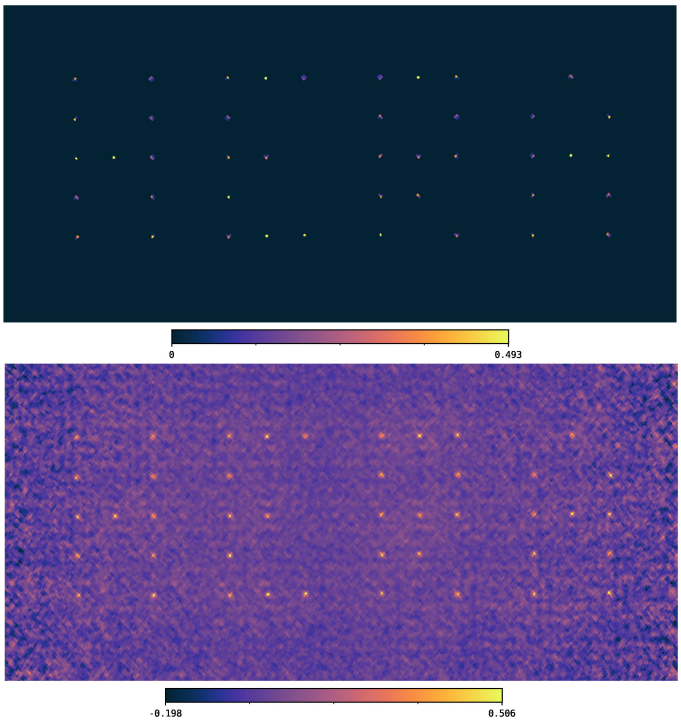

Fig. 3. Point sources (top). Point sources dotted with **P** matrix (bottom).

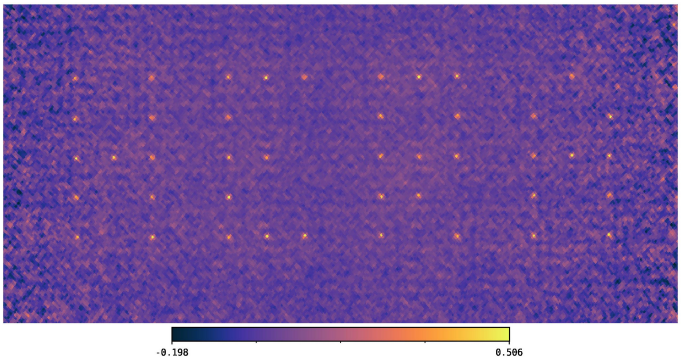

Fig. 4. DOM output (`nside`=1024).

Having such an "expectation image" should allow for a quantification of *image fidelity*. A measure of how accurately the imaging process is able to match the correct output. The expectation image before and after the matrix multiplication is shown in Figure 3. The output of DOM is shown in Figure 4. See section 4 for the full quantitative comparisons.

3. CASA and Healpix maps

While DOM has already been designed for wide-field imaging and specifically making healpix maps Gorski (2005), CASA generates images with a 2D projection. Converting these 2D images into healpix maps for comparison with DOM was accomplished by loading FITs files from CASA into a python-based framework where `astropy`, `reproject`, and `healpy` were used to recreate the images on a blank healpix canvas. There were four versions of this process that were tested using the settings ['bicubic', 'biquadratic', 'bilinear', and 'nearest-neighbor']. The method 'bicubic' yielded the lowest error; see Table 2.

The CASA images used were higher of a resolution than the more restricted healpix format because the error scaled roughly like $\sim$ *cell*/*beam* for *cell* $<$ *beam* but not *cell* $\ll$ *beam*. The discrete steps of `nside` values for healpix maps result in pixel resolutions listed in Table 1. When imaging in CASA with the task `tclean` the parameter "cell" sets the angular resolution of the image (a cell size of 1' was used for all CASA images because it provided the lowest error). For the array described in Section 2 the resolution can be estimated to be $\sim \lambda/B$ where B is the longest baseline. Using the values of 1km and 3m results in an angular resolution of roughly 10′.

Table 1. Healpix map resolutions

Healpix nside value	*resolution*
512	6.870′
1024	3.435′
2048	1.718′

To fully Nyquist sample the spatial frequency distribution then would require a pixel output resolution twice as small. The CASA documentation even suggests oversampling with 5 pixels per beam, so a healpix map of `nside` 2048 might be the minimum resolution for capturing the full instrumental response to the sky brightness distribution ($\sim 10'/5$). Higher resolution maps were produced for both CASA and DOM for testing purposes, however, calculating the **P** matrix became extremely resource intensive with respect to RAM, requiring over 700GB for an `nside` of 2048. Interestingly, the quantitative measures did respond to these increases in resolution. This will be important for higher resolution work; see Section 4.

4. Dynamic range, SNR, and image fidelity

In addition to the standard statistical measures—Root Mean Square (RMS), max, and min—a quantity related to the *dynamic range*—the Signal-to-Noise Ratio (SNR)—and image fidelity have been calculated. Dynamic range is generally determined by the subject in photography; for example, a photograph with maximum dynamic range would include the sun and shadowy areas like a cave. The dynamic range of such an image could be over 10^7. Most generally, the dynamic range is the difference between the maximum and minimum value in an image. This would apply to the sky brightness distribution, however, the dynamic range of a radio telescope and the maps it produces is not quite the same. The noise floor of the instrument is almost always above the minimum value, so the dynamic range is ideally thermally limited in practice.

The sky brightness distributions that will be observed with 21 cm precision cosmology instruments will contain large intrinsic dynamic ranges with the target signal at the low end of the range. Extremely bright Radio Frequency Interference (RFI) can cause confusion even in high polar angle sidelobes with $T_B > 10^9$ K. Galactic and extra-galactic foregrounds generally have a rising flux density at low frequency from synchrotron emission with brightness temperatures around $T_B \sim 10^2$ K. Observations will take place at night to avoid the sun $T_B > 10^5$ K. Other sources like the A-team will have to be avoided to permit sound observations of the redshifted 21 cm signal with $T_B \approx 10$ mK. These type of sources can push the dynamic range beyond the limits of what instruments can handle: $\sim 10^5$.

For this memo the testing treats dynamic range and the SNR as being effectively the same thing. The maps presented without noise would have the SNR being equal to the dynamic range.

Table 2. Map statistics

Name	Max	*Min*	*RMS*	SNR	*RE*
DOM 512	0.33	−0.18	0.03	1.02	76.55%
CASA 512	0.35	−0.19	0.05	0.89	68.07%
DOM 1024	0.51	−0.20	0.03	1.21	$3 \cdot 10^{-4}$%
CASA 1024	0.51	−0.20	0.04	1.10	4.250%

Any images with noise where it can be measured accurately might have a different value for the SNR, however, they will both be considered as difference between the maximum value and σ_{RMS} in a blank patch of the map:

$$SNR \equiv \log_{10}(max\_value) - \log_{10}(\sigma_{RMS}). \quad (4)$$

A more in depth discussion of this topic can be found in Burke (2019). All of the images presented here will be a linear stretch of the values unless otherwise noted.

Image fidelity is not quantifiable for most radio astronomy applications—only when there is an adequate model of the sky brightness distribution can this be achieved. Given the images produced in this memo are from a simulation it was possible to construct the expectation map (a proxy for "ground truth") presented in Section 2. For scalar values a relative error can be calculated when the exact solution is known. For maps this notation has to be generalized by the following mathematics:

$$RE \equiv \frac{\|M - E\|_F}{\|E\|_F}, \ \|E\|_F \equiv \left(\sum_{i,j} |e_{i,j}|^2 \right)^{1/2} \quad (5)$$

where M is a map, E is the expectation, and the equation on the right is the definition of the Frobenius norm, which has been implemented in the `numpy` package.[6] The resulting quantity will be called the Relative Error (RE).

5. Histograms and difference images

At the resolution corresponding to `nside`=1024 the CASA, DOM, and expectation maps can be plotted together in a histogram of their intensity distributions. This is shown in Figure 5. The DOM and expectation points overlap to a large extent and the CASA histogram notably shows more high-flux counts. It is not clear what exactly causes the differences. The issue of wide-field imaging and the w-term will be addressed in Section 6.

The expectation histogram was fit with a Gaussian distribution to quantify the variance. The result is provided in the legend. In this area of the histogram, around zero, there should mostly be noise; however, these maps do not contain any noise, only the effects of the instrument, i.e., the dirty beam. Negative values are expected with any interferometer that has not sampled the zero spacing. Modeling the negative basin could provide an extra check for the minimum values and allow for corrections.

[6] numpy.org/doc/stable/reference/generated/numpy.linalg.norm.html

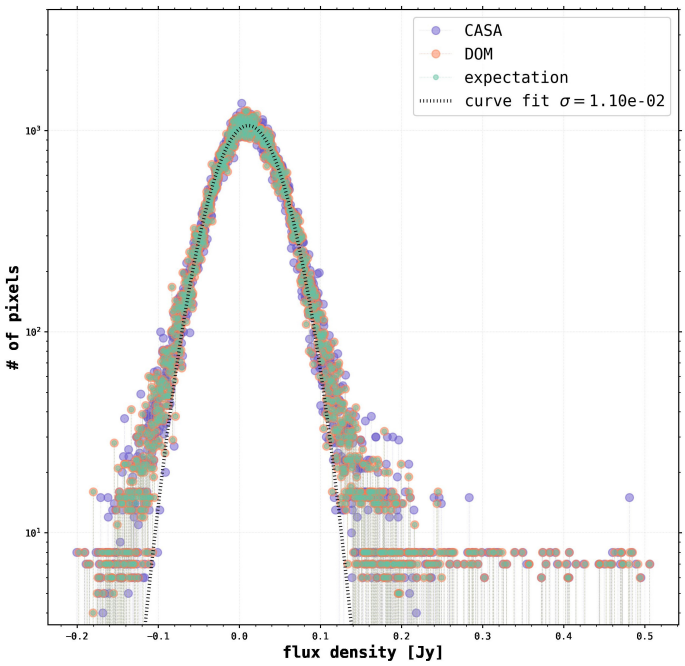

Fig. 5. Histograms of DOM, CASA, and expectation map; `bins`=2048.

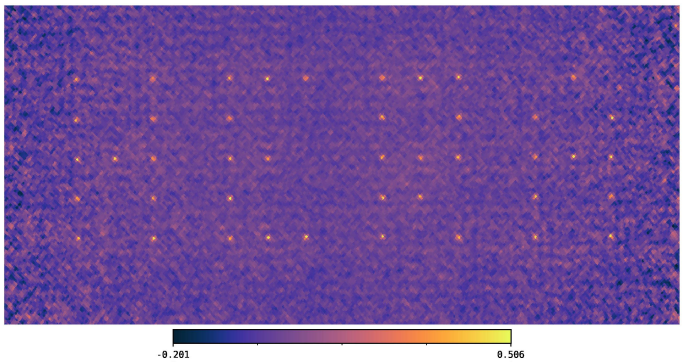

Fig. 6. CASA healpix map ('bicubic').

Additionally, three difference maps were made between DOM and the expectation, CASA and the expectation, and DOM and CASA. The results with CASA are open to errors from the healpix transfer process, although this was largely mitigated by basing the simulation on a healpix map. Qualitatively, there are little differences. The values in Table 2 show the RE values for DOM and CASA; larger errors for text-based simulated visibilities are omitted (~ 40% for both CASA and DOM).

The three difference maps containing residuals are found in Figure 7. The most interesting feature is the non-linear looking transformation that appears in the middle, which is patterned for bicubic and has curving arcs for bilinear. The normalization scheme used in the matrix **D** appears to some degree in all three images. While this scheme improves the images in the widefield zone—it unfortunately up-weights small variations on the periphery. Note that this normalization was factored into DOM, CASA, and the expectation map—although at different stages; nevertheless, because the pattern on the edges is in all the difference images, it lends support to the notion that the normalization is magnifying a combination of wide-field, interpolation regarding coordinate systems, and dirty beam-related issues.

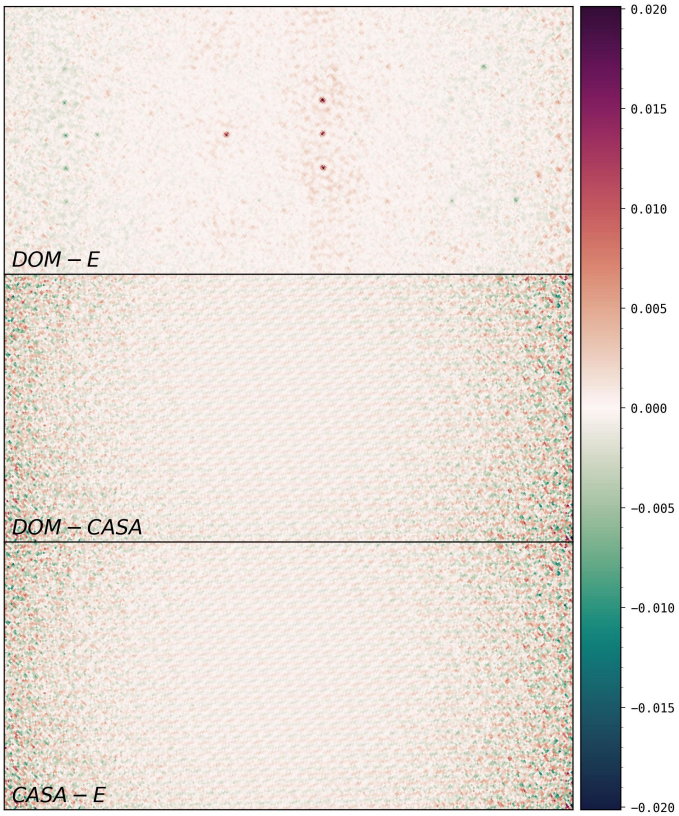

Fig. 7. Difference maps; see annotations (E is expectation, Figure 3). The colorbar amplitude for the top panel is about 4 dex less.

6. Wide-field imaging and the *w*-term

For small fields of view—within the synthesized beam of a radio interferometer—it is easier to jump from the full 3D visibility relation between the sky and the *uvw*-plane Thompson (2017)

$$\mathcal{V}(u,v,w) = \frac{1}{\Omega}\iint \mathcal{A}(l,m)\mathcal{I}(l,m)e^{-j2\pi[ul+vm+w(\Omega-1)]}\mathrm{d}l\mathrm{d}m, \qquad \Omega = \sqrt{1-l^2-m^2} \tag{6}$$

to a 2D version with only *uv* components. This allows for a FFT inversion to produce a dirty image, which can then be followed by a deconvolution method. The full field-of-view offered by an interferometer is determined by the size of an individual antennae: it is the resolution of what is called the primary beam. To produce dirty images of this wider field of view necessitates extra work. 21 cm precision cosmology is looking for a wide-field diffuse target signal—so it going to be required.

One of the benefits of DOM is an automatic wide-field imaging compatibility. The current algorithm of choice for wide-field imaging is W-projection. This is detailed along with the following criterion used to determine if this type of visibility treatment is needed in Cornwall (2008):

$$N_F = \frac{D^2}{B\lambda} \tag{7}$$

where D is the antennae size, B is the longest baseline, λ is the observation wavelength, and the result N_F is known as the *Fresnel Number*. When the Fresnel Number is less than 1 there will be (non-coplanar) effects that need to be mitigated with an appropriate wide-field inversion method. DOM solves this problem by performing a *direct* linear inversion; this would be computationally prohibitive for large fields of view, accordingly DOM is intended for use in a faceted observation program Dillon (2015)—making it akin to a hybrid of full 3D inversion and the older (pre-W-projection) technique of faceting.[7]

It is possible small differences between the images, e.g., see Figures 5 and 7, are due to issues related to the *w*-term and wide-field imaging. A quick calculation of the Fresnel number for the HERA array with $D \approx 14$ m and $B \sim 1$ km yields 0.07, which is indicative of potential problems. Using the `gridder` option set to W-projection in CASA generated worse results, in terms of RE, by a factor of ~ 2. An extremely non-coplanar array—with *w*-term magnitude on the order of the *uv*-plane—was also tested: this required W-projection in CASA with `wprojplanes` > 100 to recover a map similar to the DOM result, i.e., containing all point sources. A full exploration of this nuanced topic is beyond the scope of this memo.

7. Towards advanced Fourier filtering

The main advantage of DOM is not supposed to be superior image fidelity nor dynamic range, although they would be welcome features. To fully make use of this new mapping method requires using the covariance matrix it generates to do something novel CASA is unable to do. A simple example would be adding Gaussian noise into the simulated visibilities, carrying out the DOM procedure, and then untangling the noise from the result using the covariance matrix to guide a precision Fourier filtering algorithm. Reckoning by the radiometer equation, adding a Gaussian distribution with $\Delta\mathcal{V} \sim 1$ Jy is equivalent to a reasonable system temperature of $T_{sys} \sim 1000$ K

$$\Delta\mathcal{V} \approx \frac{2k_B T_{sys}}{A_{eff}\sqrt{\delta\nu\tau}} \tag{8}$$

with an integration time of $\tau = 11$ seconds, bandwidth $\delta\nu = 80$ KHz, and an effective aperture $A_{eff} \approx \epsilon N\pi(D/2)^2 \approx 7000\,\mathrm{m}^2$ assuming an aperture efficacy $\epsilon = {}^1\!/_2$ Tillman (2016). A schematic pipeline, for this type of Fourier filtering, is given in Figure 8, where the power spectrum is shown from before and after to illustrate how an algorithm like `skimage` non-local means might be augmented via a full covariance matrix.

8. Linear deconvolution

Normally, after forming a dirty image like those in Figures 4 and 6, these intermediary images are then further refined by performing a deconvolution to remove the effects of the instrument (i.e., $\mathcal{A}$ in Equation 6, which is the dirty beam). Two of the most prevalent methods used are CLEAN and Maximum Entropy, although there are many others. These methods are both non-linear and have a mathematical motivation with varying levels of confidence. Ideally, a linear deconvolution would be performed whereby the effects of the telescope beam would be removed exactly instead of by workarounds. The current methods will probably not be retired because even if a linear deconvolution algorithm becomes feasible it does not mean it will be worth the computational costs for all situations.

Within the framework of DOM a linear deconvolution can be carried out relatively easily when $\mathbf{D} = \mathbf{I}$; see Equation 3. To show

[7] These facets will be approximately the size of the maps presented in this memo: $\sim 10^\circ$ on a side.

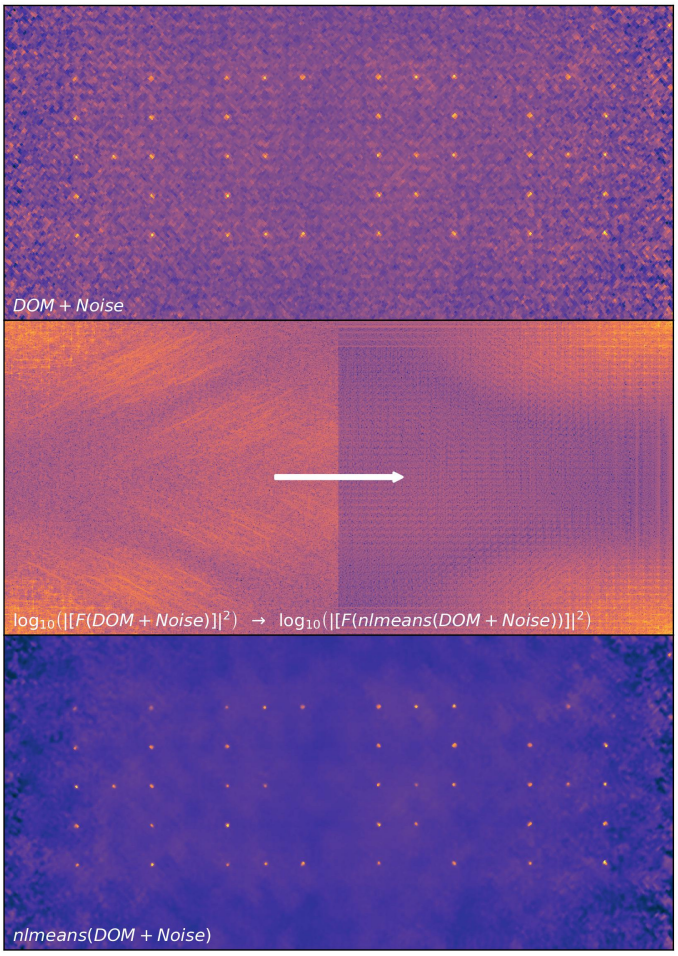

Fig. 8. Fourier filter schematic for Figure 4 with additional Gaussian noise added. See annotations; F represents a 2D Fourier transform. (The arrow shows how the original power spectrum is modified—*this is a stand-in intermediary for the result from* `skimage` *non-local means, which is similar to a full covariance matrix filter.*)

this as a proof of concept the expectation image in Figure 3 was deconvolved linearly. Once the matrix $\mathbf{P}$ had been obtained the inverse of this matrix could be solved for using `numpy` allowing $\mathbf{x}$ to be solved for; see Equation 3.[8] The results of this somewhat trivial exercise are show in in Figure 9; for comparison, a non-linear deconvolution is shown in Figure 10 where the CASA map was CLEANed. Applying this method to the actual dirty images (to remove dirty beams) is left for future work.

9. Conclusions

While the imaging process in radio astronomy can often be fraught with non-linear algorithms and various approximations, making it seem more like an art than a science, this memo has sought to compare and contrast a novel technique which offers unparalleled access to a linear map making paradigm. For 21 cm precision cosmology—where the goal is likely to be a statistical measure with well defined error bars, instead of a visually appealing image—a linear algorithm, for optimal map making, is an important tool.[9] Knowing how DOM stacks up against the

[8] The output of this operation did not recover the identity $\mathbf{P}^{-1}\mathbf{P} = \mathbf{I}$, which could be due to numerical precision-related errors.

[9] Most instruments working in this field are not designed to produce aesthetic images; this can be quantified by the so-called "photogenic resolution" Christensen (2014), which for radio interferometers and HERA respectively is $r_p = B/D \approx 10^2$ (it should ideally be over 10^3).

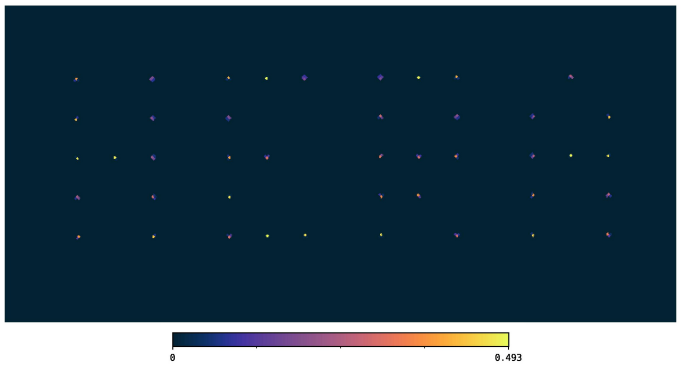

Fig. 9. Linear deconvolution of Figure 3.

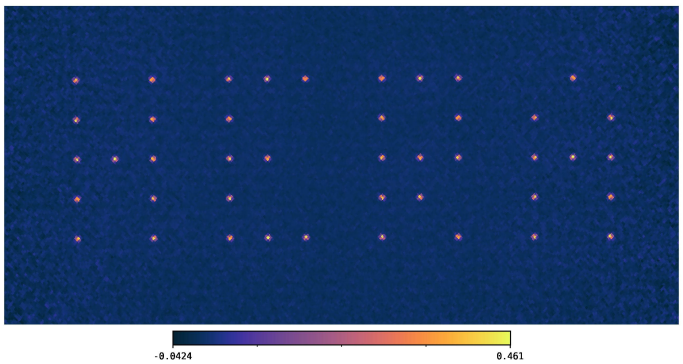

Fig. 10. Non-linear deconvolution of Figure 6; `niter`=1000.

competition in radio astronomy is crucial because, in particular, radio interferometry has many idiosyncrasies, which could potentially generate non-linear dependencies between the time series visibilities and the output map; both the ionosphere, at low frequency, and the atmosphere are good examples. Only the surface has been scratched here. The promise of DOM is clear and there are other applications like linear deconvolution where DOM might contribute to the larger radio astronomy community along with more potential uses that have not been considered.

Acknowledgements. Warm regards to all of the many open source software communities related to the work herein—including, but not limited to: `astropy`, `numpy`, `CASA/NRAO`, `python`, `matplotlib`, `slurm`, `scipy`, `skimage`, `jupyter`, `healpy`, `reproject`, `cmocean`, and `linux`. Most importantly, I have to acknowledge the insights and support from Prof. Pober, Jacob Burba, and Zhilei Xu. *Funding for this project came from NSF REU*. Thank you!

References

Morales, M. & Wyithe, S. 2010, Ann. R. Astro., 48:127-71.
Zaroubi, S. 2012, https://arxiv.org/abs/1206.0267
Dillon, J. *et al.* 2015, Phys. Rev. D, 91(2), 023002(26).
Xu, Z. *et al.* 2022, https://arxiv.org/abs/2204.06021
DeBoer, D. *et al.* 2017, Pub. of the Astro. Soc. of the Pac., 129:045001.
Tegmark, M. 1997, The Astro. Jour. 480 : L87–L90.
Gorski, K. *et al.* 2005, The Astro. Jour. 622 : 759–771.
Burke, B. *et al.* 2019, Cambridge University Press, (New York, New York).
Thompson, R. *et al.* 2017, 3rd ed., Springer, (Cham, Switzerland).
Cornwall, T. *et al.* 2008, https://arxiv.org/abs/0807.4161
Tillman, R. *et al.* 2016, Astronom. Instrum., 5, 1650004.
Christensen, L. *et al.* 2014, Comm. Astro. with the Pub. Jo., 14, 20.