P9 Distributed Image Reconstruction for the new Radio Interferometers

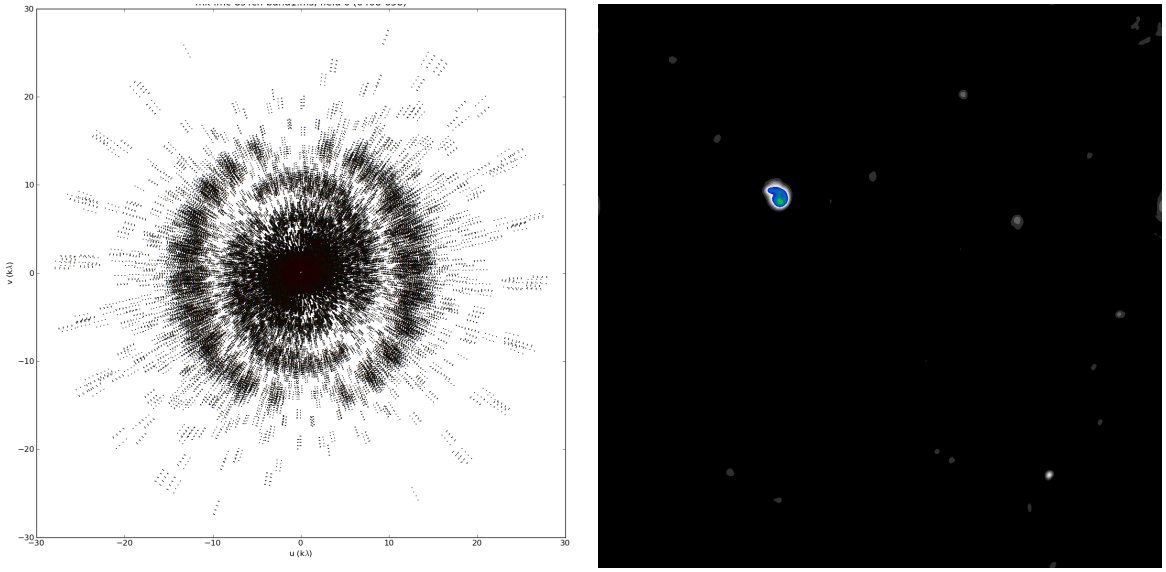
Jonas Schwammberger

August 9, 2019

Abstract

Contents

1 Introduction to radio interferometric imaging



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

Figure 1: Example of an image reconstruction for visibility measurements of the MeerKAT radio interferometer

The figure 4 gives an example of the radio interferometric imaging. We wish to reconstruct an image, shown in 4b from the measurements of the interferometer 4a. The reconstructed image contains two object types: Point sources and extended emissions. Point sources, stars for example, are far-away emissions that are concentrated in a single pixel. Extended emissions like hydrogen clouds emit radio waves over a large area of the sky and span over several pixels in the reconstructed image.

The measurements of the radio interferometer are in the Fourier space. It does not measure how much emission was observed at a single pixel in the sky image, it measures the amplitude and phase of a Fourier component of the sky image. This means each dot in the figure 4a represents one Fourier measurement (amplitude and phase) of the observed image. The dots in the center, measurements with a low uv-value, measure the low-frequency components of the sky image. They contain the information about structures spanning over several pixels. Dots further away from the center measure the high-frequency components, they contain the information about the edges of the image and are responsible for a ”sharp” image. In the radio astronomy literature, the measured Fourier components are called visibilities and for the sake of consistency, we will be calling the measured Fourier components visibilities in this work.

As we will see, we cannot retrieve the observed image by simply calculating the inverse Fourier transform. We require a reconstruction algorithm that finds the most likely observed image given the measurements. How close the most likely image is to the truly observed one depends on the reconstruction algorithm we choose. As it is often the case, how expensive an image reconstruction is in terms of runtime costs (Money spent on the necessary CPU’s, time spent waiting for the results etc.) also depends on the reconstruction algorithm. We often face a trade-off between reconstruction quality and runtime costs. This work focuses on distributing the radio interferometric image reconstruction. We search for an effective way to use distributed computing hardware and test how scalable the reconstruction algorithm is with a real-world observation of the MeerKAT radio interferometer.

Introduction to how a radio interferometer works What the fundamental problems are and how we can solve them in theory and what the implications are in practice how we can This section gives a short introduction to interferometry and show the fundamental problems in the radio interferometric image reconstruction. In section 1.2 we give an introduction how we can solve the radio interferometric image reconstruction in theory,

and what the implications are in practice. Section 1.3 introduces how the common architecture for radio interferometric image reconstruction, and how the reconstruction algorithm is divided into sub-tasks in practice.

1.1 From electromagnetic waves over visibilities to images

We give a short introduction into how the electromagnetic wave gets measured by the interferometer, turned into visibilities and finally processed into an image. Figure 3 shows a radio source and its electro-magnetic (e-m) wave arriving at the antennas of the interferometer. It then shows the three processes involved to arrive at an image: Correlation, calibration and image reconstruction.

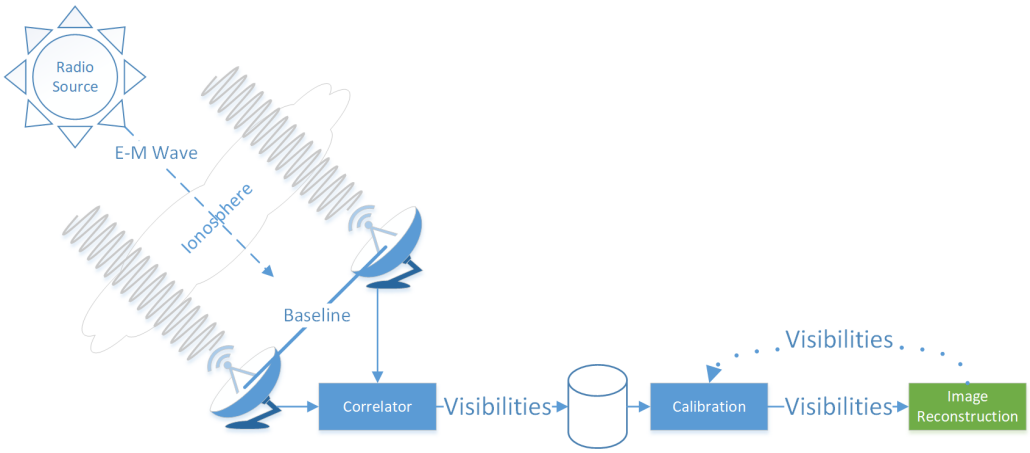


Figure 2: Radio interferometer system

First, we have a source in the sky that is emitting e-m waves in the radio frequency. The waves travel to earth, through the earth's ionosphere and finally to our interferometer. Along its path, the e-m waves may get distorted from various sources. For example, it may receive a phase shift by the ionosphere.

Then, the e-m wave arrives at our interferometer. We call each antenna pair a baseline. Each baseline will end up measuring a single visibility. The figure 3 shows a single baseline of the interferometer. The distance between the antennas and their orientation to the e-m wave will determine where we sample the uv-plane. Short baselines measure the uv-plane close at the origin, while long baselines sample the uv-space further away from the origin. Remember that the samples away from the uv-origin contain the information about edges and other details of our image. With a longer baseline the interferometer measures more highly resolved details, regardless of the antenna dish-diameter[1].

Not yet visibilities. This happens in the correlator Correlator, amplitude and phase of the e-m wave at a frequency. saved for further processing

Calibration. Effects from the ionosphere. Imperfections of the instrument, like varying antenna sensitivity. Calibration is complex, and not part of the project

Image reconstruction generally happens after calibration. The focus of this project. self-calibration

1.1.1 The measurement equation

As we discussed so far, the radio interferometer measures visibilities of the sky image, and we wish to find the observed image from the measurements. Put formally, we wish to invert the following system of linear

[1] Remember that this is the reason why we build radio interferometers. We do not need impossibly large dish diameters for a high angular resolution. We just need large distances between smaller antennas

equations (1.1), where V is the visibility vector[2], F is the Fourier transform matrix and I is the pixel vector of the observed image.

$$V = FI \tag{1.1}$$

We wish to find the observed image I, while we only know the visibility vector V and the Fourier transform matrix F. This is what we call the measurement equation. In most context for this project, looking is an adequate view of the image reconstruction problem. We will show why we cannot find the observed image I by simply calculating the inverse Fourier transform. However, when we need to efficiently apply the Fourier transform, we need to know F in more detail. As we will see, radio interferometers have some difficulties hidden in the Fourier transform matrix, which are difficult to handle efficiently. First, let us abandon the vector notation of (1.1), and represent the measurement equation with integrals (1.2).

$$V(u,v) = \int\int I(l,m)e^{2\pi i[ul+vm]}\,dl\,dm \tag{1.2}$$

This is essentially the same problem. The main difference is that we do not represent the Fourier transform as a matrix F, but as integrals $\int\int e^{2\pi i[ul+vm]}$, where u, v are the coordinates in Fourier space and l, m are the angles away from the image center. A single pixel represents the intensity of the radio emission from the direction l, m. Note that the measurement equation (1.2) shows the fact that the visibilities are measured in a continuous Fourier space. If the Fourier space would also be discrete, we could replace the integrals with sums.

However, the measurement equation (1.2) is inaccurate in the sense that it ignores many effects that distort the signal. For example, it does not account for the distortion by the ionosphere, or the distortion introduced by real-world antennas. The measurement equation (1.2) shown here does not represent the real world. But depending on the instrument and the observation, these distortions may be negligible, and the measurement equation (1.2) is a good approximation.

When there is a distortion source that cannot be ignored, it has to be modelled in the measurement equation. As such there is no unified measurement equation for all radio interferometric observations, let alone radio interferometers. The equation shown in (1.2) can be seen as the basis that gets extended as necessary[1, 2, 3, 4].

For example, the measurement equation (1.2) is only accurate for small field of view observations, when l and m are both small angles. For wide field of view observations, we need to account for the fact that the visibilities have a third term w, and we arrive at the wide field of view measurement equation (1.3).

$$V(u,v,w) = \int\int \frac{I(l,m)}{c(l,m)} e^{2\pi i[ul+vm+w(c(x,y)-1)]}\,dl\,dm\,, \quad c(l,m) = \sqrt{1-l^2-m^2} \tag{1.3}$$

The third w-term has two effects on the measurement equation. It introduces a phase shift in the Fourier transform $e^{2\pi i[\ldots+w(c(l,m)-1)]}$, and a normalization factor of the image $\frac{I(l,m)}{c(l,m)}$. Note that when the angles are small, i.e. $l^2 + m^2 \ll 1$ then the wide field of view measurement equation (1.3) reduces to our original (1.2). This is another way of saying that for small field of views, the measurement equation (1.2) is a good approximation under the right conditions.

In this project, we use the wide field of view measurement equation (1.3). But as we mentioned in the beginning of this section, for most contexts, it is not important whether we ignore the w-term of the visibilities or not. It is important when we design an efficient implementations for applying the wide field of view Fourier

[2]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.

transform, because the w-term keeps us from using the Fast Fourier Transform (FFT). In every other case, we can ignore this technicality. Because even more complicated measurement equation still have a linear relationship between visibilities and image [1, 2, 3, 4]. We can view the whole reconstruction problem as a system of linear equations (1.1), where the matrix F takes care of how exactly the measurements and pixels relate in this case.

1.2 The ill-posed image reconstruction problem

So far, we discussed how the interferometer measures in Fourier space, and we wish to find the observed image that matches the measurements. In other words, we wish to find a solution to a system of linear equation (1.4), where V are the measurements, x is the observed image and F is the Fourier transform matrix. We also discussed that F can be complicated in practice, but is still essentially a linear operator. Meaning we know how the inverse Fourier transform matrix F^{-1}, and the question arises: Why can't we solve the equations (1.4) by calculating the inverse Fourier transform? Or, why does $x = F^{-1}V$ not lead to the observed image?

$$\underset{x}{solve} \quad V - Fx = 0 \tag{1.4}$$

The answer is, the equations (1.4) do not have a unique solution, which makes the problem ill-posed. 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).

From linear algebra, we know that an under-determined system of linear equations, i.e. when (1.4) has more pixels than visibility measurements, then the problem is under-determined and there may be a potentially infinite number of solutions to the system. Under-determined systems arise in many similar fields, as for example in X-Ray imaging of the sun[5]. However, radio interferometers measure a large number of visibilities. We generally have more visibilities than pixels. This means the image reconstruction problem (1.4) for radio interferometers is actually over-determined.

From linear algebra, we know that an over-determined system either has one or zero solutions. At first glance it may be counter-intuitive why there are many possible solutions to (1.4) for radio interferometers. The reason why lies in two properties of the measurements: They are noisy and incomplete.

The radio interferometer measures noisy visibilities, meaning each amplitude and phase of a measurement is influenced by an unknown noise factor. Finding a reconstructed image is the same as finding the de-noised versions of the visibility measurements. This alone would make the problem ill-posed, but the visibilities actually have a second property that makes them ill-posed: incompleteness.

When we look back at figure 4a, 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 frequency visibilities get fewer and fewer samples. This means we are missing data for crucial measurements for the reconstruction.

Also note that the question whether the measurements are incomplete essentially comes down to the image resolution of the reconstruction: Since we are missing high-frequency measurements and we can choose the resolution of the image, we can also reduce the resolution of the reconstructed image until the missing frequencies become negligible. However, if we can solve the ill-posed inverse problem, we can reconstruct

an image at a higher resolution from the same measurements. The question is how do we solve the ill-posed inverse problem?

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 (2.2) 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 about the image, and therefore search the 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.

$$\underset{x}{minimize}\ \|V - Fx\|_2^2 + \lambda P(x) \tag{1.5}$$

The objective (2.3) 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 [6].

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 ($\|\nabla x\|_1$) [7], L2 ($\|x\|_2$) [8] or the L1 norm in a wavelet space ($\|\Psi x\|_1$) [9].

Finally, an optimization algorithm is necessary which can optimize the objective function (2.3) with the chosen prior function $P()$. Typical choices for optimization algorithms in image reconstructions are Interior Point Methods like Simplex, Matching Pursuit [10] and ADMM[11].

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, we limit ourselves to the objective (2.3) and explore how we can distribute the reconstruction problem.

The last question that remains is, how close are the most likely image, under a given prior, to the truly observed one? Remember that the Nyquist-Shannon sampling theorem states that our uniform sampling frequency needs to be larger than twice the highest frequency in a band-limited signal[3], and then the theorem guarantees exact reconstruction. For radio astronomy, we do not have uniformly sampled visibilities, and although we have a large number of samples, we are missing crucial parts of the Fourier space. Luckily, there

[3]For example: if we want to record human voices with the highest frequency of 20 kHz, Nyquist-Shannon states our uniform sampling frequency has to be larger than 40 kHz to guarantee exact reconstruction

is another sampling theorem that, under certain assumptions, guarantees exact reconstruction for the case of radio interferometers. Namely, the theory of compressed sensing.

1.2.2 Compressive sampling of the sky

For the sake of demonstration, let us assume the radio interferometer observes a patch of sky containing ten stars. It measures an incomplete set of random Fourier components of the ten stars, and we would like to reconstruct an image of size 256^2 pixels. The emissions from stars are concentrated into a single pixel. For compressive sampling, we need to know an space in which our reconstructed image is sparse, and we need to take measurements in a different, incoherent space.

Our reconstructed image contains only zero pixels except at the ten locations of the stars, meaning the image space for this patch of the sky is already sparse. In this case, we do not need any additional sparse space like wavelets. We can reconstruct directly in the sparse image space, and we have the first requirement met for compressive sampling.

The next requirement is that the measurement and reconstruction space (which is the image in this example), are as incoherent as possible.

Yes, incoherence.

$$\underset{x}{minimize}\ \|V - Fx\|_2^2 + \lambda \|x\|_1 \tag{1.6}$$

Intuitively, the number of samples depend on the information content, not on the bandwidth. thi

At what point are we guaranteed? The matrix A needs to fullfil the Restricted Isometry Porperty (RIP) [12, 13]. 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[14]. Approximations are also difficult to compute[15]. So we can only talk about how likely a given matrix fullfils the RIP. Matrix where each element is sampled from a random Gaussian distributîon has the highest change.

Random samples in the Fourier space also have a high chance to fulfill the RIP [16].

Not possible for Radio interferometers, because they sample in the Fourier space. Not random.

There are also extended emissions, clouds etc. Resulting in a lot of non-zero pixels. There may be better sparse spaces for radio astronomy.

But also RIP may be too strict. Exact reconstruction can also work under less strict conditions[17] active field of research.

1.2.3 Reconstruction guarantees in the real world

What does this all mean for image reconstruction? For us, we design reconstruction algorithms and cannot influence the matrix F. We do not know if it actually

The theory of compressed sensing gives us a framework. There is a data modelling task, and a task for finding efficient algorithms.

So there is a data modelling task in finding a good sparse prior. We also do not know the proper sparse space in which radio interferometric images. We know several spaces, Curvelets [18] Starlets [19], Daubechies wavelets [20]. As of the time of writing, it is currently unknown which leads to the best reconstruction. We can also learn dictionaries.

The task for efficient algorithms is finding the best way to optimize the objective with a given prior on real hardware. Fast convergence. Also difficult, because in practice the choice of prior can also influence convergence.

Compressed sensing based reconstruction algorithm. In radio astronomy.

1.3 Introduction into optimization/RI reconstruction algorithms

1.3.1 The Major/Minor cycle

1.3.2 Image reconstruction as deconvolution

2 Introduction

2.1 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 3 and how Radio Interferometers work in principle: The antennas observe the arriving electromagnetic wave, gets processed in three steps, correlation, calibration and image reconstruction.

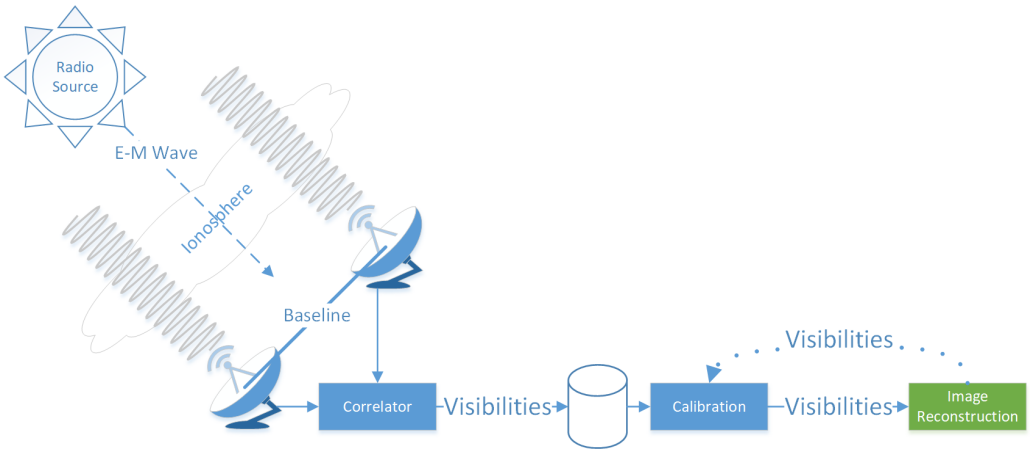


Figure 3: Radio interferometer system

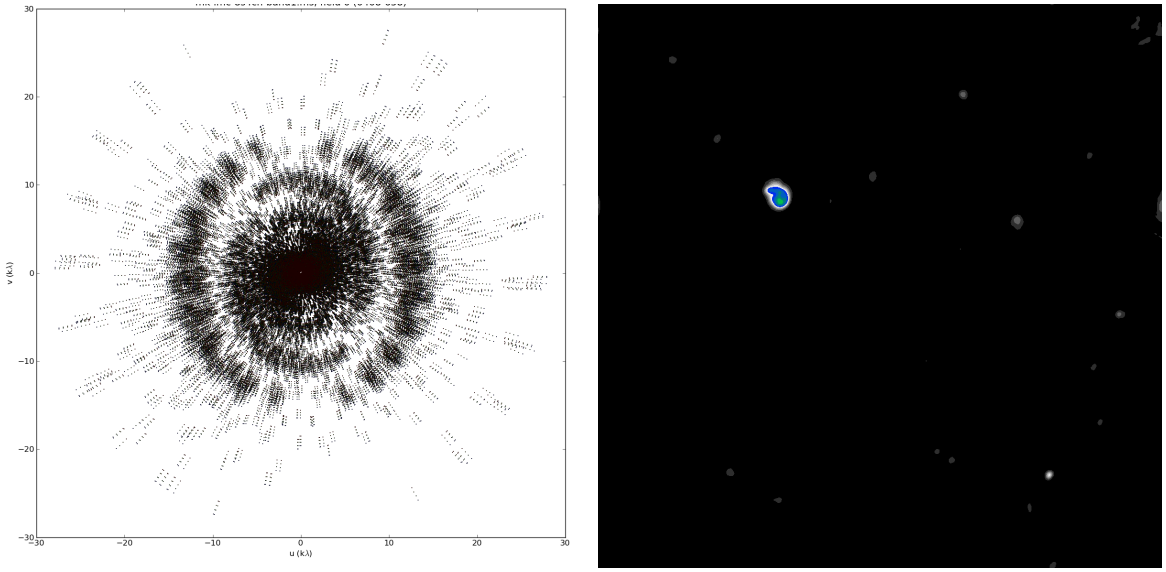
First, the electromagnetic wave gets measured by the different antennas of the interferometer. The measurements of each antenna pair get correlated into a Fourier component (called Visibility in Radio Astronomy). Each antenna pair measures a complex-valued, noisy visibility of the sky. The distance and 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 4a 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 4 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.

2.1.1 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.



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

Figure 4: Example of an image reconstruction for visibility measurements of the MeerKAT radio interferometer

In this project we use the measurement equation (2.1). It consists of three parts: The visibility measurements $V(u, v, w)$[4], 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 space, 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 4a shows an example for $V()$, while 4b shows an example for $I()$.

$$V(u, v, w) = \int\int \frac{I(x, y)}{c(x, y)} e^{2\pi i[ux+vy+w(c(x,y)-1)]}\, dx\, dy\,, \quad c(x, y) = \sqrt{1 - x^2 - y^2} \tag{2.1}$$

The radio interferometer essentially observes the sky in the Fourier space. 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 (2.1). Research in this area tries to use approximations which lets us use faster algorithms like the FFT, and correct for DDE's accurately enough [21, 22, 23]. In this project, the w-correction is the only DDE we handle.

2.2 The reconstruction problem as a system of linear equations

Even though the Fourier Transform in the measurement equation (2.1) contains a w-correction factor, it is still linear. There exist more complex measurement equations, but the relationship often stays linear in nature [1, 2, 3, 4]. This means we can represent the measurement equation as a system of linear equations (2.2),

[4] V in the equation (2.2) 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.

where F is the Fourier Transform with w-correction, x is the image we are searching, V are the calibrated visibilities. For the sake of demonstration, we ignore the noise in the visibility measurements in (2.2). To reconstruct the image x, we simply need to search for a solution to (2.2).

$$\underset{x}{find} \quad V - Fx = 0 \tag{2.2}$$

Even though we generally have more visibilities than pixels in the reconstruction, which makes (2.2) 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. 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 (2.2). 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[5]. 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 (2.2) 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 4a, 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.

2.2.1 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 (2.2) 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.

$$\underset{x}{minimize}\ \|V - Fx\|_2^2 + \lambda P(x) \tag{2.3}$$

The objective (2.3) 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 [6].

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 de-noising applications. Such as: Total Variation ($\|\nabla x\|_1$) [7], L2 ($\|x\|_2$) [8] or the L1 norm in a wavelet space ($\|\Psi x\|_1$) [9].

Finally, an optimization algorithm is necessary which can optimize the objective function (2.3) with the chosen prior function $P()$. Typical choices for optimization algorithms in image reconstructions are Interior Point Methods like Simplex, Matching Pursuit [10] and ADMM[11].

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, we limit ourselves to the objective (2.3) and explore how we can distribute the reconstruction problem.

The last question that remains is, how close are the most likely image, under a given prior, to the truly observed one? Remember that the Nyquist-Shannon sampling theorem states that our uniform sampling frequency needs to be larger than twice the highest frequency in a band-limited signal[5], and then the theorem guarantees exact reconstruction. For radio astronomy, we do not have uniformly sampled visibilities, and although we have a large number of samples, we are missing crucial parts of the Fourier space. Luckily, there is another sampling theorem that, under certain assumptions, guarantees exact reconstruction for the case of radio interferometers. Namely, the theory of compressed sensing.

2.2.2 Theory of compressed sensing

The theory of compressed sensing[12, 13] is a recently developed sampling theorem. It guarantees exact reconstruction even when we have fewer samples than necessary according to the Nyquist-Shannon theorem. The compressed sensing sampling theorem has wide ranging implications and was successfully applied in various signal processing tasks like de-noising[24], compression[25] and reconstruction[7]. This part gives a short introduction on the theory of compressed sensing. It tries to give an intuition on why we it is even possible to find the truly observed image at the optimum of the function (2.3), and why the image reconstruction can benefits from measuring in Fourier space.

The assumption underlying compressed sensing is that most natural signal are sparse in some space, only requiring a few non-zero components for representation. Empirically, this is true for many different signals. For example natural images are sparse in a the wavelet space. We only require a few non-zero components to represent the image with wavelets. This fact can obviously be exploited for image compression, where algorithms like JPEG2000 only need to save the non-zero wavelet components instead of all the pixel values.

[5]For example: if we want to record human voices with the highest frequency of 20 kHz, Nyquist-Shannon states our uniform sampling frequency has to be larger than 40 kHz to guarantee exact reconstruction

Less obvious is the application in image de-noising. Random noise in the image tends to affect all wavelet components approximately equal. Most noisy wavelet components are close to zero except the original components, which means we can de-noise the image by finding the most significant wavelet components. This is a so called "sparse prior" for de-noising ($\|\Psi x\|_1$)[6], which in practice is the L1 norm in some space where our signal is sparse.

When we assume our image is sparse in some known space, taking as many samples as required by the Nyquist-Shannon theorem seems wasteful. We acquire a large amount of data just to then save the most significant sparse components. Can we measure the sparse component directly? The answer is yes, with a compressive sampling regime. In this regime, we need to know a space where our reconstructed image is sparse, and we take measurements in a different, incoherent space. We now show an example of how compressive sampling works with an idealized radio interferometer together with an idealized signal.

2.2.3 Compressive sampling of the sky

For the sake of demonstration, let us assume the radio interferometer observes a patch of sky containing ten stars. It measures an incomplete set of random Fourier components of the ten stars, and we would like to reconstruct an image of size 256^2 pixels. The emissions from stars are concentrated into a single pixel. For compressive sampling, we need to know an space in which our reconstructed image is sparse, and we need to take measurements in a different, incoherent space.

Our reconstructed image contains only zero pixels except at the ten locations of the stars, meaning the image space for this patch of the sky is already sparse. In this case, we do not need any additional sparse space like wavelets. We can reconstruct directly in the sparse image space, and we have the first requirement met for compressive sampling.

The next requirement is that the measurement and reconstruction space (which is the image in this example), are as incoherent as possible.

Yes, incoherence.

$$\underset{x}{minimize}\ \|V - Fx\|_2^2 + \lambda \|x\|_1 \tag{2.4}$$

Intuitively, this can work.

At what point are we guaranteed? The matrix A needs to fullfil the Restricted Isometry Porperty (RIP) [12, 13]. 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[14]. Approximations are also difficult to compute[15]. So we can only talk about how likely a given matrix fullfils the RIP. Matrix where each element is sampled from a random Gaussian distributîon has the highest change.

Random samples in the Fourier space also have a high chance to fulfill the RIP [16].

Not possible for Radio interferometers, because they sample in the Fourier space. Not random.

There are also extended emissions, clouds etc. Resulting in a lot of non-zero pixels. There may be better sparse spaces for radio astronomy.

But also RIP may be too strict. Exact reconstruction can also work under less strict conditions[17] active field of research.

[6] The true sparse prior is actually not the L1 norm, but the indicator function of the wavelet components ($ind(\Psi x)$). But plugging this prior into the objective function (2.3) leads to a difficult objective to optimize, because the indicator function is both non-convex and non-differentiable. But the L1 norm is convex and is virtually guaranteed to lead to the same result in practice[26].

2.2.4 Reconstruction guarantees in the real world

What does this all mean for image reconstruction? For us, we design reconstruction algorithms and cannot influence the matrix F. We do not know if it actually

The theory of compressed sensing gives us a framework. There is a data modelling task, and a task for finding efficient algorithms.

So there is a data modelling task in finding a good sparse prior. We also do not know the proper sparse space in which radio interferometric images. We know several spaces, Curvelets [18] Starlets [19], Daubechies wavelets [20]. As of the time of writing, it is currently unknown which leads to the best reconstruction. We can also learn dictionaries.

The task for efficient algorithms is finding the best way to optimize the objective with a given prior on real hardware. Fast convergence. Also difficult, because in practice the choice of prior can also influence convergence.

Compressed sensing based reconstruction algorithm. In radio astronomy.

2.3 Fast image reconstruction in practice

We know how to solve the ill-posed image reconstruction problem in theory. We formulate a minimization problem (2.3), 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 (2.3). 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 2.3.2, and discuss the major/minor cycle architecture in more details.

2.3.1 The major/minor cycle

The major/minor cycle architecture shown in figure 5 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 (2.7)). 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

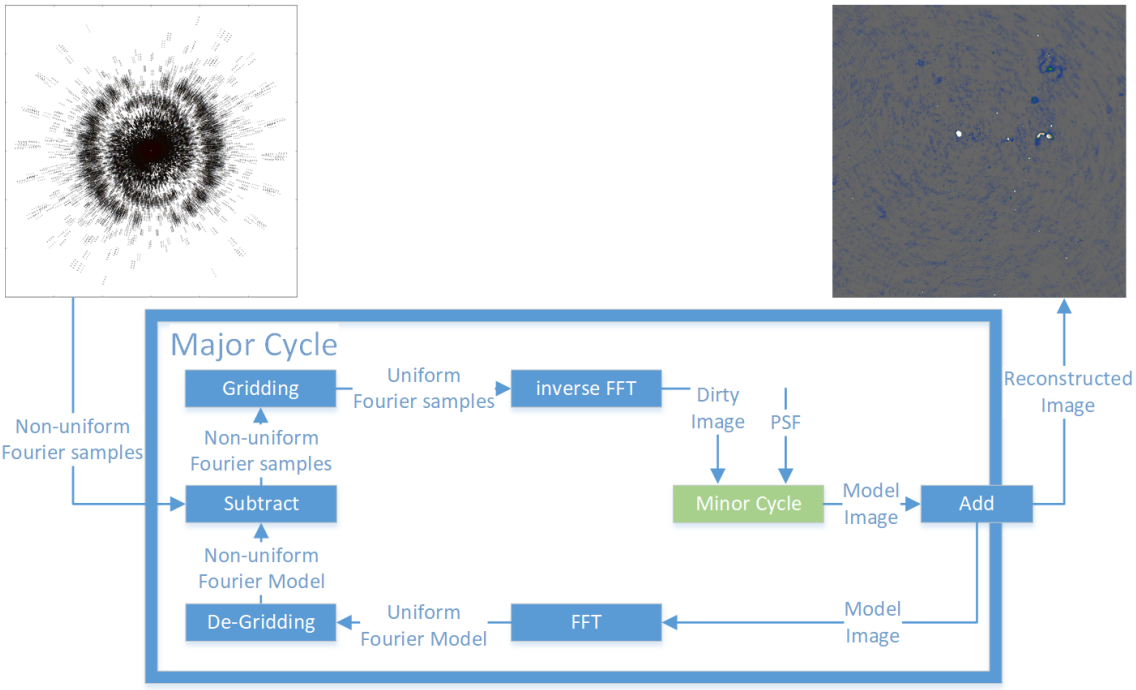


Figure 5: The Major/Minor Cycle Architecture

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.

2.3.2 Reformulating as a deconvolution problem

The Fourier transform matrix F in (2.3) 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 (2.3). 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 (2.5). This is identical to (2.3), except for the factorization of F.

$$\text{original:} \quad \underset{x}{minimize} \ \|V - MFx\|_2^2 + \lambda P(x) \tag{2.5}$$

$$\text{in-painting:} \quad \underset{V_2}{minimize} \ \|V - MV_2\|_2^2 + \lambda P(F^{-1}V_2) \tag{2.6}$$

$$\text{deconvolution:} \quad \underset{x}{minimize} \ \|I_{dirty} - x * PSF\|_2^2 + \lambda P(x) \tag{2.7}$$

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 (2.6), 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 (2.7).

Since the deconvolution formulation is vital for the major/minor cycle architecture, we have a closer look at (2.7). 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 (2.7) and the original image reconstruction problem (2.3) 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.

2.3.3 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.

2.3.4 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.

3 State of the art image reconstruction

This section introduces the state of the art image reconstruction algorithm in radio astronomy.

Two software packages, the Common Astronomy Software Applications (CASA)[27] and WSCLEAN[22] use the Major/Minor cycle architecture introduced in section 1.3.1. They have been the de-facto standard for decades and to our knowledge, are still the two most widely used implementations of image reconstruction algorithms. However, there has also been recent development to abandon the Major/Minor cycle architecture [28, 29]. They still use a gridding algorithm, but the reconstruction it does not contain Minor cycle deconvolution. At this point, it is not known whether the Major/Minor cycle architecture will be used in the future.

In this section, we introduce the

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

3.1 Gridding algorithms

Biggest part is w-term, how to handle it efficiently.

(show the problem of w-component)

Accuracy and speed.

Faceting.

w-projection algorithm [30] Convolution in the Fourier domain.

The WSCLEAN [22] did a large part.

More recently, the Image domain gridding algorithm has been developed [21]. Which can put it on the gpu, along with handling more DDE's.

3.1.1 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

3.1.2 Image Domain Gridder

Recently developed[21]. 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

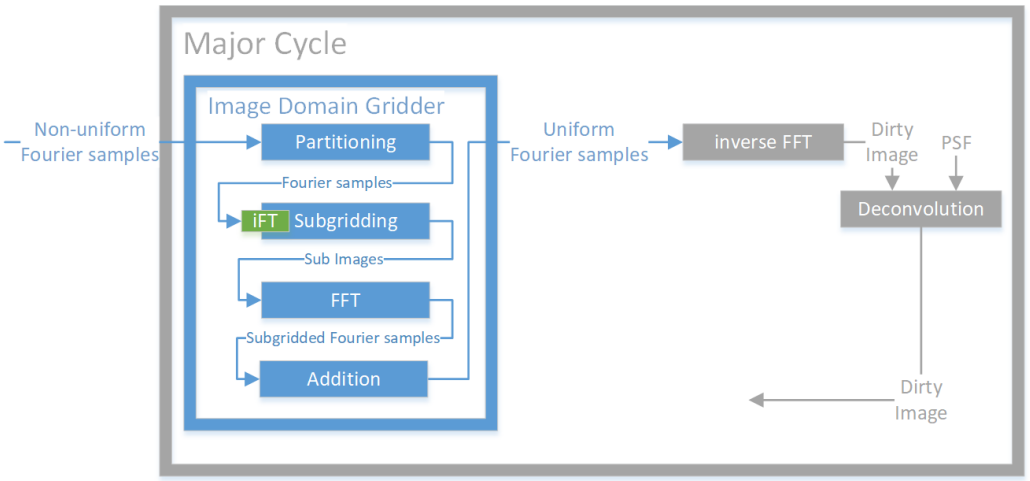


Figure 6: Image Domain Gridder in the Major Cycle Architecture

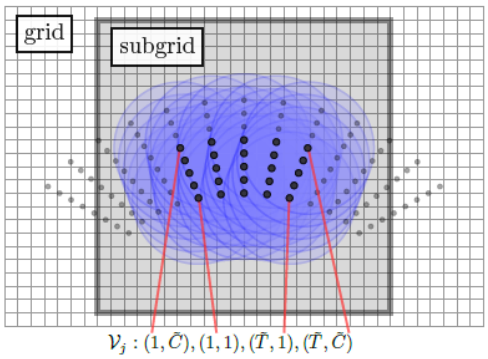


Figure 7: Subgrid

3.2 Deconvolution algorithms

3.2.1 MS-MFS-CLEAN

CLEAN basic [10].

Latest variants for multiscale and multi frequency CLEAN (MS-MFS-CLEAN)[31].

Current state-of-the-art

3.2.2 MORESANE

MORESANE [32]

3.3 Reconstruction algorithms which are not based on the deconvolution

SARA[29].

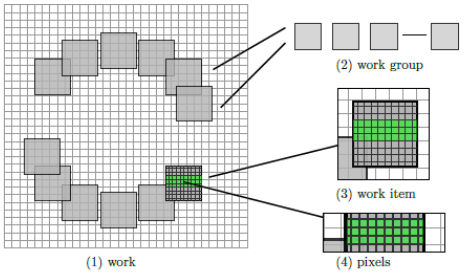


Figure 8: parallel

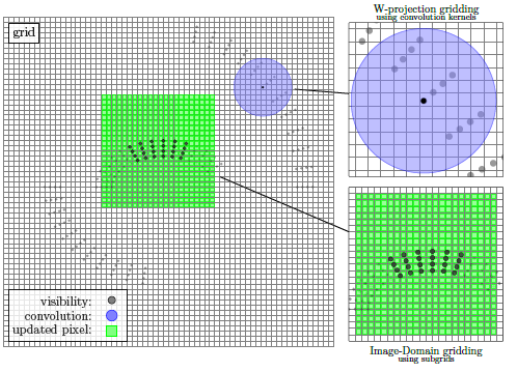


Figure 9: Image Domain Gridder in the Major Cycle Architecture

4 Simple distributed image reconstruction

How do we distribute the major cycle. We need to distribute every step, Gridding, FFT and Deconvolution.

Gridding, Large number of input data. This needs to be distributed We use the Image domain gridding introduces in section 3.1.2 and use it as the basis for the distributed gridding.

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. (cite distributed fftw)

Deconvolution is also worth distributing. CLEAN depending on the observation is the second most time consuming step. But gridding tends to be easier to distribute, so in some observations it is the most time consuming step. Split the image into patches and deconvolve each patch. Sadly not possible, we need communication. how we communicate is important.

We use a distributed Gridding and a distributed deconvolution. Which leads us to the following architecture.

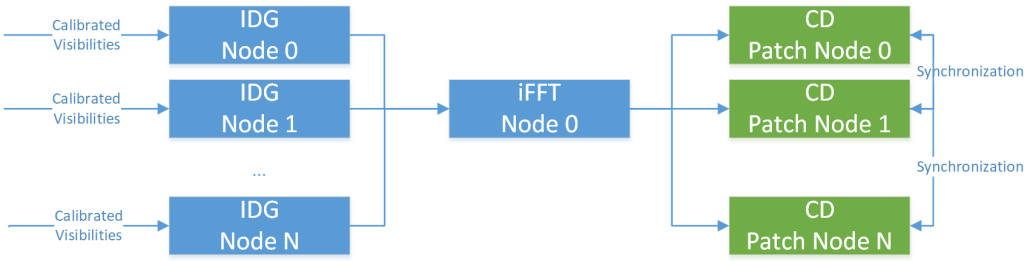


Figure 10: Distributed architecture for half a major cycle

Where each node is one computer, i.e. has its own, possibly multiple cpus and its shared memory. Split the input visibilities onto nodes. Do the gridding locally on each node Communicate the grid inverse FFT on one node. Communicate the patches of the image. Deconvolve each patch and communicate

How we distribute the IDG algorithm, and create a distributed deconvolution algorithm. Main contribution is the distributed deconvolution algorithm

4.1 Distributing the IDG algorithm

4.2 Distributed deconvolution

Main contribution We start

What we need: An optimization objective, an optimization algorithm and a regularization.

We use (4.1) as our optimization objective with data fidelity term and regularization term. We use Coordinate Descent Methods as our optimization algorithm, which we introduce in section 4.2.1, and the ElasticNet regularization, which we discuss in section 4.2.3

$$\underset{x}{minimize}\ \|I_{dirty} - x * PSF\|_2^2 + \lambda P(x) \tag{4.1}$$

4.2.1 Coordinate Descent Method

Coordinate Descent has gained more interest, because it may exploit sparse structures in the data. Optimization for large data volume

Most problems become simple when we look at them in the one dimensional case. Finding the optimum to our objective (4.1) is hard. But when we try to find the optimum of a single pixel, independent of all other pixels, the problem becomes simple. As we will see, we can find a closed form solution to find the optimum of a single pixel. Coordinate Descent then optimizes a single pixel at a time, and we iterate over the whole image with some strategy until we converge. We may need to optimize the same pixel several times.

Monotone decreasing. Robust Many different variants. How fast it is in practice depends on the regularization $P(x)$ we use. Outperforms other algorithms like Gradient based methods, FISTA and ADMM if a single iteration (minimizing a single pixel in our case) is ”cheap” to compute.

```
1  dirty = IFFT(Gridding(visibilities))
2  residuals = dirty
3
4  x = new Array
5  objectiveValue = SUM(residuals * residuals) + P(x)
6  oldObjectiveValue = objectiveValue
7
8  do
9  {
10   oldObjectiveValue = objectiveValue
11
12   //the core of the algorithm
13   pixelLocation = IterationStrategy(residuals)
14   optimalValue = Minimize(residuals, pixelLocation)
15   optimalValue = ApplyElasticNet(optimalValue)
16
17   //housekeeping
```

```
    x[pixelLocation] = optimalValue
    residuals = (dirty − Convolve(x, psf))
    objectiveValue = SUM(residuals ∗ residuals) + ElasticNet(x)
} while (oldObjectiveValue − objectiveValue) < epsilon
```

The core of the algorithm consists of three functions: $ApplyElasticNet()$ which applies the closed form solution of the ElasticNet regularization (discussed in section 4.2.3), the $Minimize()$ function which minimizes a single pixel at a time and the iteration strategy. We derive a closed form solution for minimizing a single pixel independently of each other and then discuss the basics of the iteration strategy.

For the closed form solution, we exclude the regularization for a moment and focus on the data fidelity term $\|I_{dirty} - x * PSF\|_2^2$. As we will see later, the ElasticNet regularization we chose also has a closed form solution and it does not influence the data fidelity term in our case. Let us start by writing the data fidelity term as a quadratic equation: $(I_{dirty} - pixel * PSF)^2$, where $pixel$ is a scalar. Further, let us assume we only have a single pixel in the dirty image I_{dirty} and a single pixel in the PSF. Everything is a scalar, it is easy to see that we are dealing with a parabola ($a * x^2 + b * x + c$. When we use our variable names we get $PSF^2 * pixel^2 - 2 * (I_{dirty} * PSF) * pixel + (I_{dirty})^2$). We have a closed form solution to find the optimum of a parabola, i.e. $x_{opt} = \frac{-b}{2a}$, or when we use our variable names explicitly, we get $x_{opt} = \frac{(pixel*PSF)*I_{dirty}}{PSF^2}$. If we are dealing with more than one pixel in the dirty image I_{dirty} and PSF, we multiply element-wise and sum up all the dimensions (i.e. $b = -2SUM(pixel * PSF * I_{dirty})$). We found our closed form solution in the one dimensional case and we can implement $Minimize()$.

The next point is the iteration strategy. This is where Coordinate Descent methods are robust. They are proven to converge by simply choosing randomly. Or a greedy strategy where the pixel which reduces the objective (4.1) by the largest amount. It was proven to converge for a variety of strategies. For this project, we are interested in which takes fewer computing resources to converge. Start with a greedy scheme, well known convergence guarantees at least linearly (with respect to iteration count?)[33] Cyclic scheme.

4.2.2 Efficient Coordinate Descent implementation

The question of efficient implementation. In the previous section, we have seen how coordinate descent works in general. But now we have to look at the details and see how we can implement it efficiently.

Question of convolution scheme. Circular convolution is physically not possible. The PSF of the lower left corner does not wrap around the image and influence all other corners of the image. But as we will see, for an efficient coordinate descent we need to calculate a convolution to cache. The circular convolution is the output of the FFT. This is why reconstruction algorithms sometimes choose to use circular convolution [8].

The other convolution scheme, namely zero padding. We use this scheme. It results

Iteration scheme. For a greedy iteration scheme, we would need to first:

1. find the optimum pixel value for each pixel independently (i.e. $\frac{-b}{2a}$)
2. evaluate which optimum pixel value results in the best reduction of the objective function (4.1)

For the first step, we can cache intermediate results and drastically reduce the computation. The second step can be approximated in a way that we do not need to evaluate the objective function.

Caching intermediate results Caching of intermediate results. We have seen in section 4.2.1 how we can find the optimum for a single pixel. We need to calculate $x_{opt} = \frac{-b}{2a}$ where $b = -2SUM(pixel * PSF * I_{dirty})$ and $a = SUM(PSF * PSF)$. First, we note that a depends only on the PSF which is constant, which means a is also constant. However, this is not true depending on the convolution we use.

How to cache b. Use correlation

Unsolved problem of varying PSF^2

Approximating the objective function a is constant in $\frac{-b}{2a}$

```
dirty = IFFT(Gridding(visibilities))

//pre-calculate bMap
dirtyPadded = ZeroPadding(dirty, psfSize)
fourierDirtyPadded = FFT(dirtyPadded)
fourierPsf = FFT(invert(psf))   //we require the correlation, that is why we invert
    the psf
bMap = IFFT(fourierDirtyPadded * fourierPsf)

//pre-calculate aMap. aMap stays constant over all iterations
aMap = new Array
for each pixelLocation in dirty
{
  aMap[pixelLocation] = CalcA(pixelLocation, psf)
}

do
{
  oldObjectiveValue = objectiveValue

  //the core of the algorithm
  pixelLocation = IterationStrategy(residuals)
  optimalValue = Minimize(residuals, pixelLocation)
  optimalValue = ApplyElasticNet(optimalValue)

  //housekeeping
  x[pixelLocation] = optimalValue
  residuals = (dirty - Convolve(x, psf))
  objectiveValue = SUM(residuals * residuals) + ElasticNet(x)
} while (oldObjectiveValue - objectiveValue) < epsilon
```

4.2.3 ElasticNet Regularization

L2 norm was used in other work. [8]

Formula

Effect

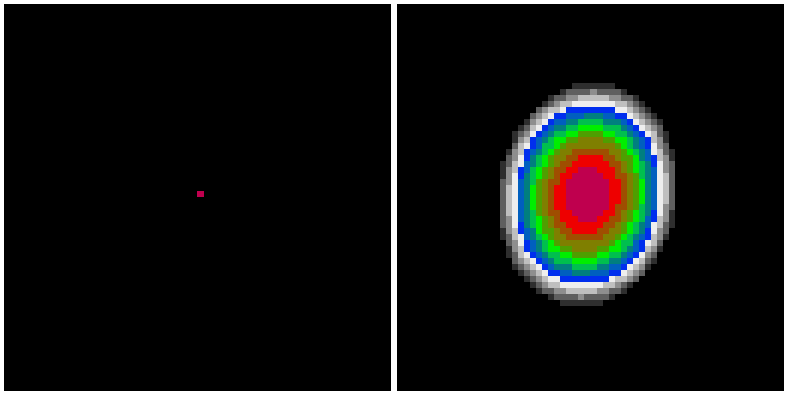
Implementation, proximal operator. Shrinkage of a division.

May even speed up convergence for correlated pixel values compared to L1 or L2[34]. But was not investigated in this project

4.3 Major Cycle convergence

Putting it all together

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



(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 11: Effect of the L1 and L2 Norm separately.

Coordinate Descent Path optimization [34] 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

4.4 Test on MeerKAT data

5 Tests on the MeerKAT LMC observation

For algorithmic testing

5.1 Wall clock time

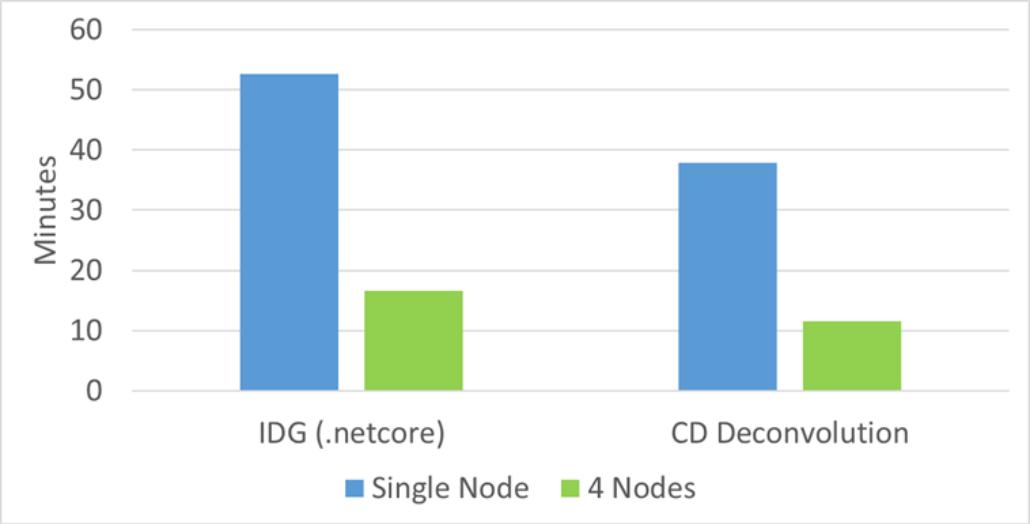


Figure 12: Wall-clock time of the distributed reconstruction

5.2 Quality

6 Conclusion

References

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

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

[3] 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.

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

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

[12] 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.

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

[14] 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.

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

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

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

[18] Jean-Luc Starck, David L Donoho, and Emmanuel J Candès. Astronomical image representation by the curvelet transform. Astronomy & Astrophysics, 398(2):785–800, 2003.

[19] Jean-Luc Starck, Fionn Murtagh, and Mario Bertero. Starlet transform in astronomical data processing. Handbook of Mathematical Methods in Imaging, pages 2053–2098, 2015.

[20] Rafael E Carrillo, Jason D McEwen, and Yves Wiaux. Sparsity averaging reweighted analysis (sara): a novel algorithm for radio-interferometric imaging. Monthly Notices of the Royal Astronomical Society, 426(2):1223–1234, 2012.

[21] 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.

[22] 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.

[23] 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.

[24] Christopher A Metzler, Arian Maleki, and Richard G Baraniuk. From denoising to compressed sensing. IEEE Transactions on Information Theory, 62(9):5117–5144, 2016.

[25] Hossein Mamaghanian, Nadia Khaled, David Atienza, and Pierre Vandergheynst. Compressed sensing for real-time energy-efficient ecg compression on wireless body sensor nodes. IEEE Transactions on Biomedical Engineering, 58(9):2456–2466, 2011.

[26] Michael Lustig, David L Donoho, Juan M Santos, and John M Pauly. Compressed sensing mri. IEEE signal processing magazine, 25(2):72, 2008.

[27] National Radio Astronomy Observations. Casa, 2019.

[28] Luke Pratley, Jason D McEwen, Mayeul d'Avezac, Rafael E Carrillo, Alexandru Onose, and Yves Wiaux. Robust sparse image reconstruction of radio interferometric observations with purify. Monthly Notices of the Royal Astronomical Society, 473(1):1038–1058, 2017.

[29] Arwa Dabbech, Alexandru Onose, Abdullah Abdulaziz, Richard A Perley, Oleg M Smirnov, and Yves Wiaux. Cygnus a super-resolved via convex optimization from vla data. Monthly Notices of the Royal Astronomical Society, 476(3):2853–2866, 2018.

[30] 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.

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

[32] 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.

[33] Zhi-Quan Luo and Paul Tseng. On the convergence of the coordinate descent method for convex differentiable minimization. Journal of Optimization Theory and Applications, 72(1):7–35, 1992.

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

List of Figures

List of Tables

7 attachment

8 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.

8.1 CLEAN: The Major Cycle Architecture

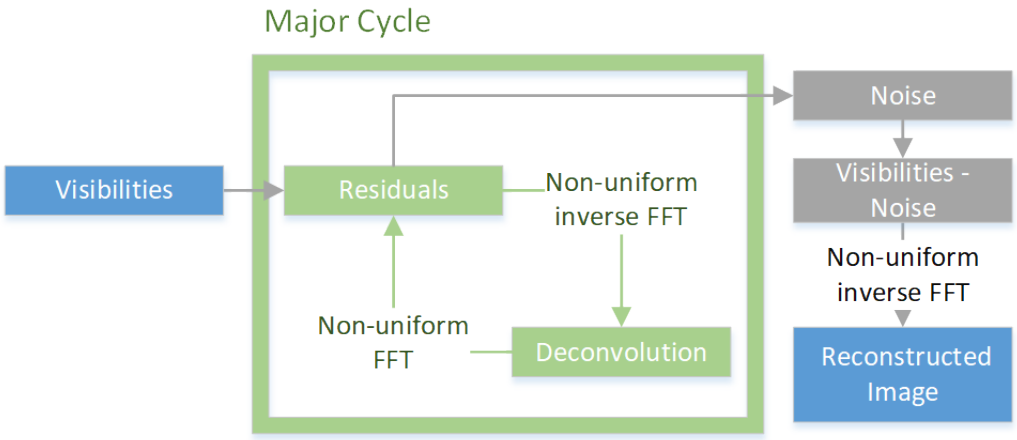


Figure 13: The Major Cycle Architecture

Figure 13 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.

8.2 Compressed Sensing Architecture

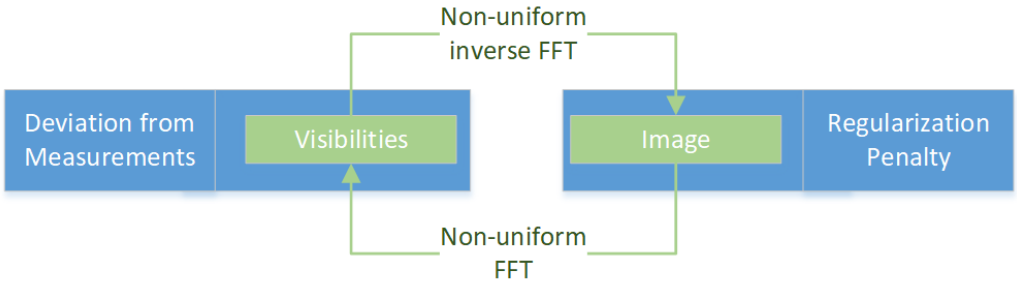


Figure 14: State-of-the-art Compressed Sensing Reconstruction Architecture

Figure 14 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.

8.3 Hypothesis for reducing costs of Compressed Sensing Algorithms

Compressed Sensing Algorithms are not bound to the Architecture presented in section 8.2. 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.

8.4 State of the art: WSCLEAN Software Package

8.4.1 W-Stacking Major Cycle

8.4.2 Deconvolution Algorithms

CLEAN MORESANE

8.5 Distributing the Image Reconstruction

8.5.1 Distributing the Non-uniform FFT

8.5.2 Distributing the Deconvolution

9 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[22, 23], continually reducing the runtime costs of the operation. Recently, Veeneboer et al[21] 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.

9.1 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.

10 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 15), which contains three operations: Gridding, FFT and Deconvolution.

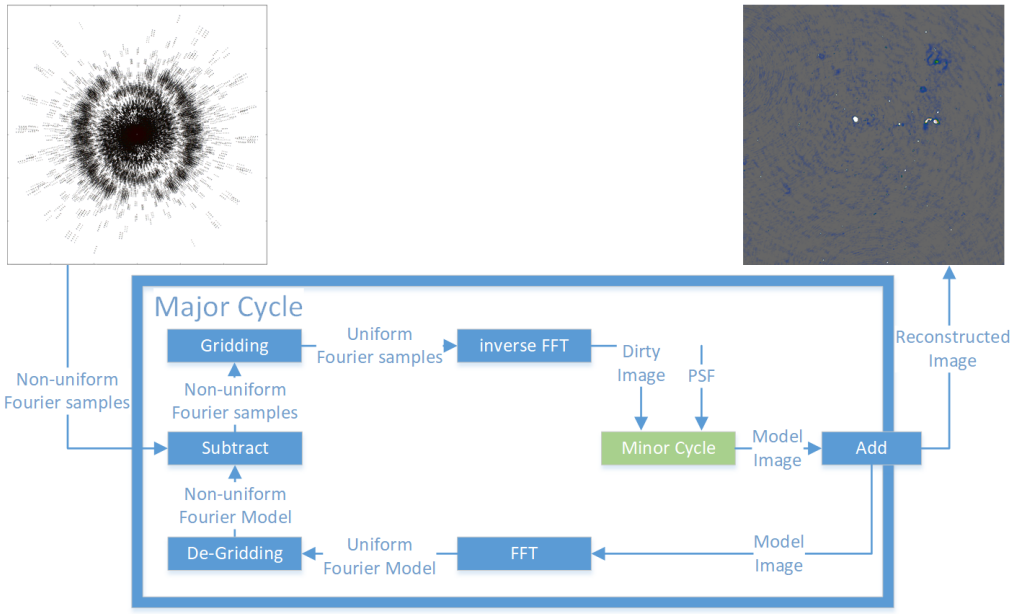


Figure 15: The Major Cycle Architecture of image reconstruction algorithms

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.

10.1 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[21] 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[32]. 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.

10.2 First steps towards a distributed Algorithm

In this project, we make the first steps towards a distributed Major Cycle architecture (shown in figure 15) 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.

11 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, August 9, 2019

Jonas Schwammberger