P9 Distributed Image Reconstruction for the new Radio Interferometers

Jonas Schwammberger

June 14, 2019

Abstract

Contents

1 Introduction

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.

However, Radio Interferometers do not measure the pixels of the sky image. Instead, Radio Interferometers measure an incomplete set of Fourier components. The sky image has to be reconstructed from the measurements. This forms an ill-posed inverse problem: There are potentially many different images that fit the measurements, and from the measurements alone we cannot decide which image was actually observed.

Algorithms that solve the ill-posed Image Reconstruction problem. Find most likely images. Extensive Research in the last decades.

SKA wants to create Radio Interferometers on a new scale. MeerKAT is the precursor of SKA-Mid. Create a new scale of Fourier measurements. Push towards distributing The Image Reconstruction Problem has to be solved with distributed computing.

Algorithms were developed before the advent of distributed computing. Distribution so far has been difficult. Only small number of nodes.

Target to distribute the image reconstruction First tests

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.

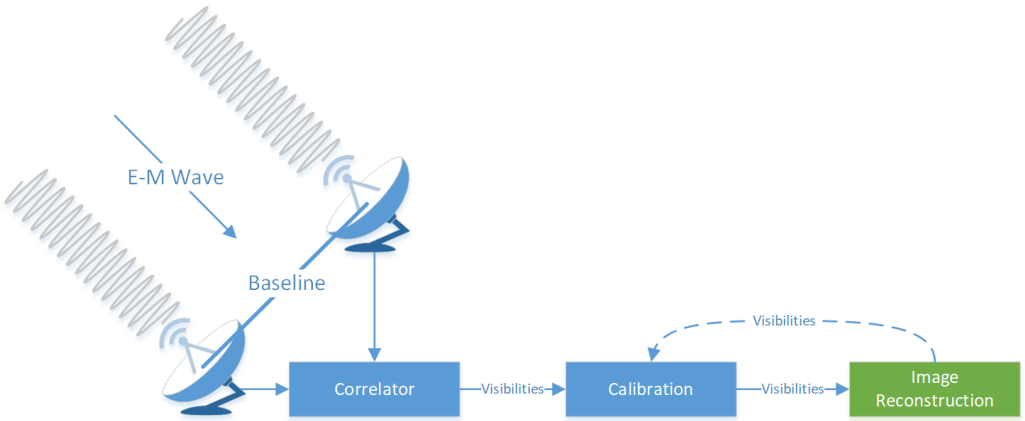


Figure 1: Radio Interferometer System

First, the electromagnetic wave gets measured by the different antennas of the interferometer. The measurements of each antenna pair get correlated into a complex-valued Fourier Component (called Visibility in Radio Astronomy). Each antenna pair measures a noisy amplitude and phase of a single Visibility (Fourier Component) of the sky image. 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. 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 Visibilities are calibrated, we can average the measurements and reduce the data volume by several factors. Typically, averaging is done to reduce the runtime costs of the Image Reconstruction step.

The Image Reconstruction step takes the calibrated, and potentially averaged down Visibility data and finds the most likely observed image of the sky. Although the Interferometer produces a potentially large number of Visibilities, they are incomplete: For example an Interferometer with dish-antennas is typically blind to the microwave background radiation. The largest structures in the image it can detect depends on the shortest baseline. Since the antennas have to be at least the dish-diameter placed apart, the Interferometer is simply blind to very large structures in the sky image, like the microwave background radiation. This property makes Image Reconstruction Problem ill-posed. In the Image Reconstruction step, we therefore have to find the most likely image which fits the measurements.

1.1.1 Earth's rotation and arbitrary large Number of Visibilities

To improve the final image, we want to measure as many different Visibilities as possible. Modern Radio Interferometers use the the earth's rotation to change the baselines. When the earth rotates, it modifies the length of the baseline and by extent, what Visibilities get measured. Modern Interferometer can produce an almost arbitrary large number of measurements by just increasing the observation time.

The data volume can be averaged down in the Calibration step. However, with self-calibration, the Image Reconstruction is tasked with solving both the most likely image and the calibration parameters at the same time. This further improves the reconstruction quality[**?**], but requires the Image Reconstruction step to handle all the Visibility data.

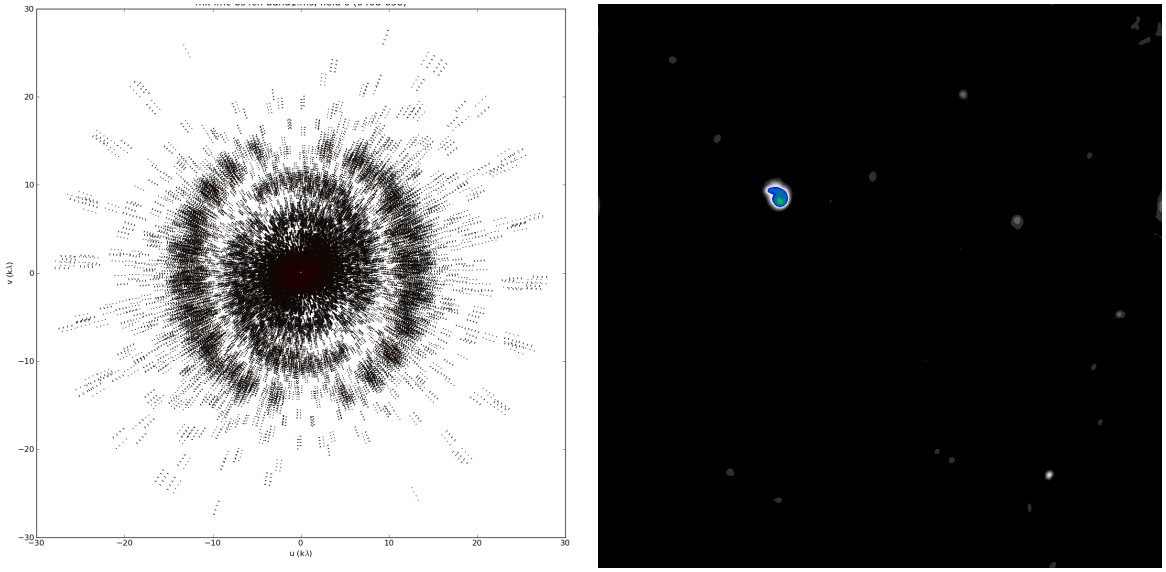
Modern Radio Interferometers can produce an almost arbitrary large number of measurements. The reconstruction quality benefits from a large number of different Visibility measurements. The two main limiting factors however, are the cost of data storage and the scalability of the Image Reconstruction algorithm.

1.2 The Image Reconstruction Problem

For Radio Interferometers, Image Reconstruction forms an ill-posed inverse problem. There are potentially many different images that fit the measurements. The Image Reconstruction is tasked with finding the most likely image $I()$ given the (calibrated) Visibility measurements $V()$. Figure 2 shows an example from a MeerKAT observation for $V()$ and $I()$. The image 2a shows the incomplete sampling in the Fourier space. The sample density decreases away from the center[1]. The observed image 2b contains two classes of structure, point sources and extended emissions. Point sources arise from stars and other distant objects, their emissions are concentrated around a single pixel. Extended emissions arise from hydrogen clouds and other sources which span an area of several arc-seconds of the sky. Or more formally, we want to invert the measurement equation (1.1) of the radio interferometer and retrieve the observed image $I()$. As we will see, this is an ill-posed inverse problem, and we have a large number of images that fit the measurements.

Let us look at the measurement equation (1.1) in more detail. It consists of three parts: The Visibility measurements $V(u, v, w)$, the observed image with a normalization factor $\frac{I(x,y)}{c(x,y)}$ and the Fourier Transform $e^{2\pi i[...]}$. u

[1] The center has actually no samples, since the MeerKAT Radio Interferometer uses antenna's with dishes. Due to the resolution, it might not be visible in the image 2a.



(a) Measurements $V()$ in the uv-plane. *(b) The observed image $I()$.*

Figure 2: The Image Reconstruction Problem

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.

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

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. However, the inverse Fourier Transform does not give us the observed image $I()$, because the measurements $V()$ are incomplete.

Note that the Visibilities $V(u, v, w)$ are three dimensional, while the image $I(x, y)$ only has two. However, 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 [1, 2, 3]. In this project, the w-correction is the only DDE we handle.

1.2.1 System of Linear Equations

Even though the Fourier Transform in equation (1.1) contains a w-correction factor, $V()$ and $I()$ have still a linear relationship. This means we can represent the Image Reconstruction Problem as a system of linear equations: (1.2). F is the Fourier Transform with w-correction, x is the image we are searching and V are the measured Visibilities[2].

[2] 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.

All the relevant deductions can be made in this.

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

Under-Determined system, because V is incomplete. We cannot get the observed image x Add prior information. We know the image contains hydrogen clouds and distant stars. We can use this information and find the most likely image.

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

This is a minimization problem, where we have the data term and regularization term. Data term forces image to be close to the measurements. Regularization term penalizes unlikely images. What are the chances of

As we discussed in section 1.1, Radio Interferometers can produce an almost arbitrary large number of Visibilities. In practice, we often reconstruct images with far fewer pixels than Visibility measurements, making the equation (1.2) an over-determined problem. Meaning equation (1.2) is either consistent and has a unique solution x, or is inconsistent with no solutions. Sadly, the Visibility measurements are subject to noise. The equation (1.2) is inconsistent does not have a solution. To solve the Image Reconstruction Problem, we need to account for noise in the measurements, which leads us to a new inequality (1.4).

$$\underset{x}{find} \ \|V - Fx\|_2^2 < \epsilon \tag{1.4}$$

We use the L2 norm and find the solution x which, when the Fourier Transform is applied, has the smallest distance to the measurements V. The L2 norm leads to a strictly-convex function, meaning (1.4) has only one minimum, and we can use convex optimization techniques to find it. So the question arises, is the observed image $I()$ located at the global minimum of (1.4)? Although we have more measurements than pixels in the reconstruction, the set of Visibilities are still incomplete. They shift the observed image some distance away from the global minimum of (1.4). From the measurements alone, we cannot locate the observed image. We know it is near the global minimum, but we do not know where it is. The equation (1.4) has many candidate solutions x, one of which is the observed image. From the measurements alone, we cannot decide which candidate was the observed image, and (1.4) is an ill-posed inverse problem.

Intuitively, the incomplete measurements allow the pixels x in the reconstruction to form physically implausible solutions. By including prior information, we can distinguish between likely and unlikely candidates of (1.4), and hopefully find the observed image. We can use the Theory of Compressed Sensing allows us to encode prior information about the image, and find the most likely solution to the ill-posed Image Reconstruction Problem.

1.2.2 Theory of Compressed Sensing

RIP of F

Recently, the Compressed Sensing sampling theorem has emerged[4, 5], and has applications in data compression, de-noising[6], and image reconstruction. For our intends and purposes, the Theory of Compressed Sensing gives us a way to encode prior information into the Image Reconstruction Problem. In essence, we choose a dictionary of basis functions D to represent our image $x = D\alpha$. With the appropriate dictionary D, the most likely solution is the one with the fewest non-zero components in α, and we arrive at the minimization objective (1.5) to solve our Image Reconstruction Problem. But first, let us look at how we chose the dictionary D and how we include prior information in more detail.

We can always represent the image in a different space, for example with Haar-wavelets. In that case D in $x = D\alpha$ is a dictionary of Haar-Wavelets, α is the Haar-Wavelet component vector and x are the pixel values of our image. Natural images are sparse in Haar-Wavelet space. Few non-zero components in α If the natural image is corrupted with noise, it tends to affect all α's equally, meaning it is not sparse any more. However, the most significant components are still the original ones. So by searching for a solution which is sparse in Haar-Wavelet space, we can de-noise the image.

For our Image Reconstruction Problem, we need a dictionary D in which our observed image can be sparsely represented. If we choose the right D, the Theory of Compressed Sensing guarantees us the most likely image is the observed one. We search for the image x which is as close to the measurements as possible, but also has the fewest non-zero components α. We can formulate this as a LASSO minimization problem (1.4).

$$\underset{x}{minimize}\ \|V - Fx\|_2^2 + \lambda \left\|D^{-1}x\right\|_1 \tag{1.5}$$

The objective contains two terms: The data term $\|V - Fx\|_2^2$, which is identical to the previous inequality (1.4), and the regularization term $\lambda \left\|D^{-1}x\right\|_1$, which enforces sparsity in our Dictionary. Reconstruction quality depends on the choice of D Radio Astronomy three different D are in wide use, Starlets, Daubechies Wavelets, and sparsity in image domain. Starlets and daubechies lead to better reconstruction quality,

So we need a Regularization Dicitionary D, which captures the prior information we have about the image. Then, we can use convex optimization techniques to minimize the objective (1.5).

1.2.3 Different Representations for the LASSO Objective

There are different ways to represent the LASSO objective (1.5). Has the data term in the Fourier space, and reconstructs the image directly. We have different design decisions here. We can choose the Data term to be either in Fourier or image space, and we can set the Variables to be in either image, uniform Fourier or in the sparse space.

These possibilities arise from the Fourier Transform Matrix F, shown in equation (1.6). The first line shows the normal Fourier transform, from image x into Visibilities V. Since x is in a uniformly-sampled space, we can factorize F into the FFT and the masking matrix M and we arrive at the second line of equation (1.6). M is essentially a matrix contains values between zero and one. Represents the degradation due to incomplete measurements.

$$\begin{aligned} V &= Fx \\ V &= M\ FFT(x) \\ V &= M\ V_2 \\ I_{Dirty} &= x \star PSF \end{aligned} \tag{1.6}$$

We can use the matrix M directly, and use it to transform from uniformly sampled Visibilities V_2 to non-uniformly sampled Visibilities V, arriving at line three of equation (1.6).

Since a multiplication in Fourier space is a convolution in image space, we can also represent the masking matrix M as a Point Spread Function PSF in image space. Degradation is a convolution. I_{Dirty} calculation is a Fourier Transform of the measurements. So we need to solve a deconvolution problem. DIRTY IMAGE

All these lines are of (1.6) all represent the Image Reconstruction Problem in different spaces. For example, if we change the data term in our LASSO objective from $\|V - Fx\|_2^2$ to a deconvolution $\|I_{Dirty} - x \star PSF\|_2^2$

we arrive at the same global minimum. But, in Radio Astronomy, the images are magnitudes smaller than the calibrated Visibility measurements V. Reducing the problem size. In practice, we have an accuracy issue with the PSF, since it is not constant. The PSF changes over the image due to DDEs like w-correction.

$$\begin{aligned} \underset{x}{minimize}\ \|V - Fx\|_2^2 + \lambda \left\|D^{-1}x\right\|_1 \\ \underset{x}{minimize}\ \|I_{Dirty} - x \star PSF\|_2^2 + \lambda \left\|D^{-1}x\right\|_1 \end{aligned} \tag{1.7}$$

We will use deconvolution in the state-of-the-art.

Explain data space, reconstruction space and regularization space,

We can set the data term in our LASSO objective to either Visibility or image space. The reconstruction space has three possible design choices, which leads to three different LASSO objectives: Analysis, in-painting and synthesis.

$$\text{analysis: } \underset{x}{minimize}\ \|V - Fx\|_2^2 + \lambda \left\|D^{-1}x\right\|_1 \tag{1.8}$$

$$\text{in-painting: } \underset{V_2}{minimize}\ \|V - MV_2\|_2^2 + \lambda \left\|D^{-1}F^{-1}V_2\right\|_1 \tag{1.9}$$

$$\text{synthesis: } \underset{\alpha}{minimize}\ \|V - FD\alpha\|_2^2 + \lambda \|\alpha\|_1 \tag{1.10}$$

Analysis, standard formulation.

To our knowledge, in-painting (1.9) is not used for the Image Reconstruction Problem. Difficulty in the Regularization term.

Synthesis reconstructs in the sparse space, increases the number of free variables, we have at least as many α than x. But, we do not need the Matrix D^{-1}, which for certain spaces may not even be defined. We can use over-complete representations. Why use over-complete? More freedom?

1.3 Radio interferometric 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. However, we can use two ”tricks” to:

1. Reduce the dimensions of the measurements V.
2. Approximate runtime intensive operations.

The first ”trick” is to reformulate the image reconstruction as a deconvolution problem, which reduces the dimensions of the measurements. The second trick is to split the minimization problem into two cycles, the major/minor cycle. The major/minor cycle architecture was developed for radio interferometric image reconstruction and has been in use for decades [7]. First, we derive the deconvolution problem and then introduce the major/minor cycle architecture.

1.3.1 Reformulating as a deconvolution problem

We can reformulate the image reconstruction (1.3) as a deconvolution problem, which reduces the dimensions of the measurement V and removes the large Fourier transform matrix F from the minimization objective.

Note that the Fourier transform matrix F in (1.3) is a product of two operations: The Fourier transform F, and the masking operator M. The Fourier transform was established in section 1.2. The masking operator is simply 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 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 (1.11). This is identical to (1.3), except for the factorization of F.

$$\text{original:} \quad \underset{x}{minimize} \ \|V - MFx\|_2^2 + \lambda P(x) \tag{1.11}$$

$$\text{in-painting:} \quad \underset{V_2}{minimize} \ \|V - MV_2\|_2^2 + \lambda P(F^{-1}V_2) \tag{1.12}$$

$$\text{deconvolution:} \quad \underset{x}{minimize} \ \|I_{dirty} - x * PSF\|_2^2 + \lambda P(x) \tag{1.13}$$

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 (1.12), 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 (1.13).

Since the deconvolution formulation is vital for the major/minor cycle architecture, we have a closer look at (1.13). 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 (1.13) 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. Together with the major/minor cycle architecture, we can introduce approximations that simplify the deconvolution problem.

1.3.2 The major/minor cycle

The major/minor cycle architecture shown in figure 3 consists of two parts: The minor cycle, which iteratively deconvolves the dirty image with the PSF (it minimizes (1.13)). The major cycle is more complex. The first half of the major cycle approximates the dirty image ($I_{dirty} \approx F^{-1}V$) for the minor cycle, while the second half takes the deconvolved image, called the ”model” image in the major/minor cycle architecture, and transforms it back to the original visibility space ($V_{model} \approx Fx$). The next major cycle then uses the residual visibilities ($V_{res} = V - V_{model}$) and repeats the process.

At first glance, it may not be clear why several major cycles are even necessary. It is precisely this cycle that allows us to introduce approximations in the minor cycle. We will go into more detail about the approximations

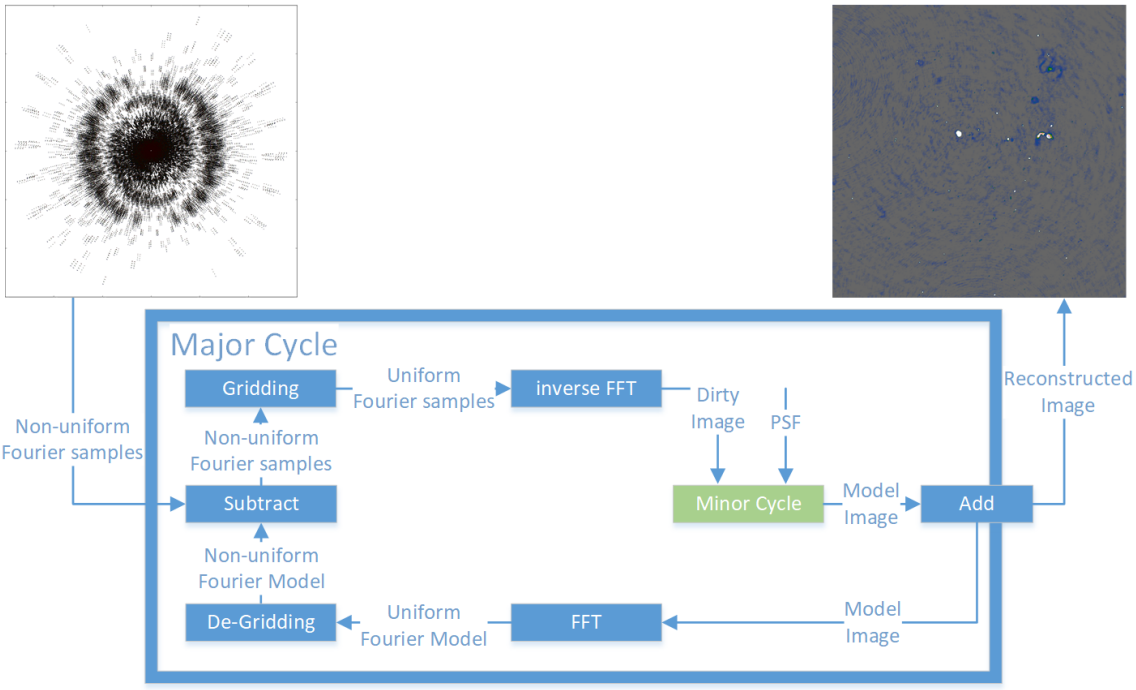


Figure 3: The Major/Minor Cycle Architecture

in section 1.3.3 and why they are important for radio interferometers. This section dives deeper into the core operation of the major cycle, the gridding.

Calculating the dirty image from the measurements ($I_{dirty} = F^{-1}V$) again needs the impractically large Fourier transform matrix F. We cannot use the inverse Fast Fourier Transform (FFT), because the visibility samples do not lie on a uniform grid, and our visibilities have a third term w. However, we can interpolate the measurements on a uniform grid, and then use the inverse FFT, which is known as the non-uniform FFT [8] in the literature. In the major/minor cycle architecture, we call the interpolation step ”gridding”.

The major cycle allows us to use approximations in the minor cycle, and in the Fourier transform.

The dirty image is handed over to the minor cycle, which finds the most likely deconvolved image. A deconvolution is still an ill-posed inverse problem. Indeed, if we ignore the error that is introduced by the gridding step, the deconvolution problem shown (**??**) is equivalent to the image reconstruction problem (1.3). Both objectives arrive at the same most-likely image x, but with a key difference: We replaced the big Fourier Transform matrix Fx with a convolution $x \star PSF$, which we can represent much more efficiently in practice. The deconvolved image is called the model image in the major/minor cycle architecture.

The major cycle is responsible for transforming the calibrated visibilities to an image. First, it takes the calibrated visibilities and interpolates them on a uniform grid. This is called the ”gridding” step. The gridding step lets us use the inverse Fast Fourier Transform (iFFT) next. At this point, we have approximated the Fourier Transform matrix F with the gridding and iFFT step. Since the radio interferometer measures an incomplete set of visibilities, the image after the iFFT is ”dirty”. The missing visibilities have the effect of convolving the observed image with a Point Spread Function PSF, resulting in the dirty image.

1.3.3 Approximations under the major cycle

Looking at the deconvolution problem (1.13) in isolation , we find a few hidden difficulties with the PSF.

For one, it spans the whole image and it is not constant over all pixels. DDE's like the w-term change the PSF for each pixel. Solving the deconvolution problem would require us to estimate the PSF for each pixel, which quickly becomes too expensive for larger images[3]. With the major cycle however, we can use an approximation of the true PSF. The error we introduce by approximating the PSF will get removed over several major cycles.

Typically, the minor cycle uses a constant PSF and cuts parts of it off, only using the most significant center. An example is shown in figure 4a and 4b. The figure 4a shows the PSF for the center pixel. It is approximately Gaussian shape in the center, but affects pixels over the whole image. Figure 4b shows the window which was used for deconvolution. It is half the size of the original PSF, and the minor cycle ignores all effects of the PSF that are further away.

he final reconstructed image is the sum of the model images over several major cycles. But this "backwards pass" is what allows us to: Reduce the error from interpolation and use a constant plus cropped PSF.

Interpolation error.

Figure 4 shows a deconvolution problem for the MeerKAT radio interferometer. The PSF for the center pixel is shown in figure 4a. The PSF is approximately a gaussian at the center, with only

(**??**), PSF is not constant. DDE like the w-term change the PSF for each pixel. But estimating a PSF is an expensive operation. In section 3 we show how the PSF is estimated in practice. Several approximations. The minor cycle does not need to calculate the exact deconvolution, and the Gridding step is also just an approximation.

PSF is not constant over the image. With the Major cycle, the minor cycle can approximate the deconvolution with a constant PSF. Estimating the PSF for one location is as expensive as half a major cycle. It is common to use a single estimation for the whole problem

Assuming we know the reconstructed image, the major cycle then just improves the gridding accuracy, similar to how [8]. Several major cycles gives us free improved accuracy

Alternatives to the major cycle

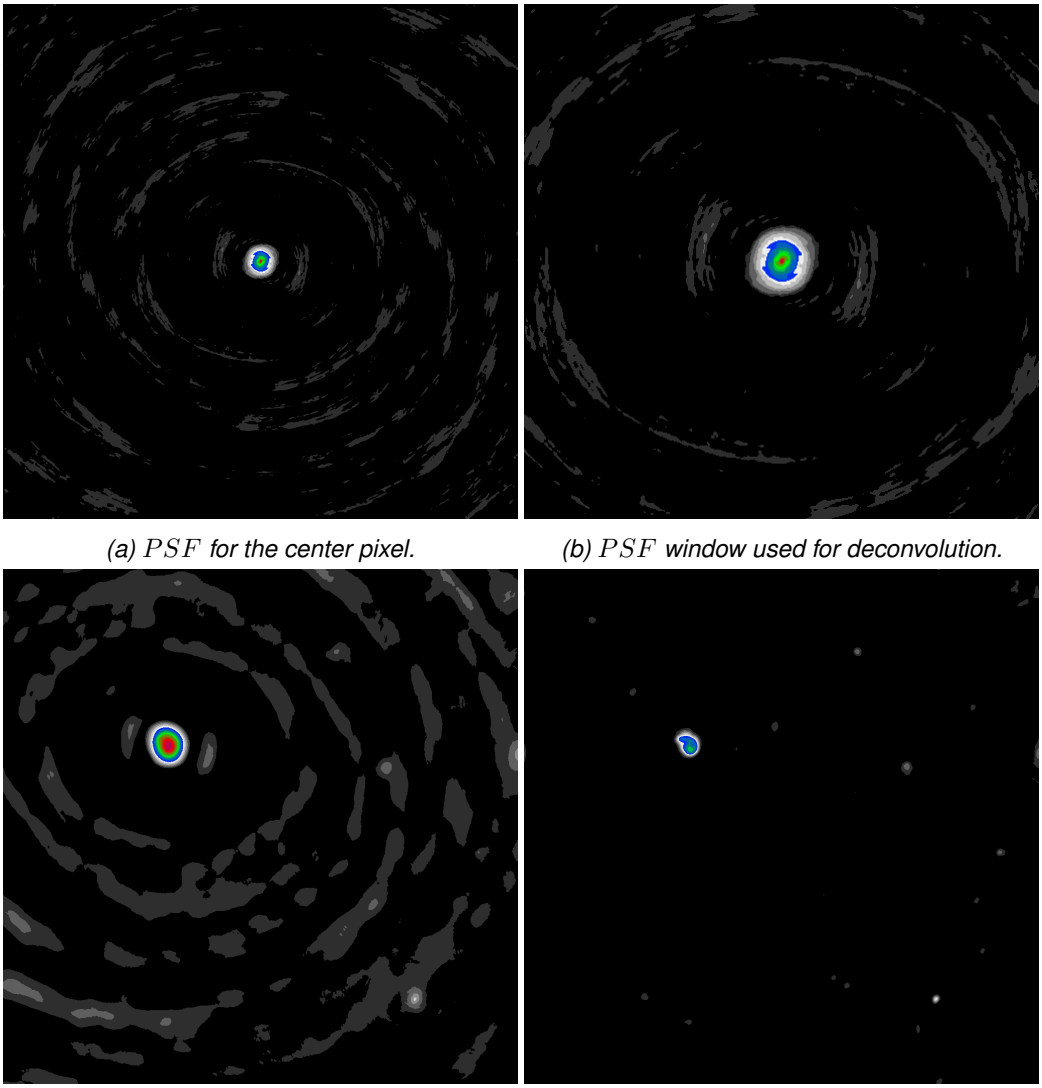
1.3.4 Alternatives to the major/minor cycle architecture

Let us first look at the Major Cycle of figure 1.3 in more detail. We can separate

The first Major Cycle takes the calibrated, non-uniformly sampled Visibilities and interpolates them on a uniform Grid. This lets us use the inverse FFT in the next step, and the Cycle arrives at the Dirty Image. At this point, the Minor Cycle takes over and deduces the most likely observed image, the Model Image. After the Minor Cycle, the Major Cycle continues with the Model Image. It calculating the FFT and the De-Gridding on the model image, and finishes with non-uniformly sampled Model Visibilities. The second Major Cycle subtracts the Model from the calibrated Visibilities and repeats the process with the residuals.

In the real world, several Major Cycles, between 2-6 are needed to reconstruct an image. At first glance, it is not clear why we cannot just use a single major cycle. The Minor Cycle minimizes the objective (1.7), it seems all we need is the PSF and the Dirty Image from the first Major Cycle. The problem is that the PSF of Radio Interferometers is not constant over the image. DDEs like the w-correction affect the PSF. In this architecture, we only approximate the deconvolution in the Minor Cycle.

[3] Calculating the PSF for a pixel requires a gridding and an iFFT step, or half a major cycle. We need around two to ten major cycles for an accurate reconstruction. Estimating a PSF for each pixel is not viable.



(a) PSF for the center pixel. *(b) PSF window used for deconvolution.*

(c) Dirty image. *(d) Reconstructed image after 5 major cycles.*

Figure 4: Deconvolution problem for the MeerKAT radio interferometer with a strong extended emission and weaker point sources in the image.

How many Minor Cycles, Major Cycles. We can use a fast approximation in the Minor Cycle. Minor Cycles are computationally cheap compared to the Major Cycle. For maximum speed, we would only want a single Major Cycle. But for the best reconstruction quality, we want as many as possible.

1.3.5 Minor Cycle: CLEAN Deconvolutions

The Minor Cycle was conceived with CLEAN in mind. CLEAN is state of the art algorithm

Contains two classes of objects: Point sources, which are essentially stars, and extended emissions, which span over several pixels.

Find the fainter sources in later iterations Because we can only estimate the psf.

2 State of the art Image Reconstruction

2.1 Gridder

2.1.1 -stacking

2.1.2 IDG

2.2 Reconstruction

2.2.1 CLEAN

2.2.2 SARA

3 Simple distributed image reconstruction

Show architecture

Different architectures, why don't we do consensus algorithms.

Message Passing Interface (MPI)

Gridding and Devonvolution

We use the recently developed Image Domain Gridder (IDG)

3.1 The Image Domain Gridder

What the gridding actually does. In theory, just an interpolation. Interpolation kernel like Spheroidal, Kaiser-Bessel functions etc. Generally the most intensive step, because the number of grid cells is generally a lot smaller than the number calibrated Visibilities.

However, here we move from the three dimensional Visibilities to two dimensions. DDE's fall down in the area of the gridder.

So how we do an interpolation

Veeneboer et al[1] developed the Image Domain Gridder. It uses Subgrids and solves each subgrid seperately. It is in the image domain, because it can do Radio Interferometer specific corrections efficiently. Furthermore, it leads to a structure which is primed for GPU processing. We use this algorithm to distribute the gridding.

W-Projection, Spheroidal are convolutions in the Fourier space.

The figure 5 shows the different parts of the image domain algorithm.

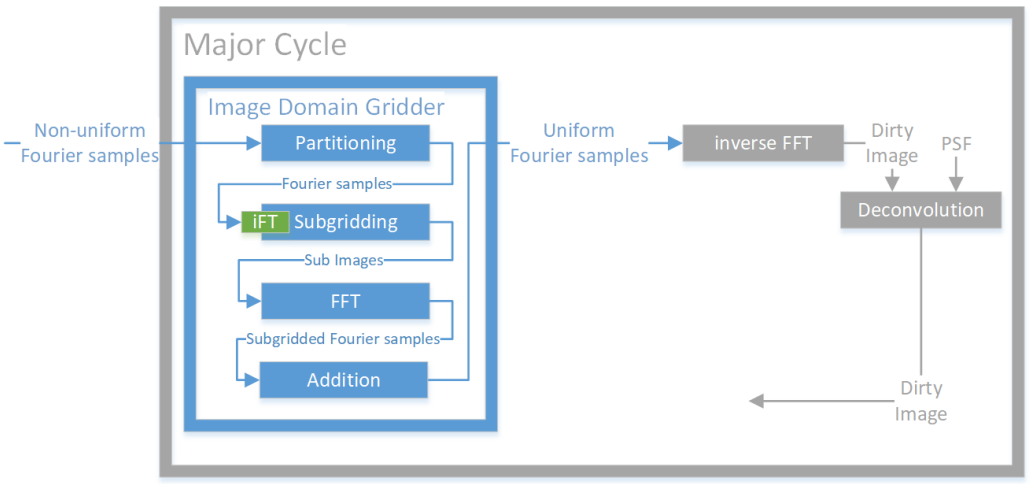


Figure 5: Image Domain Gridder in the Major Cycle Architecture

Algorithm

3.2 Coordinate Descent deconvolution

Coordinate Descent Algorithm why

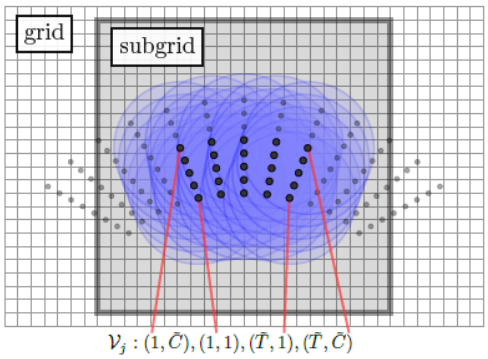


Figure 6: Subgrid

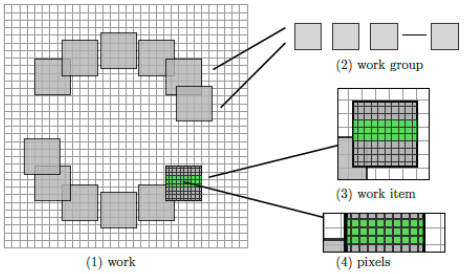


Figure 7: parallel

Variants

Implementation Correlate the dirty image Find max How to calculate a

3.2.1 ElasticNet Regularization

Why

Formula

Effect

Implementation

May even speed up convergence for correlated pixel values compared to L1 or L2[10]. 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 [10] 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

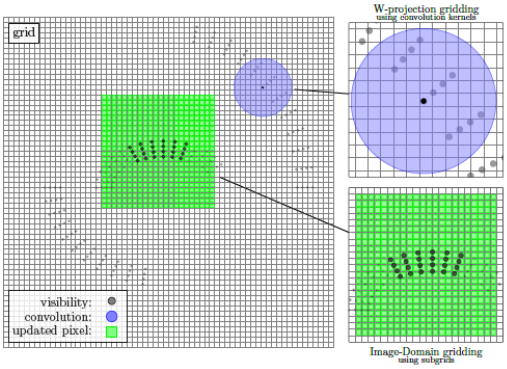
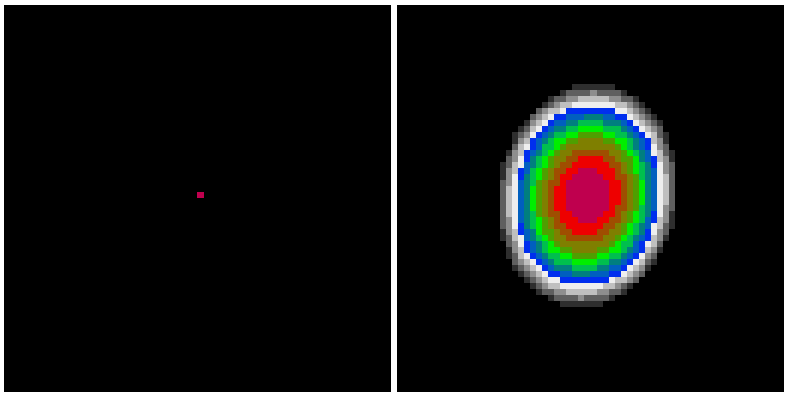


Figure 8: Image Domain Gridder in the Major Cycle Architecture



(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.

3.4 Test on MeerKAT data

4 Conclusion

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

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

[6] Lei Zhu, Yaolin Zhu, Huan Mao, and Meihua Gu. A new method for sparse signal denoising based on compressed sensing. In 2009 Second International Symposium on Knowledge Acquisition and Modeling, volume 1, pages 35–38. IEEE, 2009.

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

[8] Stefan Kunis, J Keiner, and D Potts. Nonequispaced fft. Available: www-user. tu-chemnitz. de/skunis/paper/lecturenfft. pdf.

[9] Gonzalo Mateos, Juan Andrés Bazerque, and Georgios B Giannakis. Distributed sparse linear regression. IEEE Transactions on Signal Processing, 58(10):5262–5276, 2010.

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

[11] 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

5 attachment

6 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.

6.1 CLEAN: The Major Cycle Architecture

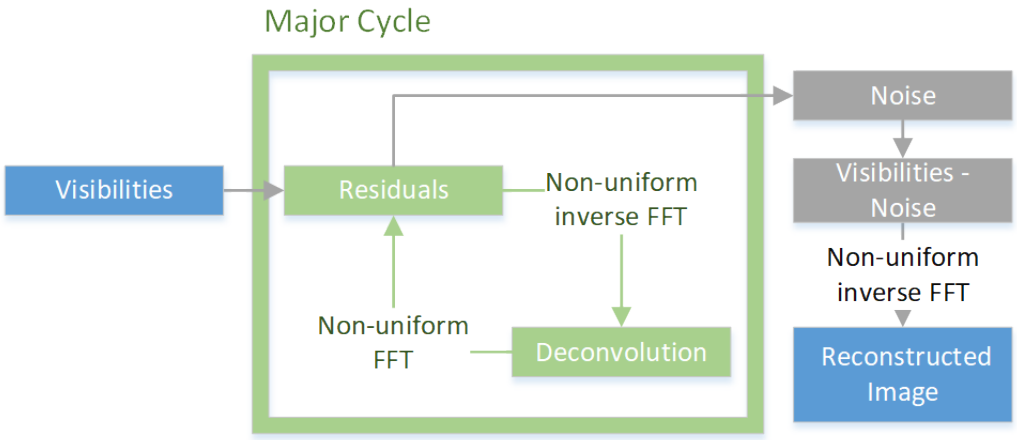


Figure 10: The Major Cycle Architecture

Figure 10 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.

6.2 Compressed Sensing Architecture

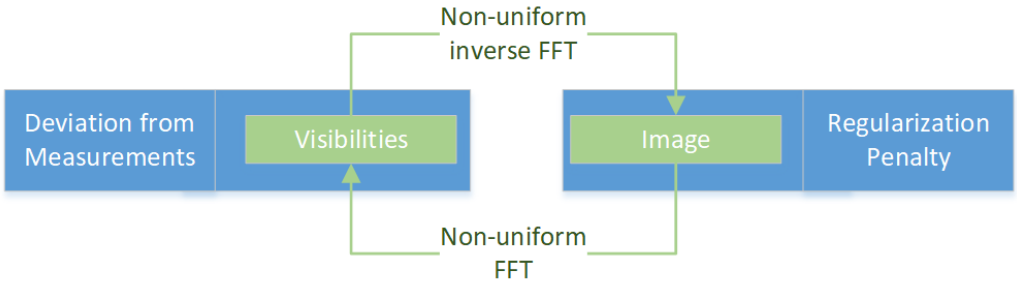


Figure 11: State-of-the-art Compressed Sensing Reconstruction Architecture

Figure 11 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.

6.3 Hypothesis for reducing costs of Compressed Sensing Algorithms

Compressed Sensing Algorithms are not bound to the Architecture presented in section 6.2. 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.

6.4 State of the art: WSCLEAN Software Package

6.4.1 W-Stacking Major Cycle

6.4.2 Deconvolution Algorithms

CLEAN MORESANE

6.5 Distributing the Image Reconstruction

6.5.1 Distributing the Non-uniform FFT

6.5.2 Distributing the Deconvolution

7 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[2, 3], continually reducing the runtime costs of the operation. Recently, Veeneboer et al[1] 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.

7.1 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.

8 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 12), which contains three operations: Gridding, FFT and Deconvolution.

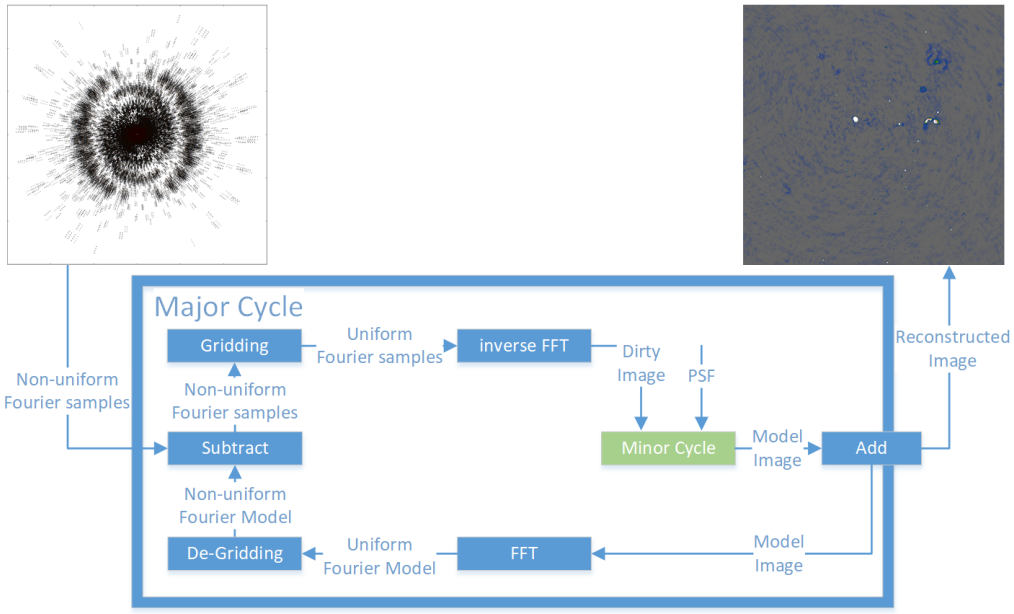


Figure 12: The Major Cycle Architecture of image reconstruction algorithms

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.

8.1 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[1] 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[11]. 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.

8.2 First steps towards a distributed Algorithm

In this project, we make the first steps towards a distributed Major Cycle architecture (shown in figure 12) 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.

9 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, June 14, 2019

Jonas Schwammberger