P9 Distributed Image Reconstruction for the new Radio Interferometers

Jonas Schwammberger

July 15, 2019

Abstract

Contents

1 Introduction

1.1 Radio interferometry system

This project is focused on distributing image Reconstruction for radio interferometers, which is only one of three steps in the pipeline from measurements to the final image. We give a quick overview over the whole pipeline in figure 1 and how Radio Interferometers work in principle: The antennas observe the arriving electromagnetic wave, gets processed in three steps, correlation, calibration and image reconstruction.

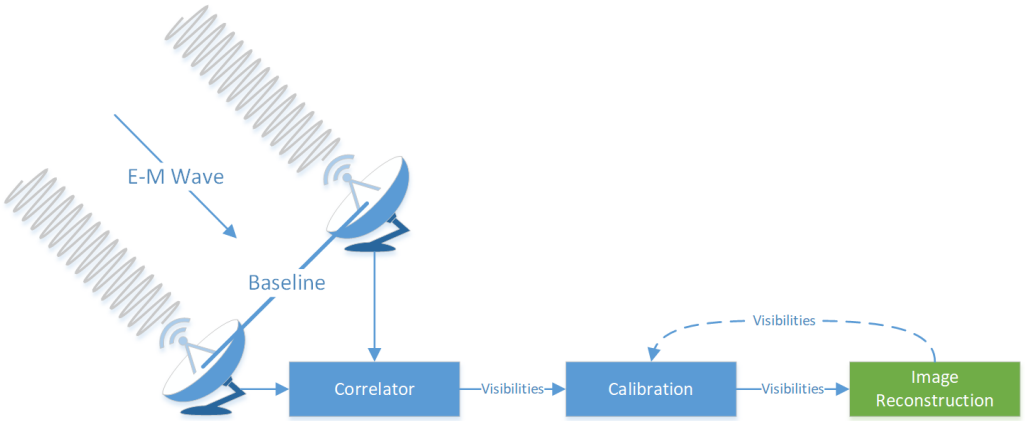


Figure 1: Radio interferometer system

First, the electromagnetic wave gets measured by the different antennas of the interferometer. The measurements of each antenna pair get correlated into a Fourier component (called Visibility in Radio Astronomy). Each antenna pair measures a complex-valued, noisy visibility of the sky. The distance and orientation of the antenna pair relative to the incoming signal, called the baseline, dictates which visibility gets measured. The longer the baseline the higher-order visibility gets measured, resulting in a higher angular resolution. The image 2a shows the sampled visibilities in the Fourier space of the MeerKAT interferometer. Every dot is a single measurement. After correlation, the visibility data is saved for later processing.

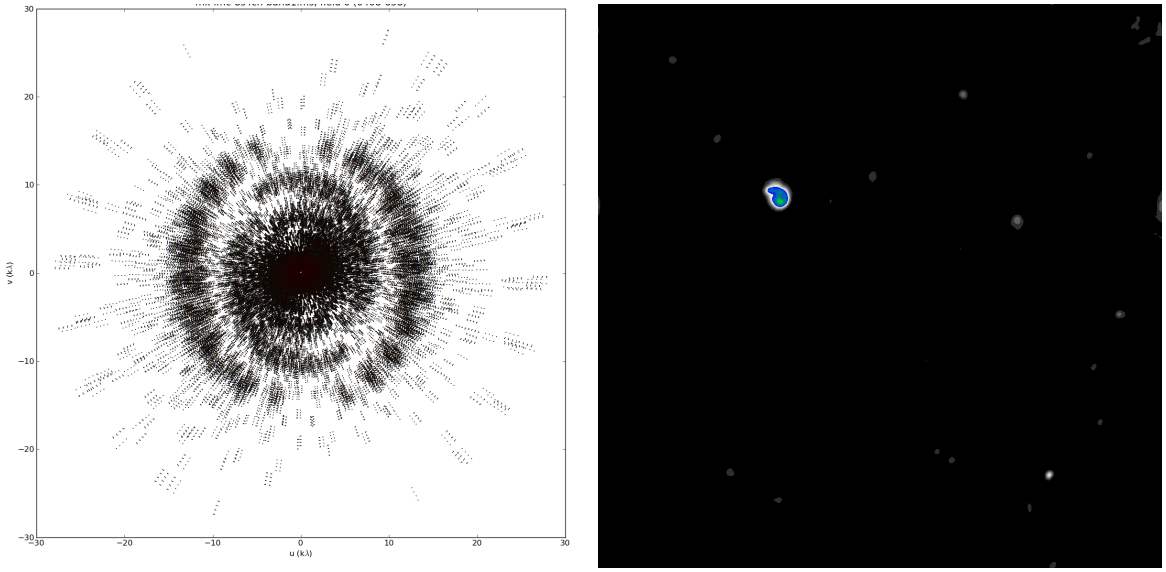
The calibration step is done after all visibility data has been recorded. This step corrects the amplitude and phase of the measurements for varying antenna sensitivities, pointing errors and other effects. Also, this step removes corrupted data from the measurements. After the calibration step, the visibilities still contain noise.

The last step is responsible for reconstructing an image from the calibrated, noisy visibilities. The figure 2 shows a real-world example of a reconstruction from the MeerKAT radio interferometer. It arrives at the reconstructed image by inverting the measurement equation, and deciding which part of the visibilities is noisy, and which part is the signal. These two problems, handling the noise and inverting the measurement equation, are central to image reconstruction for radio interferometers. They influence both the quality and the runtime costs of the reconstruction.

1.1.1 The measurement equation

The measurement equation models how the electromagnetic wave gets distorted on its path from the celestial source through the ionosphere and finally through the antenna of the interferometer[1]. It abstracts all effects we wish to correct in the image reconstruction in one equation. As such, there is no single unified measurement equation for all interferometers, and generally depends on the instrument.

In this project we use the measurement equation (1.1). It consists of three parts: The visibility measurements



(a) Measurements in the Fourier space. *(b) Reconstructed image.*

Figure 2: Example of an image reconstruction for visibility measurements of the MeerKAT radio interferometer

$V(u, v, w)$[1], the observed image with a normalization factor $\frac{I(x,y)}{c(x,y)}$ and the Fourier Transform $e^{2\pi i[...]}$. u v and w represent the axes in the Fourier Domain, while the x and y axes represent the angles away from the image center. A pixel in $I(x, y)$ represents how much radio emission was measured from a the direction x, y. The image 2a shows an example for $V()$, while 2b shows an example for $I()$.

$$V(u, v, w) = \int\int \frac{I(x, y)}{c(x, y)} e^{2\pi i[ux+vy+w(c(x,y)-1)]}\, dx\, dy\,, \quad c(x, y) = \sqrt{1 - x^2 - y^2} \tag{1.1}$$

The radio interferometer essentially observes the sky in the Fourier domain. If we want to retrieve the observed sky image $I()$, all we need to do in theory is calculate the inverse Fourier Transform. Note however that the visibilities $V(u, v, w)$ are three dimensional, while the image $I(x, y)$ only has two. Also note that the third component w only depends on the directions x and y. In a sense, the Visibilities $V()$ and the image $I()$ have a two dimensional Fourier relationship ($V(u, v, w) = \int\int I(x, y)e^{2\pi i[ux+vy]}\, dx\, dy$), but with a directionally dependent correction factor $e^{2\pi i[...+w(c(x,y)-1)]}$.

The third component w is an example of a Directionally Dependent Effect (DDE) which have a tendency to increase the runtime costs of the image reconstruction. The w-component keeps us from using the Fast Fourier Transform (FFT) for the measurement equation (1.1). Research in this area tries to use approximations which lets us use faster algorithms like the FFT, and correct for DDE's accurately enough [2, 3, 4]. In this project, the w-correction is the only DDE we handle.

1.2 System of linear equations

Even though the Fourier Transform in the measurement equation (1.1) contains a w-correction factor, it is still linear. There exist more complex measurement equations, but the relationship often stays linear in nature [1, 5, 6, 7]. This means we can represent the measurement equation as a system of linear equations (1.2), where F is the Fourier Transform with w-correction, x is the image we are searching, V are the calibrated

[1] V in the equation (1.2) is a vector. We use the lower-case v to denote the axis in the Fourier space uvw, and the upper-case letter to denote the visibility vector.

visibilities. For the sake of demonstration, we ignore the noise in the visibility measurements in (1.2). To reconstruct the image x, we simply need to search for a solution to (1.2).

$$\underset{x}{find} \quad V - Fx = 0 \tag{1.2}$$

Even though we generally have more visibilities than pixels in the reconstruction, which makes (1.2) over-determined, there is no unique solution to it. There are potentially many candidate images that solve the equation, which makes the image reconstruction for radio interferometers an ill-posed inverse problem.

Ill-posed inverse problems arise in many different areas, such as image reconstruction[], de-noising[], and signal splitting[]. The problem is considered ill-posed when it has one of the following properties:

1. No solution exists.
2. There are solutions, but no unique solution exists.
3. The solution behaviour does not change continuously with the initial condition (For example: a small change in the measurements lead to a very different reconstructed image).

The image reconstruction problem for radio interferometers is ill-posed because of the second property: There is no unique solution to (1.2). At first glance, this might be counter-intuitive. The issue of not having a unique solution makes sense for under-determined systems of linear equations, when there are fewer visibilities than pixels in the reconstruction, as for example in x-ray image reconstruction[8]. Linear algebra tells us that there is a potentially infinite number of solutions to an under-determined system. However, as already mentioned, radio interferometers produce a large number of visibilities. The system (1.2) is generally over-determined. Yet, there are still many candidate images that fit the visibility measurements.

The reason why lies in the visibility measurements V. When we look back at figure 2a, it is clear to see that the interferometer does not sample the visibilities in a uniform way. There are regions with a high sample density. The density decreases when we move further away from the center. The higher mode visibilities get fewer and fewer samples. For high angular resolution images (when a pixel represents the emission from a small angle), the missing visibilities at higher mode are responsible for making the problem ill posed. For low angular resolution images on the other hand, the missing high mode visibilities become negligible and the problem is well-posed.

The ill-posed reconstruction problem arises from the high angular resolution of the reconstructed image. Or in other words, by solving the ill-posed inverse problem, we can acquire an image with higher angular resolution from the same visibility measurements. The rest of this introduction will continue with how we solve the ill-posed image reconstruction problem in practice, while section 1.3 will introduce what algorithms and architectures are used in practice.

1.2.1 Adding a regularization

For ill-posed inverse problems, there are two viewpoints for the same idea. From the viewpoint of optimization, we can solve the ill-posed image reconstruction problem (1.2) by adding a regularization. The regularization creates a system of linear equations with a unique solution. From the viewpoint of Bayesian Statistics, we include prior knowledge and search for most likely image given the measurements. For this project, both terms describe the same idea and we use regularization and prior interchangeably. We know that the reconstructed image from radio interferometers contain a mixture of stars (point sources located in a single pixel) and extended emissions like hydrogen clouds. By adding this prior knowledge to the reconstruction problem, we can find the most likely image given the measurements. As we will see, under the right prior, we can create a reconstruction algorithm that is almost guaranteed to find the truly observed image in theory.

There are different ways to include regularization in the reconstruction problem. In this project, we use the following method: We cast the image reconstruction problem into an optimization objective consisting of a data fidelity term and a regularization term. A reconstruction algorithm therefore consists of an optimization objective, a prior function and an optimization algorithm.

$$\underset{x}{minimize}\ \|V - Fx\|_2^2 + \lambda P(x) \tag{1.3}$$

The objective (1.3) has a data term $\|V - Fx\|_2^2$, which forces the most likely image to be as close to the measurements as possible, and the regularization term $P(x)$, which penalizes unlikely images according to our prior function. The parameter λ represents how much we penalize unlikely images and by extend, much noise we expect in the reconstruction. The parameter λ is either left to the user to define, can be estimated from the data [9].

The prior function $P()$ represents our prior knowledge about the image. It assigns a high penalty for unlikely images. In in radio interferometric image reconstructions tend to use similar prior functions as for image denoising applications. Such as: Total Variation ($\|\nabla x\|_1$) [10], L2 ($\|x\|_2$) [11] or the L1 norm in a wavelet domain ($\|Wx\|_1$) [12].

Finally, an optimization algorithm is necessary which can optimize the objective function (1.3) with the chosen prior function $P()$. Typical choices for optimization algorithms in image reconstructions are Interior Point Methods like Simplex, Matching Pursuit [13] and ADMM[14].

A reconstruction algorithm for radio astronomy consists of an optimization objective, a prior function and an optimization algorithm. It is a three dimensional design space. Not every prior is suitable for every optimization algorithm. The choice of optimization objective influences both what prior and what optimization algorithm we can use. Although there are a different choices for the optimization objective, in this project we limit ourselves to the objective (1.3).

The last question that remains is, how likely is it that the most observed image is We do not have the samples necessary for the Nyquist-Shannon theorem, where we need at least twice as much samples for the highest frequency in the signal. (Radio interferometers measure more than enough data, but as we already discussed, the sampling distribution is not uniform, and we are missing the information of the higher mode visibilities). The theory of Compressed Sensing[15, 16] gives us a theoretical framework for analysis. It shows that exact reconstruction is not only possible, but almost guaranteed in theory. Depends on the properties of the Fourier transform matrix F and the prior function $P()$

Compressible signals. Natural image are compressible in a wavelet domain. Meaning that they are sparse, only need a few non-zero coefficient in the wavelet domain. search the Interesting point is that noise tends to affect Sparse priors L0 norm in a wavelet domain. For exact reconstruction, all we actually need is the information on the location and magnitude of the non-zero wavelets. When we know a sparse space, we can use this as regularization for the image reconstruction. We can do an exact reconstruction from far fewer samples than stated by the Nyquist-Shannon rate. Properties of F. Measurement and reconstruction space need to be as incoherent as possible. Why, because we learn the most about the non-zero wavelet bases. Fourier Space and Image space are maximally incoherent.

To prove exact reconstruction of signals, the matrix F needs to fulfill the restricted isometry property. Necessary to prove exact reconstruction. Approximately orthonormal on sparse signals. (If we randomly choose ten columns of F, how much do they correlate. We want as little correlation as possible.) Calculate the RIP is NP-hard[17]. Approximations are also difficult to compute[18]. So we can only talk about how likely a given matrix fulfills the RIP. Matrix where each element is sampled from a random Gaussian distribut̂ion has the highest change. Not possible for Radio interferometers, because they sample in the fourier domain. Random samples in the Fourier domain also have a high chance to fulfill the RIP [19]. Depends on the antenna layout. MeerKAT quasi random antenna placement.

Active field of research, how we prove exact reconstruction in theory [20]. For our project, it is not as important. Currently, what the best prior function is for radio astronomy is still an open question.

1.3 Image reconstruction in practice

We know how to solve the ill-posed image reconstruction problem in theory. We formulate a minimization problem (1.3), specify a prior function that capture our prior knowledge, and find the optimum image with an appropriate optimization algorithm. In practice however we have a hard time representing the dense Fourier Transform matrix F in the equation (1.3). It is the size of number of visibilities times pixels in the reconstruction. Even older radio interferometers easily produce several million visibilities, with a million pixels in the reconstructed image. We cannot represent F explicitly in memory.

The major/minor cycle architecture is widely used for reconstruction algorithms in radio astronomy. We know the image is magnitudes smaller than the original measurements. The major cycle is responsible for an efficient transform from visibility in image space, while the minor cycle solves the ill-posed inverse problem in the image space. Turns out the ill-posed inverse problem of finding an image for the incomplete set of visibilities is equal to a deconvolution in image space. A deconvolution is still ill posed. We show the derivation in section 1.3.2, and discuss the major/minor cycle architecture in more details.

1.3.1 The major/minor cycle

The major/minor cycle architecture shown in figure 3 consists of three operations: Gridding, Fast Fourier Transform (FFT) and the minor cycle.

of two parts: The minor cycle, which iteratively deconvolves the dirty image with the PSF (it minimizes (1.6)). The first half of the major cycle estimates the dirty image. It consists of two steps: the gridding and the Fast Fourier Transformation (FFT). The gridding step takes the incomplete and non-uniformly sampled visibilities and interpolates them on a uniform grid. Then the inverse FFT can be calculated on the interpolated visibilities and we arrive at the dirty image.

Minor Cycle deconvolution. Creates the deconvolved ”model” image.

After the Minor Cycle, the second half of the Major Cycle continues. It takes the model image and transforms it back to the original measurement space. FFT and de-gridding. Arriving at the model visibilities.

Then we subtract the model visibilities from the original measurements and start the next Major Cycle with the residual visibilities. The final reconstructed image is the result of several major cycles. Depending on the observation and deconvolution algorithm we require more or fewer major cycles. Expected number around 5

Most expensive operations is the gridding and de-gridding followed by the deconvolution. Gridding is also where we do corrections like the w-term. So interpolation becomes specific for the field of radio astronomy.

Depending on the observation, the second most expensive step is the deconvolution.

1.3.2 Reformulating as a deconvolution problem

The Fourier transform matrix F in (1.3) is a product of two operations: The Fourier transform F, and the masking operator M. The masking operator is a matrix with zero entries for all Fourier components invisible to the instrument. So far, M was implicitly contained in F of (1.3). To derive the deconvolution problem, we represent M explicitly. For the sake of demonstration, let us assume our visibility measurements $V(u, v, w)$ lie on a discrete grid. $V(u, v, w)$ is zero for all components that the interferometer could not measure. We

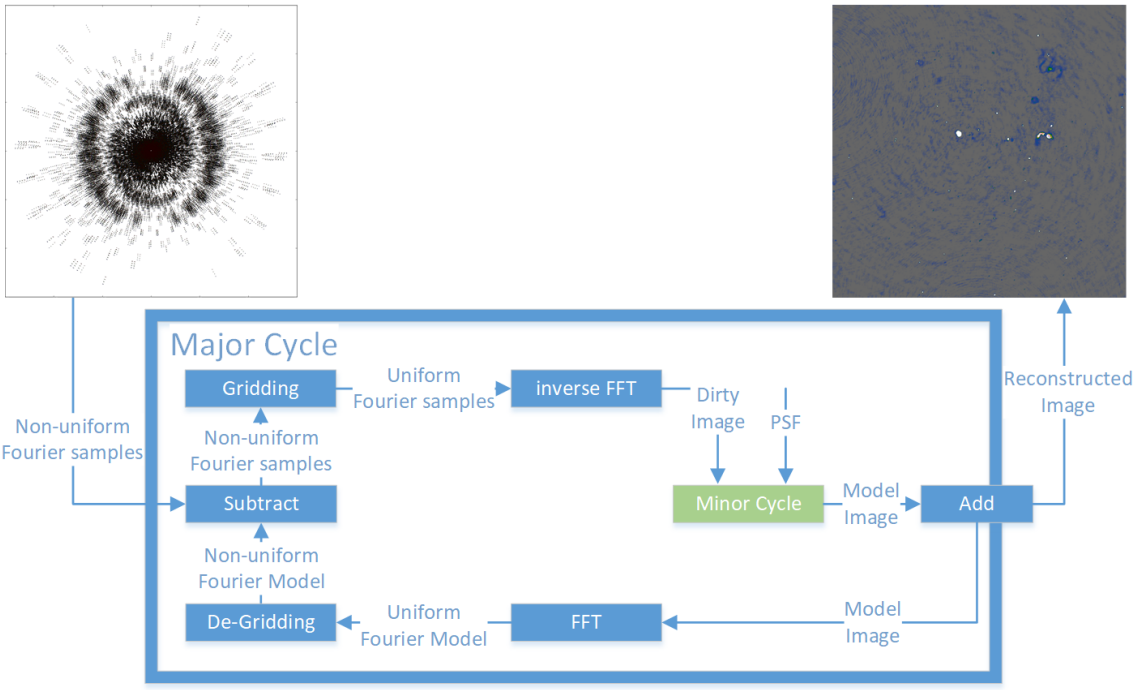


Figure 3: The Major/Minor Cycle Architecture

then can represent the transformation from image to visibility space with the Fourier transform F followed by a masking operation M, and we arrive at the image reconstruction problem (1.4). This is identical to (1.3), except for the factorization of F.

$$\text{original:} \quad \underset{x}{minimize} \ \|V - MFx\|_2^2 + \lambda P(x) \tag{1.4}$$

$$\text{in-painting:} \quad \underset{V_2}{minimize} \ \|V - MV_2\|_2^2 + \lambda P(F^{-1}V_2) \tag{1.5}$$

$$\text{deconvolution:} \quad \underset{x}{minimize} \ \|I_{dirty} - x * PSF\|_2^2 + \lambda P(x) \tag{1.6}$$

Note that M represents the degradation, the corruption introduced by incomplete sampling in the visibility space. M is the important operator. The measurements V, or the reconstructed image x can be in any space we wish. For example, we do not actually need to reconstruct the image in image space. In theory, we can reformulate an equivalent problem (1.5), in which we in-paint the missing visibilities. Or, we can use the Fourier transform on the visibility measurements V and the masking operator M, which leads us to the deconvolution problem (1.6).

Since the deconvolution formulation is vital for the major/minor cycle architecture, we have a closer look at (1.6). The effect of incomplete sampling in Fourier space is equal to a convolution with a Point Spread Function PSF in image space. I.e. $PSF = F^{-1}M$. The measurements are now represented as the ”dirty” image, $I_{dirty} = F^{-1}V$. We try to find the deconvolved image x, while only knowing the convolution kernel PSF and the convolved, dirty image I_{dirty}. This is still an ill-posed inverse problem. We have potentially many different deconvolutions that fit the dirty image, and we search the most likely candidate according to some prior $P(x)$.

The deconvolution (1.6) and the original image reconstruction problem (1.3) are equivalent. Both arrive at the same result. But the deconvolution problem is easier to handle in practice: I_{dirty} and PSF are generally more compact representations of V and M. There is one last issue: Calculating the dirty image from the measurements ($I_{dirty} = F^{-1}V$) again needs the impractically large Fourier transform matrix F. This is solved in the major/minor cycle algorithm.

1.3.3 Approximations under the major cycle

Why is the major cycle even necessary. At first glance, it seems like we may be done after the first major cycle. Because the deconvolution problem.

But the PSF is actually just an approximation. The PSF changes over the image.

Plus the gridding also introduces an error which gets reduced over several Major Cycles.

1.3.4 Alternatives to the major/minor cycle

Not formulating as a deconvolution, but takes the gridding and FFT step. Fastest approximation. Tends to use many major cycles.

2 State of the art image reconstruction

Long time research. Field that is cracking up

Major cycle architecture. Discuss the algorithms Split into two parts. We discuss the gridding first. It is responsible

2.1 Gridding algorithms

Biggest part is w-term, how to handle it efficiently. Important for wide field of view observation. Meaning really important for the SKA precursors like MeerKAT and MWA, and for the eventual SKA instruments.

(show the problem of w-component)

Accuracy and speed.

Faceting.

w-projection algorithm [21] Convolution in the Fourier domain.

The WSCLEAN [3] did a large part.

More recently, the Image domain gridding algorithm has been developed [2]. Which can put it on the gpu, along with handling more DDE's.

2.1.1 w-stacking

Idea of w-stacking, creating stacks of w-stack. So each visibility with similar w-component is in the same stack. Turns out we can then move

2.1.2 Image Domain Gridder

Recently developed[2]. It works by partitioning the input visibilities into subgrids, and then calculates the interpolation and w-correction for each subgrid.

Called image domain, because a convolution in Fourier space is a multiplication in image space, interpolation kernels can are applied in image space. uses the idea of subgrids.

Algorithm

2.2 Deconvolution algorithms

2.2.1 MS-MFS-CLEAN

CLEAN basic [13].

Latest variants for multiscale and multi frequency CLEAN (MS-MFS-CLEAN)[22].

Current state-of-the-art

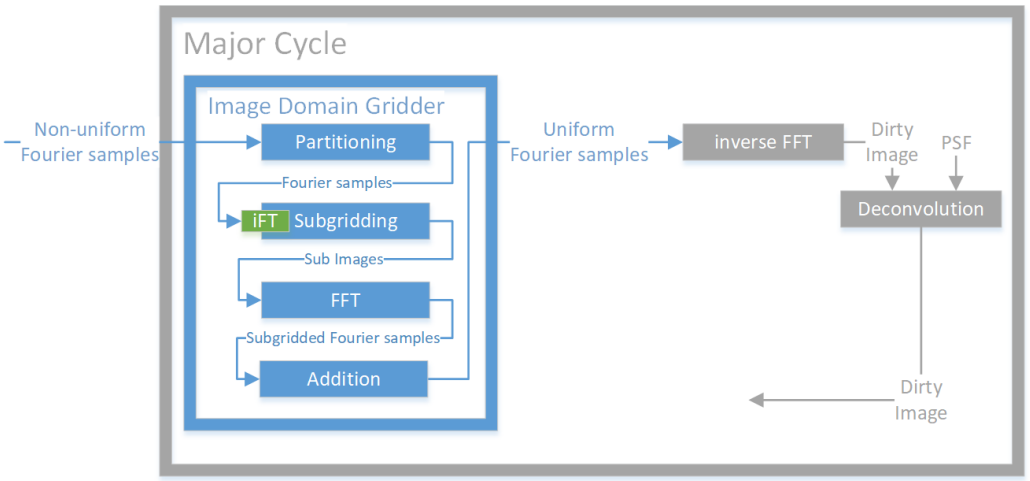


Figure 4: Image Domain Gridder in the Major Cycle Architecture

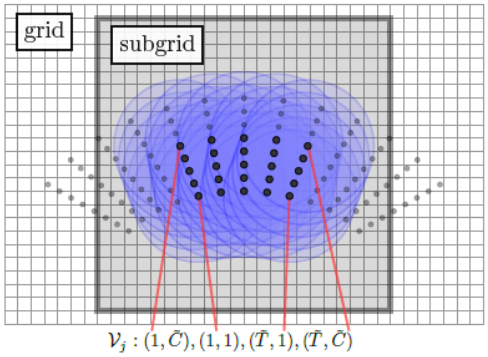


Figure 5: Subgrid

2.3 Reconstruction algorithms which are not based on the deconvolution

SARA

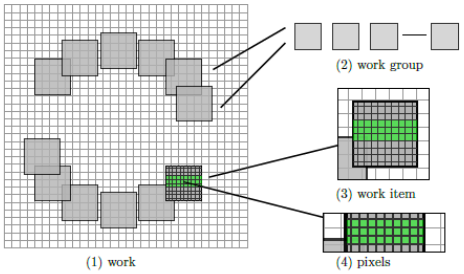


Figure 6: parallel

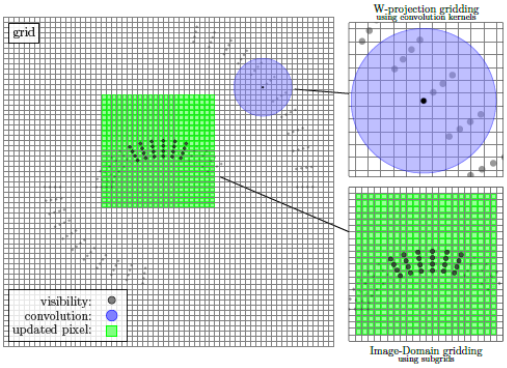


Figure 7: Image Domain Gridder in the Major Cycle Architecture

3 Simple distributed image reconstruction

Distribute the major cycle we need several major cycle iterations. We cannot execute the second cycle before the first. So we need to distribute the steps in the major cycle, namely the gridding, FFT and deconvolution step.

Gridding is generally the most time intensive step. Large number of visibilities, forces us to partition the input data. Meaning each node has only part of the data. This work assumes we have a copy of the full grid on each node. and exchange We use the Image Domain Gridding algorithm.

The FFT is generally not worth distributing, if we can keep all the data in memory. When the gridding is done, in our setup, the grid is small enough to keep in memory.

Deconvolution is potentially the second most time consuming step after the Gridding. Depends on the image content. Deconvolution produces the model image and the residual image. Distributed deconvolution requires communication. How we communicate is key. Distributing the deconvolution generally leaves us with two possible architectures: Either we split the deconvolution problem into patches, or we create a consensus algorithm. Consensus algorithms work by each node having a copy of the model image, and as a group, decide on the equal image. Not very useful for Meerkat, since we potentially do the work several times. We then want to distribute the work as evenly as possible between the nodes. That is why we chose to split the data onto nodes.

This leads us to the following architecture.

Rest of the Major Cycle. We need to get the model image patches from the nodes, FFT on node 0, and then we communicate the grid.

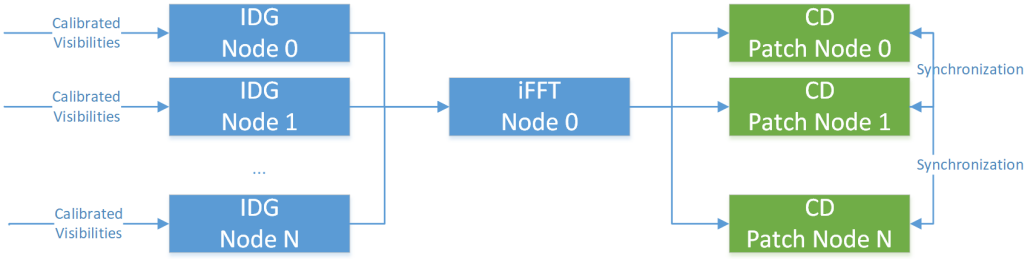


Figure 8: Distributed architecture for half a major cycle

3.1 Distributing the IDG Algorithm

It is simple, unlike w-stacking, we do not need to partition the data. So we can do just communicate the grid

3.2 Distributed deconvolution algorithm

Coordinate Descent Algorithm why.

Basic algorithm.

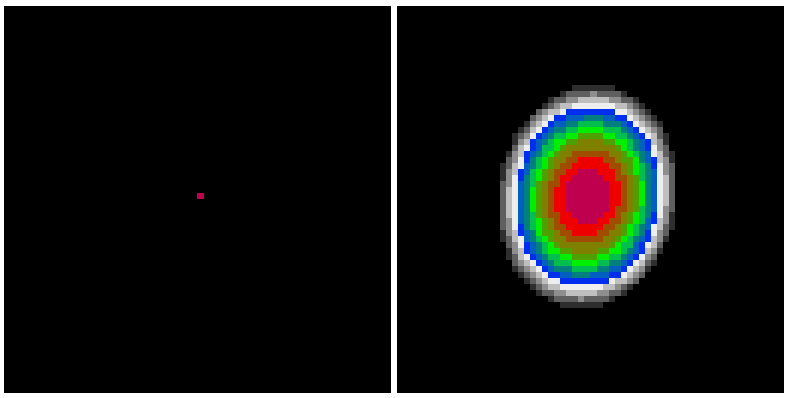
Implementation Correlate the dirty image Find max of objective How to calculate a in -b/2a

Optimizations: Why we only need to find the pixel which we can modify the most.

3.2.1 ElasticNet Regularization

L2 norm was used in other work. [**?**]

Formula



(a) Effect of the pure L1 norm (λ = 1.0) on a single point source.

(b) Effect of the pure L2 norm (λ = 1.0) on a single point source.

Figure 9: Effect of the L1 and L2 Norm separately.

Effect

Implementation

May even speed up convergence for correlated pixel values compared to L1 or L2[23]. But was not investigated in this project

3.3 Major Cycle convergence

Putting it all together

We have the Minor Cycle, which is easy to converge.

Coordinate Descent Path optimization [23] Danger that CD takes too many pixel into a Major Cycle. Lower bound per iteration, PSF sidelobe can still be too low, danger when many psf sidelobes overlap

3.4 Test on MeerKAT data

4 Tests on the MeerKAT LMC observation

For algorithmic testing

4.1 Wall clock time

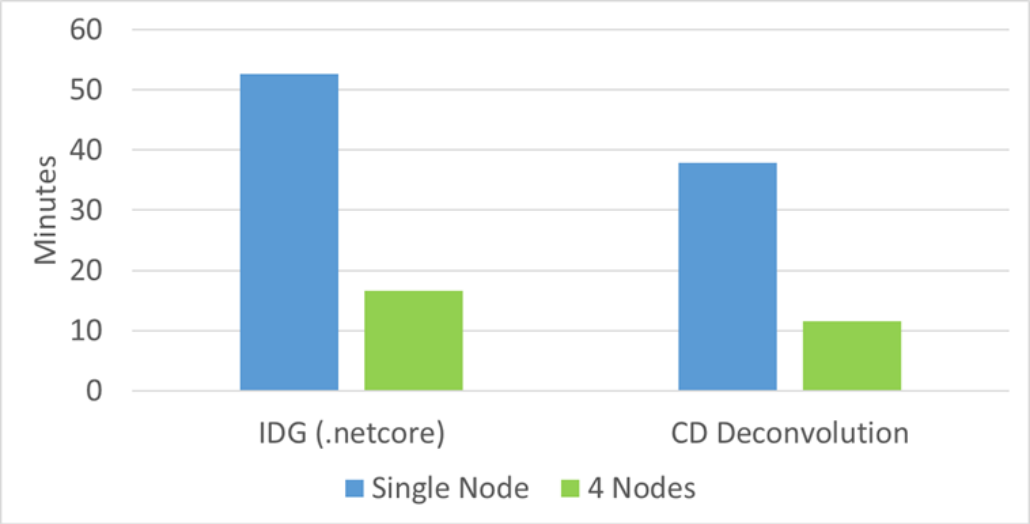


Figure 10: Wall-clock time of the distributed reconstruction

4.2 Quality

5 Conclusion

References

[1] Oleg M Smirnov. Revisiting the radio interferometer measurement equation-i. a full-sky jones formalism. Astronomy & Astrophysics, 527:A106, 2011.

[2] Bram Veenboer, Matthias Petschow, and John W Romein. Image-domain gridding on graphics processors. In 2017 IEEE International Parallel and Distributed Processing Symposium (IPDPS), pages 545–554. IEEE, 2017.

[3] AR Offringa, Benjamin McKinley, Natasha Hurley-Walker, FH Briggs, RB Wayth, DL Kaplan, ME Bell, Lu Feng, AR Neben, JD Hughes, et al. Wsclean: an implementation of a fast, generic wide-field imager for radio astronomy. Monthly Notices of the Royal Astronomical Society, 444(1):606–619, 2014.

[4] Luke Pratley, Melanie Johnston-Hollitt, and Jason D McEwen. A fast and exact w-stacking and w-projection hybrid algorithm for wide-field interferometric imaging. arXiv preprint arXiv:1807.09239, 2018.

[5] Oleg M Smirnov. Revisiting the radio interferometer measurement equation-ii. calibration and direction-dependent effects. Astronomy & Astrophysics, 527:A107, 2011.

[6] Oleg M Smirnov. Revisiting the radio interferometer measurement equation-iii. addressing direction-dependent effects in 21 cm wsrt observations of 3c 147. Astronomy & Astrophysics, 527:A108, 2011.

[7] Oleg M Smirnov. Revisiting the radio interferometer measurement equation-iv. a generalized tensor formalism. Astronomy & Astrophysics, 531:A159, 2011.

[8] Simon Felix, Roman Bolzern, and Marina Battaglia. A compressed sensing-based image reconstruction algorithm for solar flare x-ray observations. The Astrophysical Journal, 849(1):10, 2017.

[9] Keith Miller. Least squares methods for ill-posed problems with a prescribed bound. SIAM Journal on Mathematical Analysis, 1(1):52–74, 1970.

[10] Yves Wiaux, Laurent Jacques, Gilles Puy, Anna MM Scaife, and Pierre Vandergheynst. Compressed sensing imaging techniques for radio interferometry. Monthly Notices of the Royal Astronomical Society, 395(3):1733–1742, 2009.

[11] André Ferrari, David Mary, Rémi Flamary, and Cédric Richard. Distributed image reconstruction for very large arrays in radio astronomy. In 2014 IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM), pages 389–392. IEEE, 2014.

[12] Julien N Girard, Hugh Garsden, Jean Luc Starck, Stéphane Corbel, Arnaud Woiselle, Cyril Tasse, John P McKean, and Jérôme Bobin. Sparse representations and convex optimization as tools for lofar radio interferometric imaging. Journal of Instrumentation, 10(08):C08013, 2015.

[13] JA Högbom. Aperture synthesis with a non-regular distribution of interferometer baselines. Astronomy and Astrophysics Supplement Series, 15:417, 1974.

[14] Rafael E Carrillo, Jason D McEwen, and Yves Wiaux. Purify: a new approach to radio-interferometric imaging. Monthly Notices of the Royal Astronomical Society, 439(4):3591–3604, 2014.

[15] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on information theory, 52(2):489–509, 2006.

[16] David L Donoho. Compressed sensing. IEEE Transactions on information theory, 52(4):1289–1306, 2006.

[17] Andreas M Tillmann and Marc E Pfetsch. The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing. IEEE Transactions on Information Theory, 60(2):1248–1259, 2013.

[18] Abhiram Natarajan and Yi Wu. Computational complexity of certifying restricted isometry property. arXiv preprint arXiv:1406.5791, 2014.

[19] Ishay Haviv and Oded Regev. The restricted isometry property of subsampled fourier matrices. In Geometric Aspects of Functional Analysis, pages 163–179. Springer, 2017.

[20] Emmanuel J Candes and Yaniv Plan. A probabilistic and ripless theory of compressed sensing. IEEE transactions on information theory, 57(11):7235–7254, 2011.

[21] Tim J Cornwell, Kumar Golap, and Sanjay Bhatnagar. The noncoplanar baselines effect in radio interferometry: The w-projection algorithm. IEEE Journal of Selected Topics in Signal Processing, 2(5):647–657, 2008.

[22] Urvashi Rau and Tim J Cornwell. A multi-scale multi-frequency deconvolution algorithm for synthesis imaging in radio interferometry. Astronomy & Astrophysics, 532:A71, 2011.

[23] Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for generalized linear models via coordinate descent. Journal of statistical software, 33(1):1, 2010.

[24] Arwa Dabbech, Chiara Ferrari, David Mary, Eric Slezak, Oleg Smirnov, and Jonathan S Kenyon. Moresane: Model reconstruction by synthesis-analysis estimators-a sparse deconvolution algorithm for radio interferometric imaging. Astronomy & Astrophysics, 576:A7, 2015.

List of Figures

List of Tables

6 attachment

7 Larger runtime costs for Compressed Sensing Reconstructions

The MeerKAT instrument produces a new magnitude of data volume. An image with several million pixels gets reconstructed from billions of Visibility measurements. Although MeerKAT measures a large set of Visibilities, the measurements are still incomplete. We do not have all the information available to reconstruct an image. Essentially, this introduces ”fake” structures in the image, which a reconstruction algorithm has to remove. Additionally, the measurements are noisy.

We require an image reconstruction algorithm which removes the ”fake” structures from the image, and removes the noise from the measurements. The large data volume of MeerKAT requires the algorithm to be both scalable and distributable. Over the years, several reconstruction algorithms were developed, which can be separated into two classes: Algorithms based on CLEAN, which are cheaper to compute and algorithms based on Compressed Sensing, which create higher quality reconstructions.

CLEAN based algorithms represent the reconstruction problem as a deconvolution. First, they calculate the ”dirty” image, which is corrupted by noise and fake image structures. The incomplete measurements essentially convolve the image with a Point Spread Function (PSF). CLEAN estimates the PSF and searches for a deconvolved version of the dirty image. In each CLEAN iteration, it searches for the highest pixel in the dirty image, subtracts a fraction PSF at the location. It adds the fraction to the same pixel location of a the ”cleaned” image. After several iterations, the cleaned image contains the deconvolved version of the dirty image. CLEAN accounts for noise by stopping early. It stops when the highest pixel value is smaller than a certain threshold. This results in a light-weight and robust reconstruction algorithm. CLEAN is comparatively cheap to compute, but does not produce the best reconstructions and is difficult to distribute on a large scale.

Compressed Sensing based algorithms represent the reconstruction as an optimization problem. They search for the optimal image which is as close to the Visibility measurements as possible, but also has the smallest regularization penalty. The regularization encodes our prior knowledge about the image. Image structures which were likely measured by the instrument result in a low regularization penalty. Image structures which were likely introduced by noise or the measurement instrument itself result in high penalty. Compressed Sensing based algorithms explicitly handle noise and create higher quality reconstructions than CLEAN. State-of-the-art Compressed Sensing algorithms show potential for distributed computing. However, they currently do not scale on MeerKATs data volume. They require too many computing resources compared to CLEAN based algorithms.

This project searches for a way to reduce the runtime costs of Compressed Sensing based algorithms. One reason for the higher costs is due to the non-uniform FFT Cycle. State-of-the-art CLEAN and Compressed Sensing based algorithms both use the non-uniform FFT approximation in a cycle during reconstruction. The interferometer measures the Visibilities in a continuous space in a non-uniform pattern. The image is divided in a regularly spaced, discrete pixels. The non-uniform FFT creates an approximate, uniformly sampled image from the non-uniform measurements. Both, CLEAN and Compressed Sensing based algorithms use the non-uniform FFT to cycle between non-uniform Visibilities and uniform image. However, a Compressed Sensing algorithm requires more non-uniform FFT cycles for reconstruction.

CLEAN and Compressed Sensing based algorithms use the non-uniform FFT in a similar manner. However, there are slight differences in the architecture. This project hypothesises that The previous project searched for an alternative to the non-uniform FFT cycle. Although there are alternatives, there is currently no replacement which leads to lower runtime costs for Compressed Sensing. Current research is focused on reducing the number of non-uniform FFT cycles for Compressed Sensing algorithms.

CLEAN based algorithms use the Major Cycle Architecture for reconstruction. Compressed Sensing based algorithms use a similar architecture, but with slight modifications. Our hypothesis is that we may reduce the number of non-uniform FFT cycles for Compressed Sensing by using CLEAN’s Major Cycle Architecture.

7.1 CLEAN: The Major Cycle Architecture

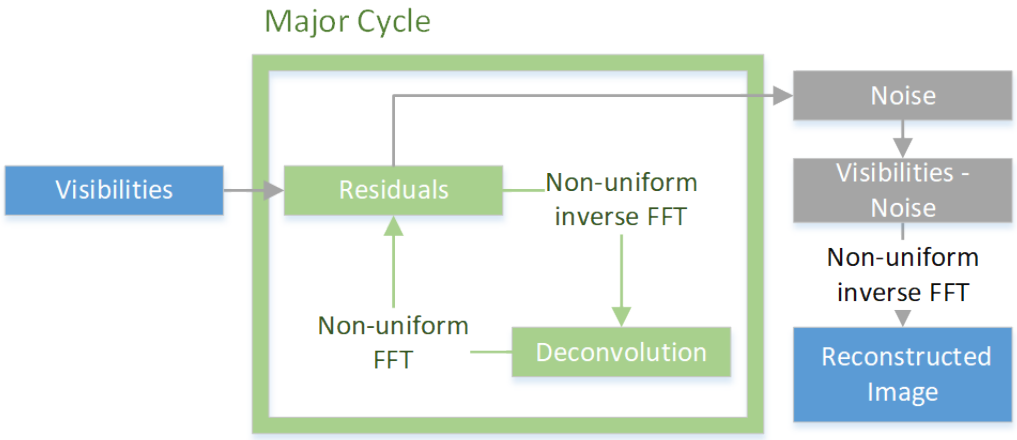


Figure 11: The Major Cycle Architecture

Figure 11 depicts the Major Cycle Architecture used by CLEAN algorithms. First, the Visibilities get transformed into an image with the non-uniform FFT. The resulting dirty image contains the corruptions of the measurement instrument and noise. A deconvolution algorithm, typically CLEAN, removes the corruption of the instrument with a deconvolution. When the deconvolution stops, it should have removed most of the observed structures from the dirty image. The rest, mostly noisy part of the dirty image gets transformed back into residual Visibilities and the cycle starts over.

In the Major Cycle Architecture, we need several deconvolution attempts before it has distinguished the noise from the measurements. Both the non-uniform FFT and the deconvolution are approximations. By using the non-uniform FFT in a cycle, it can reconstruct an image at a higher quality. For MeerKAT reconstruction with CLEAN, we need approximately 4-6 non-uniform FFT cycles for a reconstruction.

7.2 Compressed Sensing Architecture

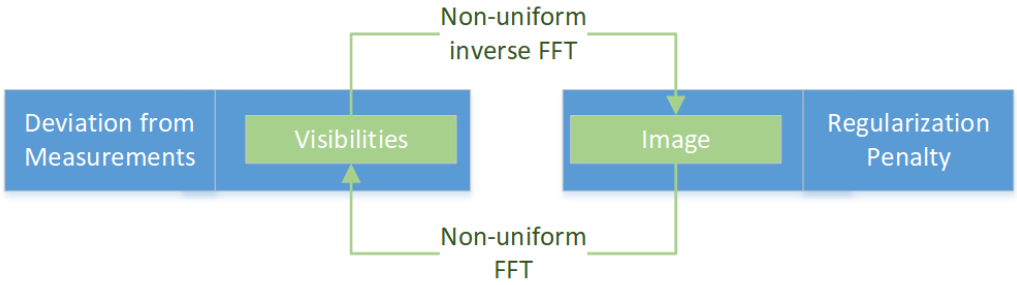


Figure 12: State-of-the-art Compressed Sensing Reconstruction Architecture

Figure 12 depicts the architecture used by Compressed Sensing reconstructions. The Visibilities get transformed into an image with the non-uniform FFT approximation. The algorithm then modifies the image so it reduces the regularization penalty. The modified image gets transformed back to Visibilities and the algorithm then minimizes the difference between measured and reconstructed Visibilities. This is repeated until the algorithm converges to an optimum.

In this architecture, state-of-the-art Compressed Sensing algorithms need approximately 10 or more non-uniform FFT cycles to converge. It is one source for the higher runtime costs. For MeerKAT reconstructions

the non-uniform FFT tends to dominate the runtime costs. A CLEAN reconstruction with the Major Cycle Architecture already spends a large part of its time in the non-uniform FFT. Compressed Sensing algorithms need even more non-uniform FFT cycle on top of the ”Image Regularization” step being generally more expensive than CLEAN deconvolution. There is one upside in this architecture: State-of-the-art algorithms managed to distribute the ”Image Regularization” operation.

7.3 Hypothesis for reducing costs of Compressed Sensing Algorithms

Compressed Sensing Algorithms are not bound to the Architecture presented in section 7.2. For example, we can design a Compressed Sensing based deconvolution algorithm and use the Major Cycle Architecture instead.

Our hypothesis is: We can create a Compressed Sensing based deconvolution algorithm which is both distributable and creates higher quality reconstructions than CLEAN. Because it also uses the Major Cycle architecture, we reckon that the Compressed Sensing deconvolution requires a comparable number of non-uniform FFT cycles to CLEAN. This would result in a Compressed Sensing based reconstruction algorithm with similar runtime costs to CLEAN, but higher reconstruction quality and higher potential for distributed computing.

7.4 State of the art: WSCLEAN Software Package

7.4.1 W-Stacking Major Cycle

7.4.2 Deconvolution Algorithms

CLEAN MORESANE

7.5 Distributing the Image Reconstruction

7.5.1 Distributing the Non-uniform FFT

7.5.2 Distributing the Deconvolution

8 Handling the Data Volume

The new data volume is a challenge to process for both algorithms and computing infrastructure. Push for parallel and distributed algorithms. For Radio Interferometer imaging, we require specialized algorithms. The two distinct operations, non-uniform FFT and Deconvolution, were difficult algorithms for parallel or distributed computing.

The non-uniform FFT was historically what dominated the runtime []. Performing an efficient non-uniform FFT for Radio Interferometers is an active field of research[3, 4], continually reducing the runtime costs of the operation. Recently, Veeneboer et al[2] developed a non-uniform FFT which can be fully executed on the GPU. It speeds up the most expensive operation.

In Radio Astronomy, CLEAN is the go-to deconvolution algorithm. It is light-weight and compared to the non-uniform FFT, a cheap algorithm. It is also highly iterative, which makes it difficult for effective parallel or distributed implementations. However, compressed sensing based deconvolution algorithms can be developed with distribution in mind.

8.1 Fully distributed imaging algorithm

Current imaging algorithms push towards parallel computing with GPU acceleration. But with Veeneboer et al's non-uniform FFT and a compressed sensing based deconvolution, we can go a step further and create a distributed imaging algorithm.

9 Image Reconstruction for Radio Interferometers

In Astronomy, instruments with higher angular resolution allows us to measure ever smaller structures in the sky. For Radio frequencies, the angular resolution is bound to the antenna dish diameter, which puts practical and financial limitations on the highest possible angular resolution. Radio Interferometers get around this limitation by using several smaller antennas instead. Together, they act as a single large antenna with higher angular resolution at lower financial costs compared to single dish instruments.

Each antenna pair of an Interferometer measures a single Fourier component of the observed image. We can retrieve the image by calculating the Fourier Transform of the measurements. However, since the Interferometer only measures an incomplete set of Fourier components, the resulting image is ”dirty”, convolved with a Point Spread Function (PSF). Calculating the Fourier Transform is not enough. To reconstruct the from an Interferometer image, an algorithm has to find the observed image with only the dirty image and the PSF as input. It has to perform a deconvolution. The difficulty lies in the fact that there are potentially many valid deconvolutions for a single measurement, and the algorithm has to decide for the most likely one. How similar the truly observed image and the reconstructed images are depends largely on the deconvolution algorithm.

State-of-the-art image reconstructions use the Major Cycle architecture (shown in Figure 13), which contains three operations: Gridding, FFT and Deconvolution.

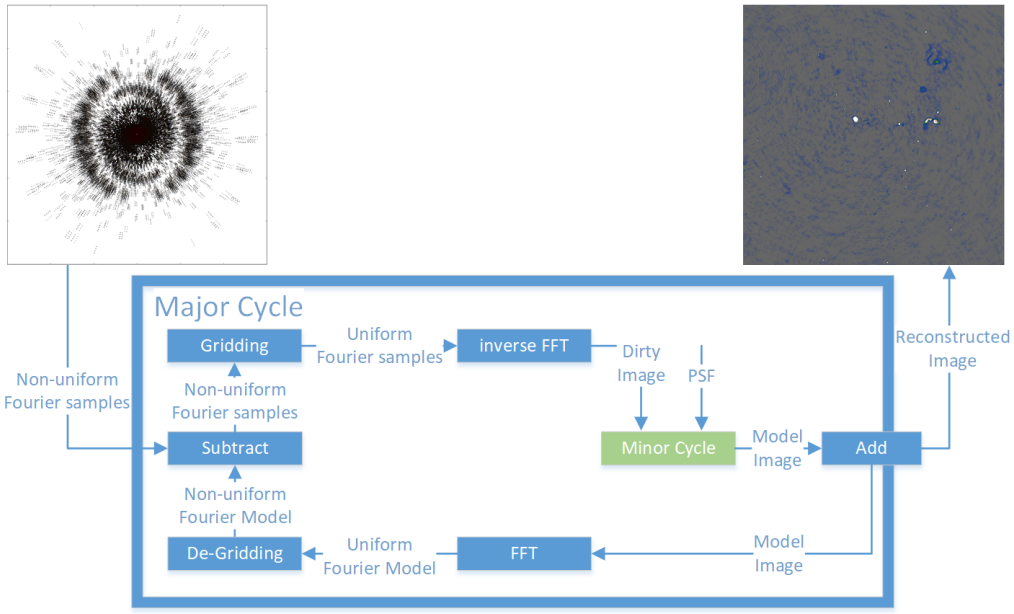


Figure 13: The Major Cycle Architecture of image reconstruction algorithms

The first operation in the Major Cycle, Gridding, takes the non-uniformly sampled Fourier measurements from the Interferometer and interpolates them on a uniformly spaced grid. The uniform grid lets us use FFT to calculate the inverse Fourier Transform and we arrive at the dirty image. A deconvolution algorithm takes the dirty image plus the PSF as input, producing the deconvolved ”model image”, and the residual image as output. At this point, the reverse operations get applied to the residual image. First the FFT and then De-gridding, arriving at the non-uniform Residuals. The next Major Cycle begins with the non-uniform Residuals as input. The cycles are necessary, because the Gridding and Deconvolution operations are only approximations. Over several cycles, we reduce the errors introduced by the approximate Gridding and Deconvolution. The final, reconstructed image is the addition of all the model images of each Major Cycle.

9.1 Distributed Image Reconstruction

New Interferometer produce an ever increasing number of measurements, creating ever larger reconstruction problems. A single image can contain several terabytes of Fourier measurements. Handling reconstruction problems of this size forces us to use distributed computing. However, state-of-the-art Gridding and Deconvolution algorithms only allow for limited distribution. How to scale the Gridding and Deconvolution algorithms to large problem sizes is still an open question.

Recent developments make a distributed Gridder and a distributed Deconvolution algorithm possible. Veeneboer et al[2] found an input partitioning scheme, which allowed them to perform the Gridding on the GPU. The same partitioning scheme can potentially be used to distribute the Gridding onto multiple machines. For Deconvolution, there exist parallel implementations for certain algorithms like MORESANE[24]. These can be used as a basis for a fully distributed image reconstruction.

In this project, we want to make the first steps towards an image reconstruction algorithm, which is distributed from end-to-end, from Gridding up to and including deconvolution. We create our own distributed Gridding and Deconvolution algorithms, and analyse the bottlenecks that arise.

9.2 First steps towards a distributed Algorithm

In this project, we make the first steps towards a distributed Major Cycle architecture (shown in figure 13) implemented C#. We port Veeneboer et al's Gridder, which is written in C++, to C# and modify it for distributed computing. We implement a simple deconvolution algorithm based on the previous project and create a first, non-optimal distributed version of it.

In the next step, we create a more sophisticated deconvolution algorithm based on the shortcomings of the first implementation. We use simulated and real-world observations of the MeerKAT Radio Interferometer and measure its speed up. We identify the bottlenecks of the current implementation and explore further steps.

From the first lessons, we continually modify the distributed algorithm and focus on decreasing the need for communication between the nodes, and increase the overall speed up compared to single-machine implementations. Possible Further steps:

- Distributed FFT
- Replacing the Major Cycle Architecture
- GPU-accelerated Deconvolution algorithm.

A state-of-the-art reconstruction algorithm has to correct large number of measurement effects arising from the Radio Interferometer. Accounting for all effects is out of the scope for this project. We make simplifying assumptions, resulting in a proof-of-concept algorithm.

10 Ehrlichkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende schriftliche Arbeit selbstständig und nur unter Zuhilfenahme der in den Verzeichnissen oder in den Anmerkungen genannten Quellen angefertigt habe. Ich versichere zudem, diese Arbeit nicht bereits anderweitig als Leistungsnachweis verwendet zu haben. Eine Überprüfung der Arbeit auf Plagiate unter Einsatz entsprechender Software darf vorgenommen werden.
Windisch, July 15, 2019

Jonas Schwammberger