Imaging Pipeline Software

Bobbie Ware 17962233

Math705 Research Project

1.0 Abstract

Aperture synthesis is the process of taking data from interferometry telescope arrays and producing an image of the sky. This research project covers the gathering of knowledge on the required topics, and then using this knowledge to create software capable of performing these processes. The software will be tested against ideal models to ensure the correct results are being produced, and multiple versions of the software will be created to improve runtime performance.

Table of Contents

[1.0 Abstract 2](#_Toc11694726)

[2.0 Acknowledgements 3](#_Toc11694727)

[3.0 Introduction 6](#_Toc11694728)

[4.0 Literature Review 7](#_Toc11694729)

[4.1 Synthesis Imaging 7](#_Toc11694730)

[4.2 Fast Fourier Transform (FFT) 9](#_Toc11694731)

[4.3 Gridding 10](#_Toc11694732)

[4.4 Deconvolution 11](#_Toc11694733)

[5.0 Methods 12](#_Toc11694734)

[5.1 Design of pipeline 12](#_Toc11694735)

[5.2 Implementing 13](#_Toc11694736)

[5.3 Testing and Improving 13](#_Toc11694737)

[6.0 Software Implementation 15](#_Toc11694738)

[6.1 Gridder 15](#_Toc11694739)

[6.2 Inverse Fast Fourier Transform 16](#_Toc11694740)

[6.3 Deconvolution 16](#_Toc11694741)

[7.0 Results 18](#_Toc11694742)

[7.1 Gridding 18](#_Toc11694743)

[7.2 Fourier Transform 19](#_Toc11694744)

[7.3 Deconvolution 21](#_Toc11694745)

[7.4 Summary of the versions 22](#_Toc11694746)

[7.4.1 Version 1 22](#_Toc11694747)

[7.4.2 Version 2 22](#_Toc11694748)

[7.4.3 Version 3 22](#_Toc11694749)

[7.5 Results of Testing 23](#_Toc11694750)

[8.0 Discussion of Results 25](#_Toc11694751)

[8.1 Output Comparison 25](#_Toc11694752)

[8.2 Deconvolution Output 25](#_Toc11694753)

[8.3 Concurrent Updates issue 26](#_Toc11694754)

[8.4 Analysis of Runtimes 26](#_Toc11694755)

[8.5 Gridder Runtime Tests 28](#_Toc11694756)

[9.0 Conclusion 29](#_Toc11694757)

[9.1 Synthesis Imaging 29](#_Toc11694758)

[9.2 Gridder 29](#_Toc11694759)

[9.3 Fourier Transform 29](#_Toc11694760)

[9.4 Deconvolution 29](#_Toc11694761)

[9.5 Performance Improvements 29](#_Toc11694762)

[10.0 References 30](#_Toc11694763)

[11.0 Appendix 32](#_Toc11694764)

3.0 Introduction

An imaging pipeline for Image Synthesis is designed to take data either gathered by radio interferometry telescopes, or generated to simulate those telescopes, and form an image of an area of the sky. This software will be similar to that used for the Square Kilometer Array – a project AUT is involved in.

This project will follow a Design Science methodology where software will be the generated artefact used to experimentally investigate image synthesis. The project will start with the gathering of knowledge on the techniques used in an imaging pipeline. It will also involve gaining knowledge on solutions dealing with complications such as concurrency control and the mapping of visibilities to a grid. The project will involve implementing the techniques in the form of a Java program, and then will be tested using visibility input data available in the High Performance Computing Research Laboratory. The software will be tested against other pipelines and changes will be made to try to improve its performance.

The output from the developed software, using visibility data as input, will be an image of the sky. The images produced by the pipeline will be analysed to compare it against the known sky images for the data sets to validate whether the techniques are implemented properly and potentially look at its performance.

It is expected that an imaging pipeline will be developed with the capability of image synthesis. Also expected is to gain knowledge in the three main steps involved in the pipeline, namely gridding, (inverse) Fourier transform, and deconvolution. As well as some techniques for algorithm optimization.

4.0 Literature Review

## 4.1 Synthesis Imaging

The resolution of radio telescopes can be increased by using pairs of telescopes (baselines) and taking the product of the received signals. This resolution can be changed by increasing the separation of the baseline, rather then increasing the size of the individual telescopes. This method popularized by the work of (Ryle & Hewish, 1960) states that using these baselines the telescopes produces; “exactly the same result as that obtained by using the complete large aperture”. This technique allowed for cheaper production of much larger apertures and the eventual development of the techniques used now.

These techniques gather Fourier domain data in the form of a visibility. However, the way in which they are sampled is non-uniform, so it must be placed on a rectangular grid. This process is known as gridding and the methods used now are based on the work by (Brouw, 1975). These visibilities V(u, v) fall upon the plane in which the baselines are setup. For a wider coverage of this plane, more baselines can be added and could also be moved around.

With more modern telescopes being developed, moving them around became a substantial task and instead the rotation of the earth can be used to move these points around the plane. An image of these points on the V(u, v) plane can be seen on the left of Figure 1.

As there are gaps within the V(u, v) plane the image is a “dirty image”, this can be seen on the right of Figure 1.

Synthesis imaging is a large research field with many discoveries being made, a recent image produced of a black hole received a lot of public attention. Reading this paper (The Event Horizon Telescope Collaboration et al, 2018), showed insight into the process of developing an imaging pipeline.

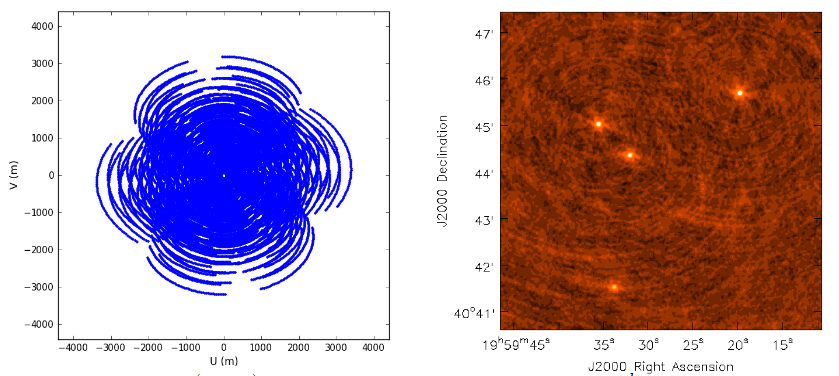


Figure 1. V(u,v) plane showing data points and a dirty image.

**Reference**

Rau, U. (2012, Sept 24). *Imaging and Deconvolution.* Retrieved from Australia Telescope National Facility: https://www.atnf.csiro.au/research/radio-school/2012/lectures/tue/RVU\_ImagingDeconvolution.pdf

## 4.2 Fast Fourier Transform (FFT)

A Fourier Transform is a process for signal-processing and analysis. (Brigham, 1988) states that the extent of the use the process is as follows; “biomedical engineering, imaging, analysis of stock market data, spectroscopy, metallurgical analysis, nonlinear systems analysis, mechanical analysis, geophysical analysis, simulation, music synthesis”. It is widely regarded as one of the most important algorithms because of its impact in so many areas.

Simply put a Fourier Transform is used to show different parts of a continuous signal. However, for Interferometry an Inverse Fourier Transform is used as we are taking the data from the Fourier Plane and creating an image from it.

While it is possible to perform a basic Fourier Transform using computers, the method has a run time of O(). Instead, we can perform a Fast Fourier Transform with a run time closer to O() as it is a divide and conquer method. Such a method is based upon the work of (Hogg, MacDonald, Conway, & Wade, 1969). The algorithm used was first discovered by Gauss and later rediscovered by (Cooley & Tukey, 1965) which notes that, “Wherever possible the use of N = with r = 2 or 4 offers important advantages”. This algorithm works under the assumption that the data is in an organized array, hence the visibilities must be gridded. Using butterfly operations, the data is combined in pairs using either a decimation in time or in frequency variation.

## 4.3 Gridding

Gridding is the process of mapping the data collected onto a rectangular grid so that it may be processed by the iFFT, and then displayed as an image. Early techniques for placing the visibility data on a grid involved finding the closest grid point to visibility’s V(u, v) co-ordinate. Then, either adding them all together or averaging them out on that grid location. Early methods were used by (Hogg, MacDonald, Conway, & Wade, 1969). However, their method produced many artifacts in the image. Therefore, there were limited application for the process. An alternative method used by (Brouw, 1975) would take a weighted value based on the distance between local grid point and the point of the visibility. By designating a “support” area around the local grid point, the data can be added to these areas. An ideal gridding method was given by (O'Sullivan, 1985), with his gridding algorithm that used a sinc function and produced images with “arbitrarily small artifact levels”. However, this function would give infinite extent to the support. This is not ideal computationally for the gridder. Instead, convolution kernels with a set support area are used. These functions also have a quick fall off in the grid to help aliasing.

A simplified approach to gridding follows this process; for every visibility, find the closest grid point to the data on the V(u, v) plane. Then, using the convolution kernel, the data point is spread across the support region. By using a convolution kernel generated by a Prolate Spheroidal, the value of the data point is spread based on its distance to the center.

## 4.4 Deconvolution

Once an image is formed from the Fourier Transform is it called a ‘dirty image’, as seen in Figure 1, this is due to the effects of having limited sampling of the V(u, v) plane. The process of Deconvolution can be used to ‘CLEAN’ the image. This method by (Högbom, 1974) uses the original V(u, v) data to form a “dirty beam”, then by taking away the dirty beam from the points of the dirty image with the greatest brightness you are left with a residual image. By iteratively carrying out this process the effects of the convolution are removed to the best extent possible.

This was expanded upon by (Clark, 1980) to make it more efficient. His method involves using more FFT’s in a major and minor cycle to subtract points away from the dirty image. The minor cycle works by performing a Högbom ‘CLEAN’ on smaller beam patches, then the major cycle applies an FFT on the points found by the minor cycle and is used to subtract from the dirty image. There is also a varient of this process where (Schwab, 1984) uses the major cycle to take away from un-gridded visibilities. This helps to remove noise from and potential errors from the gridding process.

Further examples of the ‘CLEAN’ algorithm and the Maximum Entropy technique by (Skilling & Bryan, 1984) are compared by (Cornwell & Bridle, 1996).

The image in Figure 2 shows the ‘CLEAN’ image based on the images from Figure 1.

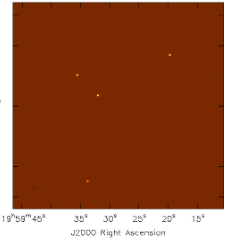
**

Figure 2. A ‘CLEAN’ image.

**Reference**

Rau, U. (2012, Sept 24). *Imaging and Deconvolution.* Retrieved from Australia Telescope National Facility: https://www.atnf.csiro.au/research/radio-school/2012/lectures/tue/RVU\_ImagingDeconvolution.pdf

5.0 Methods

The goal of the project is to gather sufficient knowledge on the topics required to produce software capable of taking data from interferometry telescopes and produce an image of the sky. Starting with a literature review, relevant knowledge regarding the topics of Image Synthesis, gridding, Fourier Transforms, and Deconvolution were gathered. Using this knowledge, an artefact will be produced that to perform the required processes. Once the pipeline is created, it will be tested against a perfect image of the sky to ensure that it is correctly carrying out the processes involved. The pipeline will then be adapted and improved upon to increase its performance.

## 5.1 Design of pipeline

The pipeline will consist of three main sections; the Gridder, the inverse Fourier Transform, and the Deconvolution. The programming language used to develop the pipeline was Java. Java is primarily taught in computer science papers at AUT and is capable of handling the size of the data used. It also supports parallelization, a feature that will be implemented to improve performance.

The grid size to be used will be 1024x1024. Providing a high enough resolution to be able to properly test against. The grid length and height must be a power of 2 due to the (Cooley & Tukey, 1965) radix 2 FFT being used. This algorithm is more performant then a standard Fourier Transform under the condition that the data is ordered, hence why the visibilities are gridded. There are two different ways of implementing an iFFT; decimation in time, or decimation in frequency. Both methods are explained and compared in the software implementation section.

The data contains Visibilities.csv which is a file that includes around 23000 data points. Each visibility consists of its locations along the V(u, v) plane, followed by its value as a complex number. The dataset also includes the Prolate Spheroidal which is used as the convolution kernel for the data with a support size of 7. However, for better precision it has been oversampled by a factor of 4. The last thing required from the data is the configuration for achieving the best accuracy for the gridder.

A double is used as the primitive data type, as in the Java language (Oracle, 1993) a double can store values from 4.94065645841246544e-324 to 1.79769313486231570e+308 including values with high decimal accuracy needed.

For Deconvolution, Högbom’s ‘CLEAN’ method will be implemented to deconvolve the true image out of the dirty image. To achieve this, a dirty beam must be constructed from the V(u, v) plane.

## 5.2 Implementing

The process of creating the pipeline will undergo many stages. This is due to multiple iterations of the software being produced as improvements are made. When the pipeline is complete, a fully formed Java project with unit testing and multiple output images will be produced.

When implementing the pipeline, constant reference to the literature will be required to ensure the processes are carrying out their operations correctly. The nature of the mathematical operations mean that high precision is needed for a meaningful image to be produced.

For a more in-depth explanation of the methods involved in the implementation, please see the following section.

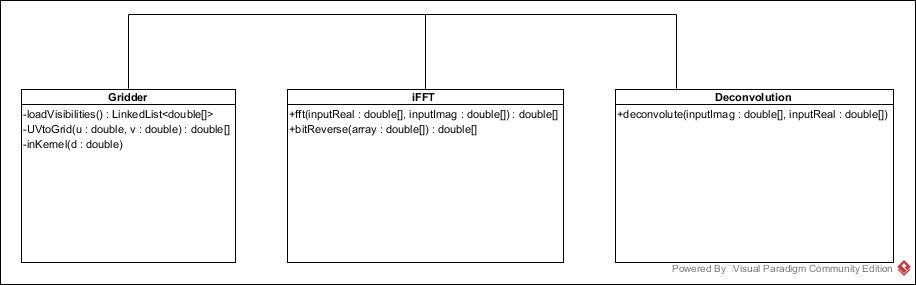
## 5.3 Testing and Improving

To test the pipeline, a dataset generated by the HPCRL at AUT will be used. By using this data, the three main processes of the pipeline can be checked to see how they are performing as it can be tested against a perfect image made from the real components of the sky. This perfect image is what the pipeline would generate if working under ideal circumstances, so any inconsistences would be evidence that improvements can be made.

As the pipeline is tested, improvements will be made with respect to its runtime. An example of this is using multiple threads to perform the gridding or the Fourier Transform. However, implementing multiple threads brings issues such as concurrent updates from the threads. A concurrency issue occurs when two threads attempt to update the same values simultaneously and only one update is saved. This would arise with overlapping visibilities in the gridder and requires a solution such as a lock on each value to prevent being accessed at the same time. Another method is to have each thread manage its own grid, combining each grid after all the visibilities have been processed.

6.0 Software Implementation

The implementation will be broken into three classes. Gridder.java, which is responsible for reading the dataset and placing the points on a grid. iFFT.java, for carrying out an Inverse Fourier Transform on the data. Deconvolution.java, implementing a ‘CLEAN’ algorithm to turn the dirty image into a more representative form. Lastly, a python script will be used to display the image. This script was given by the HPCRL at AUT. Below is a UML diagram of the Java classes.



## 6.1 Gridder

Implementing the gridder begins with initializing two 2D double arrays. These are used to store the real and imaginary values respectively. As the values are placed on the grid they will be stored in these arrays. Following that, the gridder loads the data from the .cvs file and stores it as a double array. The first two indexes are the u and v coordinates, after that is the real and then imaginary values. These double arrays are placed in an array list data structure to ensure the size of the list is not an issue. It also allows for the data to be spilt up when needed. Then going through the list, each array is transformed as per the formula mentioned before.

In order to correctly place the visibilities, two things must be used; a wavelength to meters ratio, and the UV scale. For the data generated, a frequency of 300000000 was used. To calculate the scale, we use the product of the grid size and the cell size. As discussed earlier, our grid size is 1024, and cell size is 4.848136811095360e-06 given by the dataset.

Once the visibilities are modified correctly, the process of placing the points on the grid begins. Firstly, we take the location modified by the UVScale and round it to the nearest integer to find the closest grid point. Then, using the support value of 7, we take a 7x7 grid around this center grid point as this is where the data point will be placed. Using our convolution kernel, we form a 2D array 27x27, due to our x4 oversampling, and finding the distance of the grid point to the true visibility point.

## 6.2 Inverse Fast Fourier Transform

The iFFT will be used on a 2D array. This process involves transforming every row in the image, followed by transforming every column. A decimation in time variation will be used so the arrays will be bit reversed first. A bit-reversal is done on the inputs as the in-place operations imply a bit-reversed output. The decision between decimation in time or decimation in frequency the order in which the operation take place. For decimation in time, a bit reversal occurs first, and then progressively larger distance in the butterfly operations. Decimation in frequency is when the butterfly operations get shorter in distance and then the array is bit reversed. For this implementation, decimation in time is used. There is no advantage in using one technique over another as both carry out the same computations.

The process involves one complex multiplication, addition, and subtraction. A complex multiplication consists of two real additions and four real multiplications. Complex addition and subtractions involve two real additions. For the transform, the real and the imaginary parts are spilt into two separate arrays. This is done to avoid having to use a separate object to store them.

## 6.3 Deconvolution

For this process, a Högbom ‘CLEAN’ will be implemented. This process will remove any artifacts and leave only point sources in an otherwise empty field. For this process, a dirty beam must be formed. To obtain a dirty image, the process of gridder is carried out using the original visibilty data but instead of the values given, each point is set to have a value of 1. This works by finding points of high brightness within a number of regions in the dirty image. Then, by taking away a set amount of this point using the dirty beam, and adding the point to a set of clean components. This process is done iteratively in each region until no more points that match the required brightness can be found. By using these components, a restored image can be formed. This image is only an estimate of the sky.

7.0 Results

The implementation of the pipeline resulted in three different versions of the software, the discussion of the results will first cover the output from the pipeline and then the improvements made.

## 7.1 Gridding

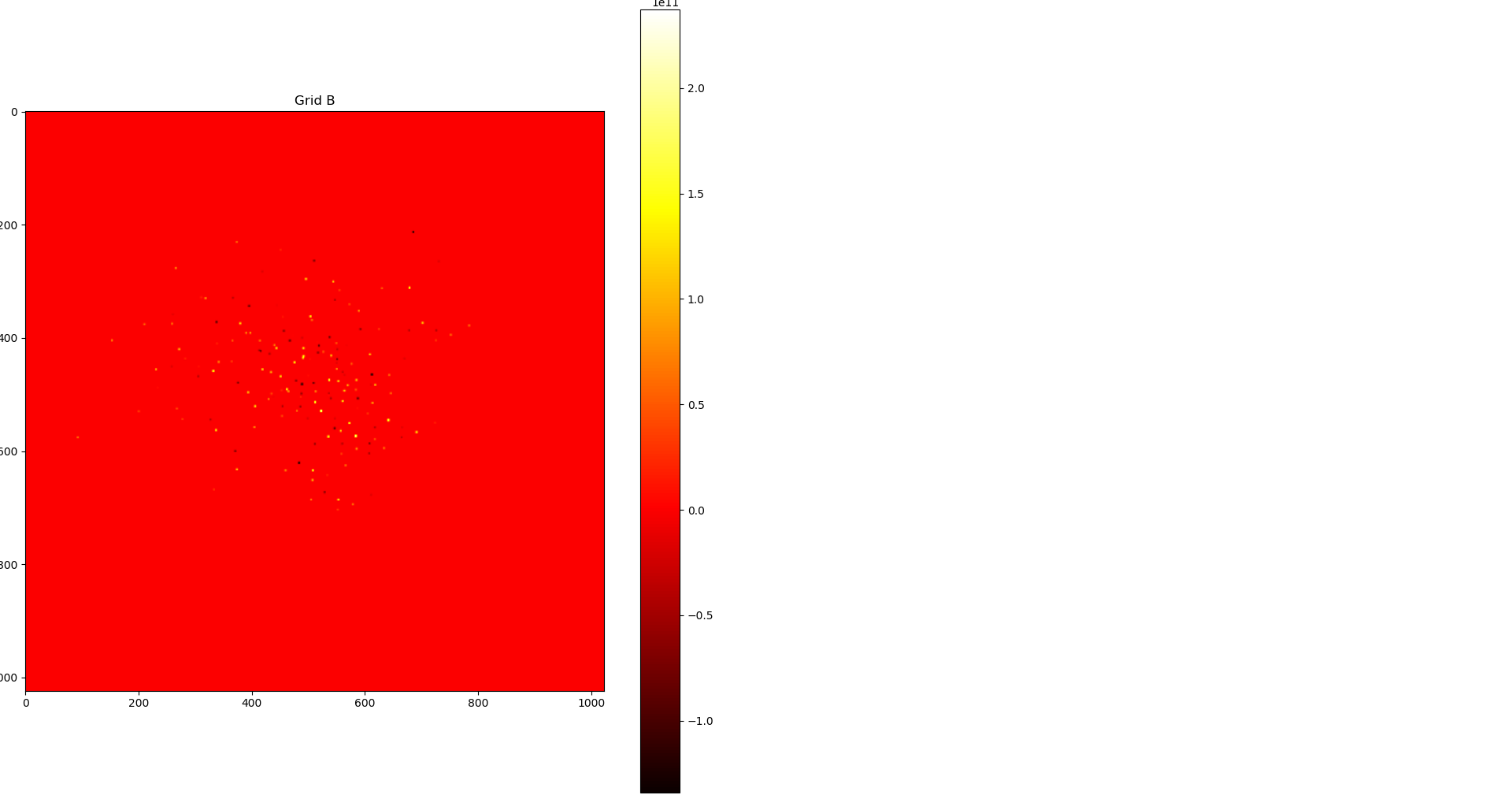
The gridder was able to successfully place all 223000 visibilities onto the grid. The first version of the gridder used a complex data type for each pixel on the grid, this was the simplest way of implementing the correct process for gridding. The image produced from the gridder can be seen in Figure 3, and a closer look can be seen in Figure 4.

Figure 3. The image produced by the gridder

As we can see in Figure 3 small points of positive or negative values are spread across the image. To compare the accuracy of the gridder the inverse Fourier transform must be carried out and if the image is accurate then the gridder is functioning correctly. What we can see is working for the gridder is the shape of the prolate spheroidal, in Figure 4 we can see round shapes of strong values that fall off around the edges. When the gridder is given a position for a visibility, it only points to one pixel, therefore a large amount of shapes with a size of around 7x7 pixels means that it has correctly spread the values.

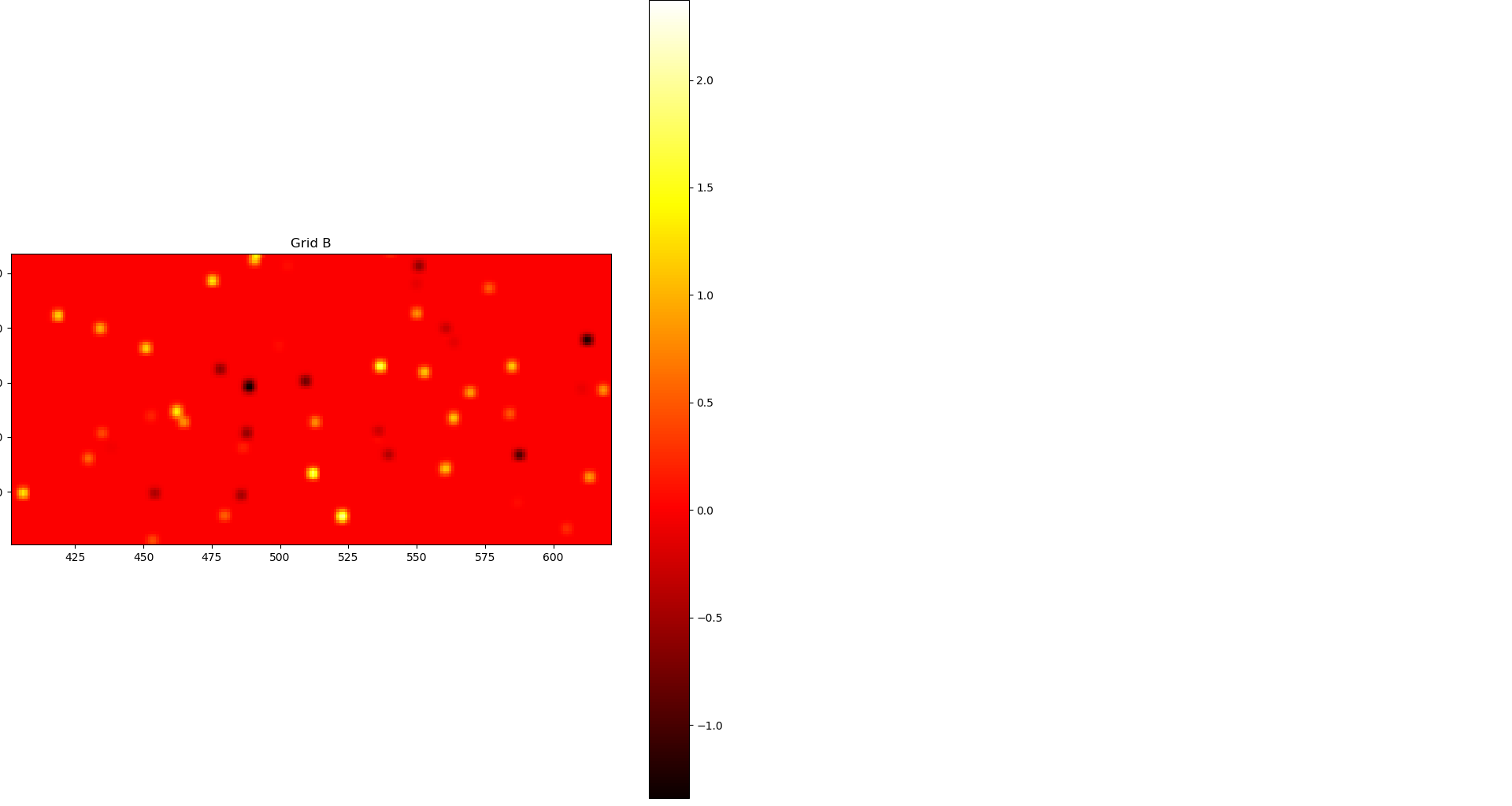


Figure 4. A closer look at part of the image in Figure 3.

## 7.2 Fourier Transform

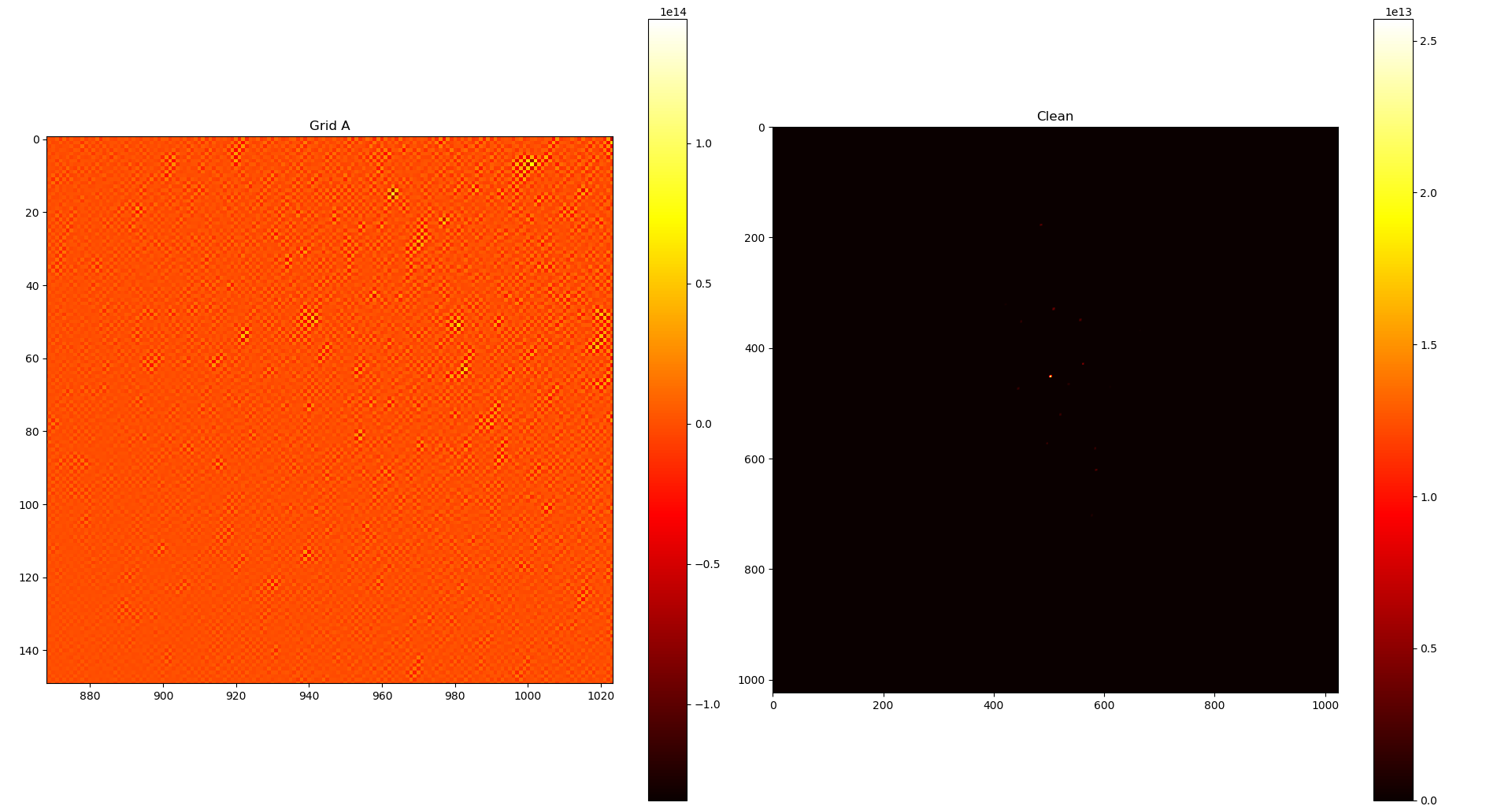
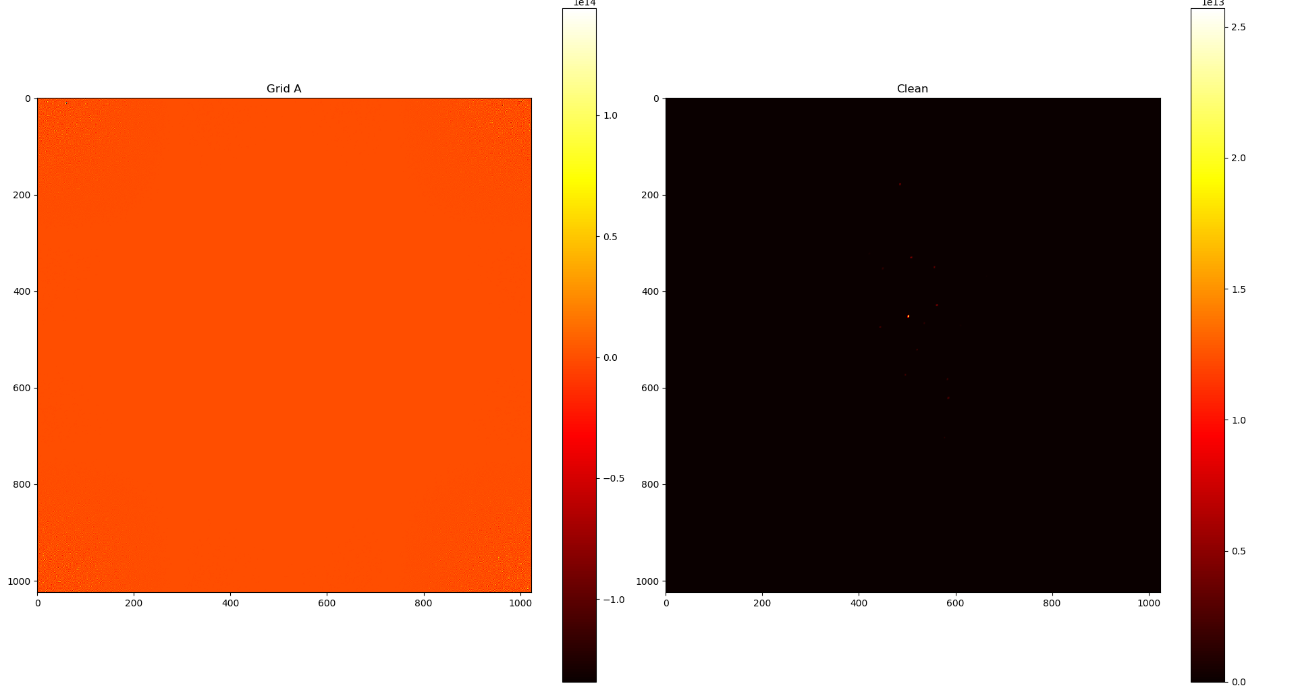
The implementation of the transform took the longest to code, it is not only the most complicated algorithm used, it also requires a lot of manipulation of the grid before and after the transform. For the specific algorithm used, all the data must be shifted so that the outmost corners are placed in the middle of the image. This is done before and after the transform. To illustrate this, the shift hasn’t been done on Figure 5 but has on Figure 6.

Figure 5. Without shifting the image is transformed incorrectly and the bright points are in the corners (Left). A zoomed in look showing the top left corner (Right).

A screenshot of a computer

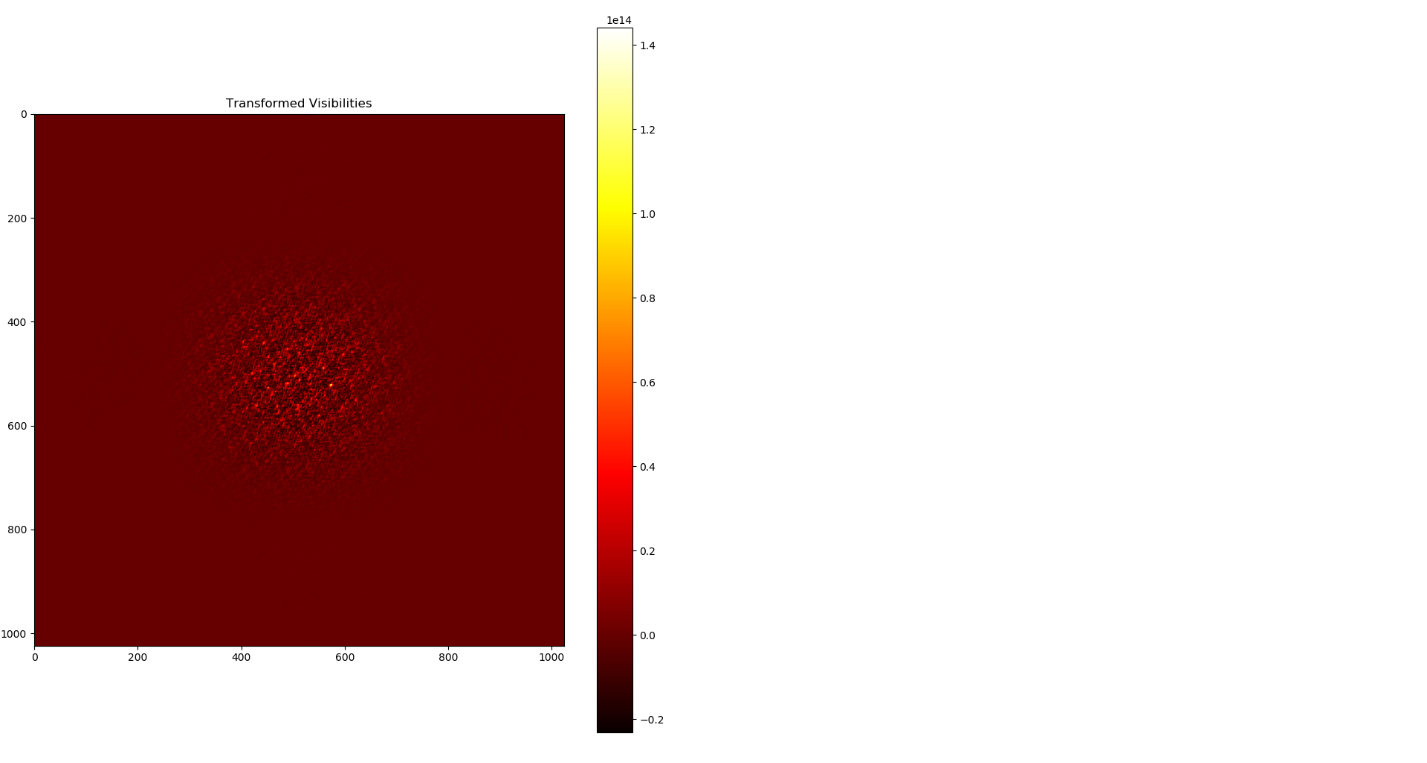
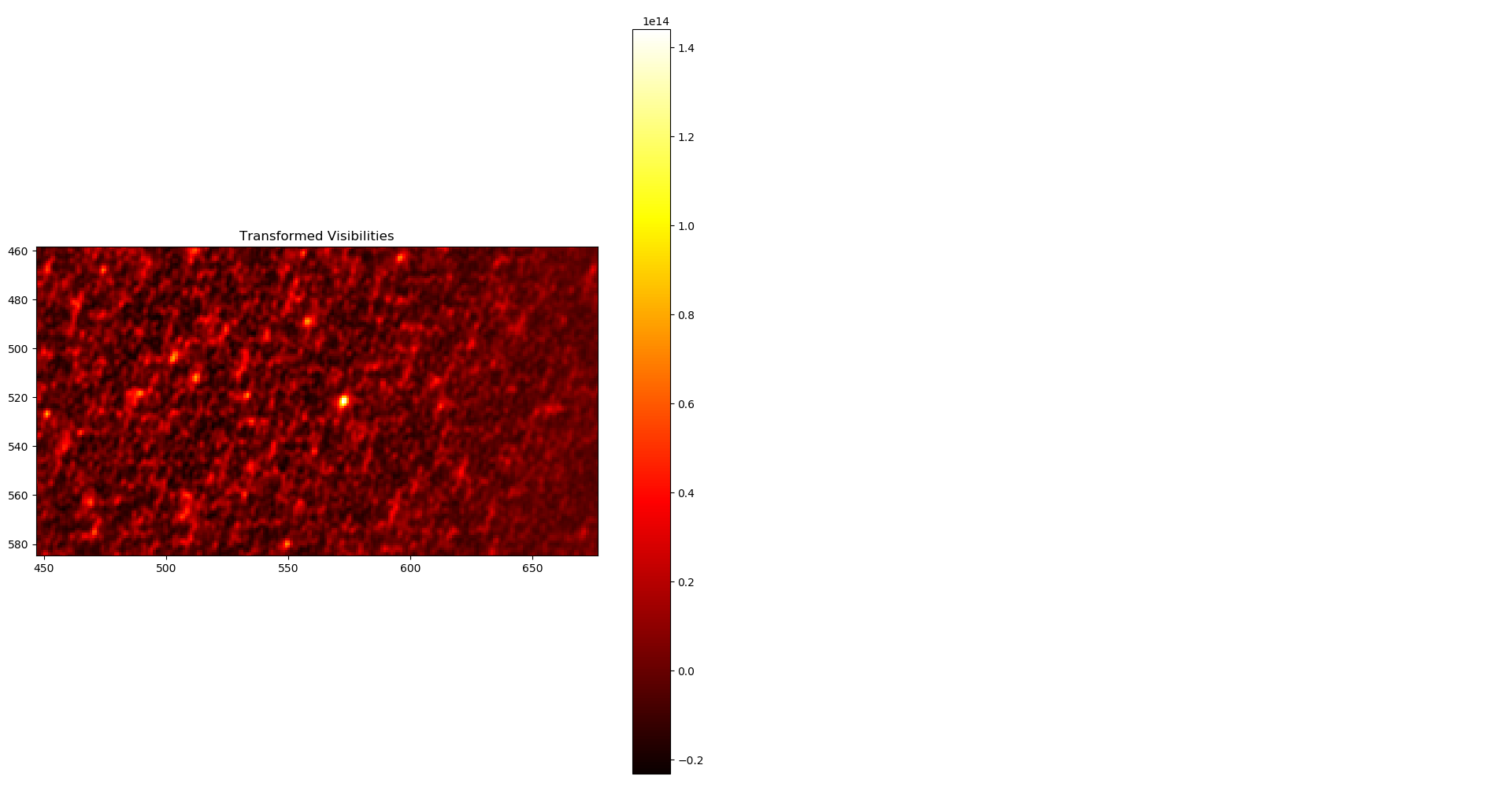
Description generated with high confidenceAs seen in Figure 7, a bright source is found near the center of the center of the image, this is potentially a star within the area of the sky imaged. The result of the transform then has a convolution correction applied so the corners are as bright as the center of the image. The image has the correct appearance for an inverse transform using our data, so we can assume the gridder and the Fourier Transform are implemented correctly.

Figure 7. Image with convolution correction applied.

Figure 6. Grid has been shifted before and after the transform (Left). A zoom in look of the middle of the image (Right).6

## 7.3 Deconvolution

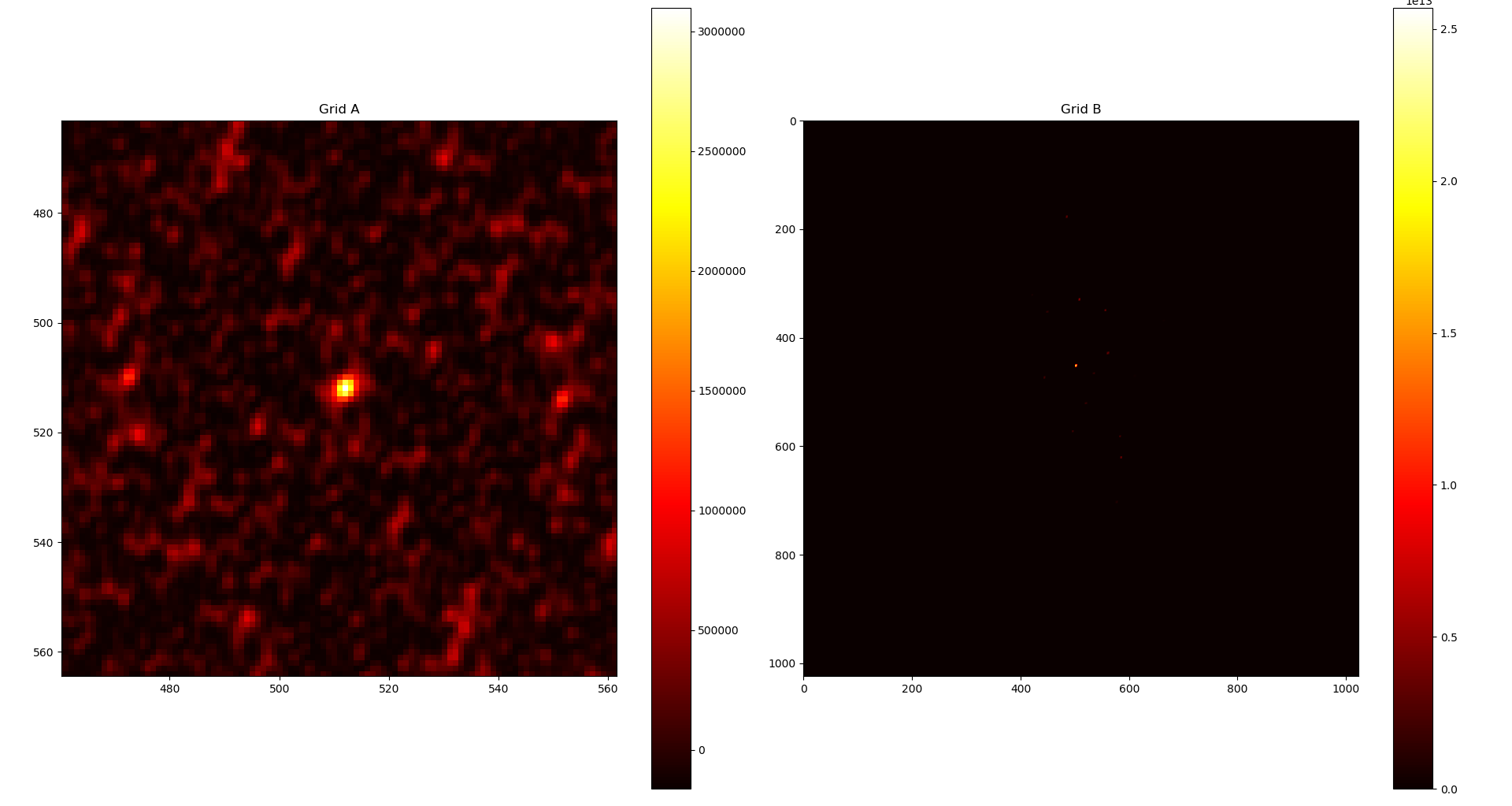
In order to carry out deconvolution of the image from the transform, we must construct a dirty beam. We get this by carry out the gridder and the Fourier transform again but instead of using the large complex numbers given, each point has a value of 1. When this process is done the dirty beam seen in Figure 7 can be seen.

Figure 8. Dirty beam found at center of the produced image

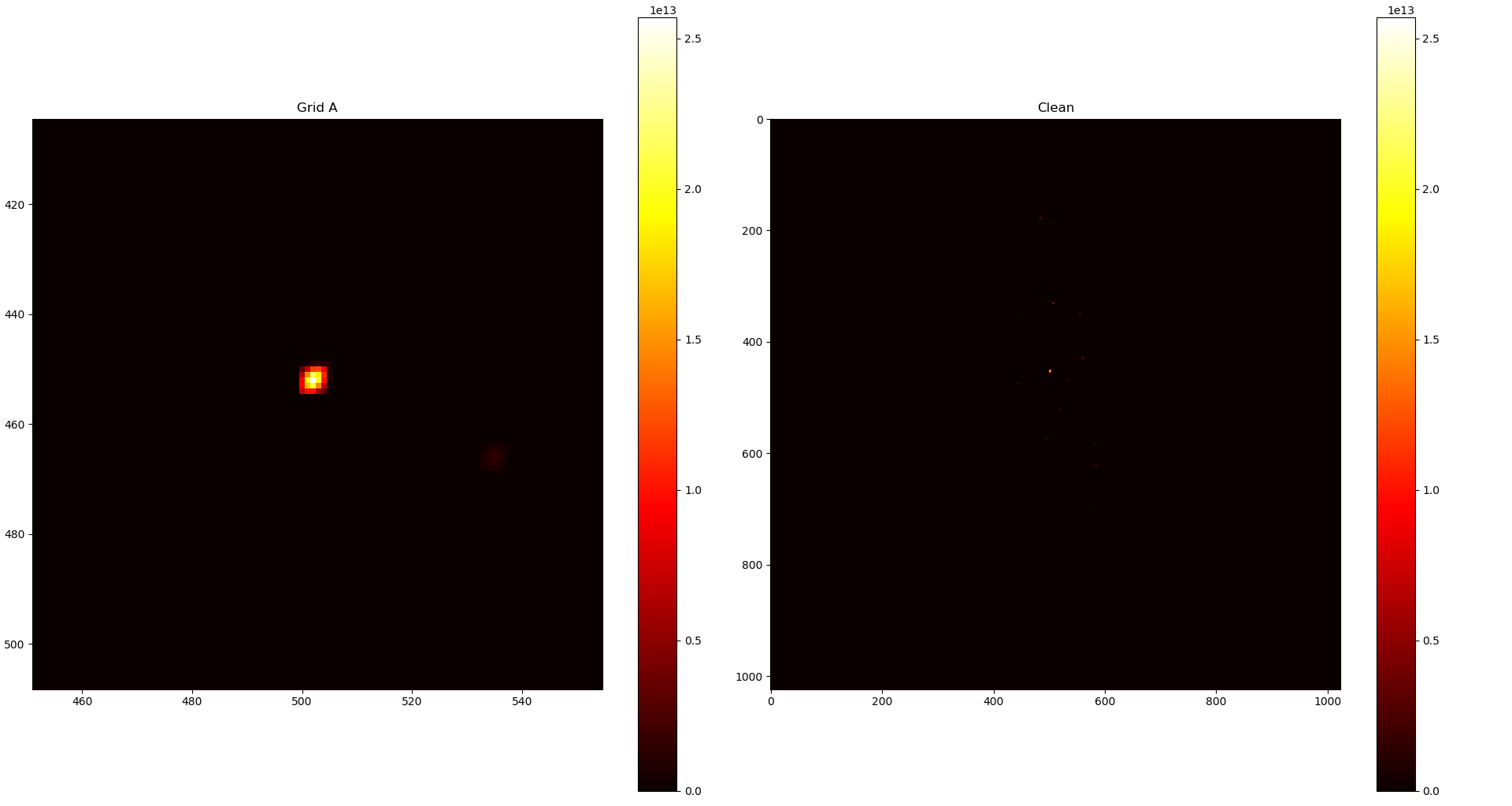
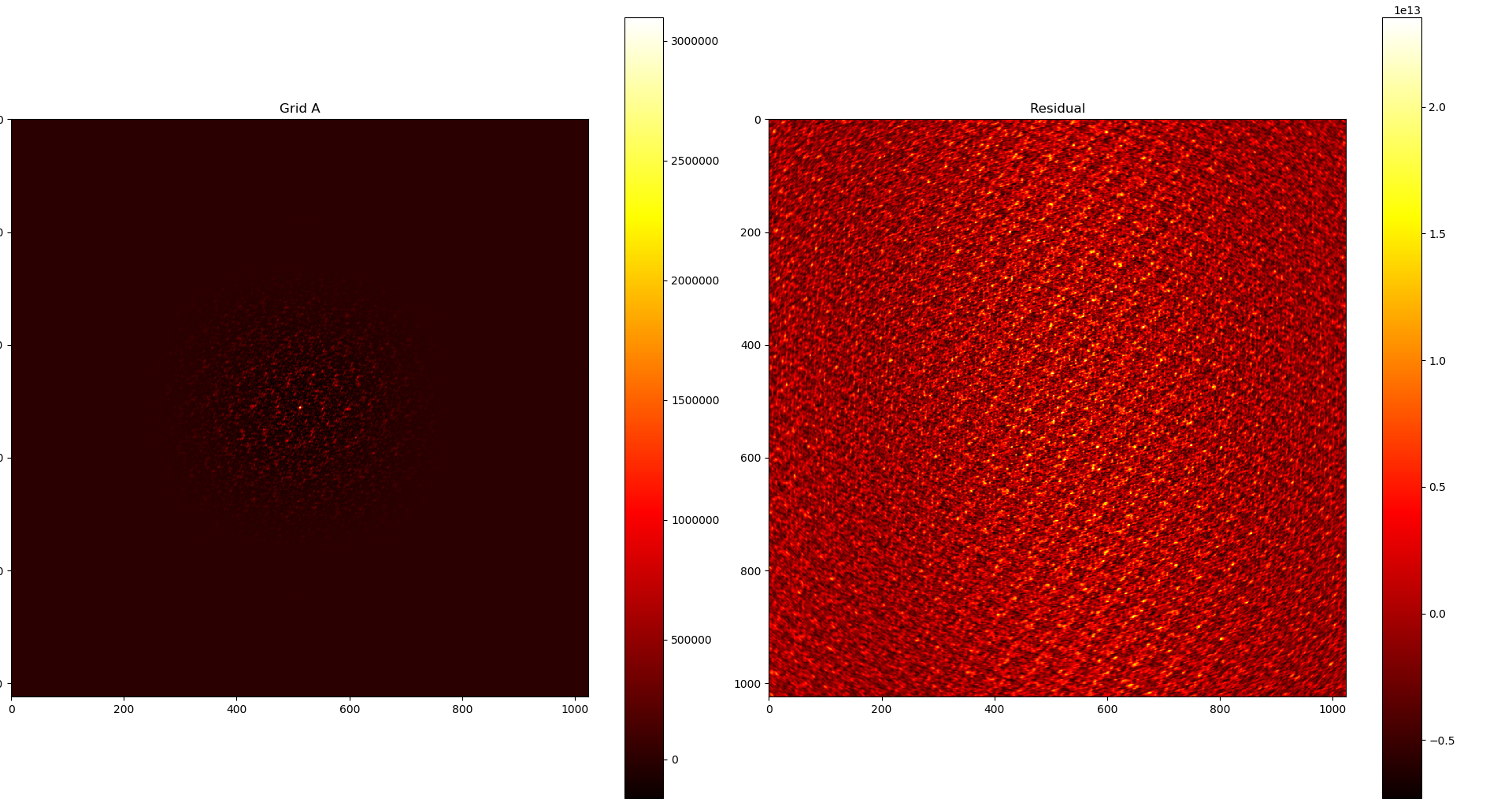
This shape is subtracted from the previous image, leaving a grid with only residual data, this can be seen in Figure 8. The CLEAN image can be seen in Figure 8, this was produced with 1000 iterations of the Högbom clean.

Figure 9. Residuals from CLEAN algorithm

Figure 10. A zoomed in section of the clean image, the whole image is Appendix A.

## 7.4 Summary of the versions

### 7.4.1 Version 1

The first version of the pipeline used a complex data type to store each pixel in the grid. This was done for simplicity when creating the first version. For the functions such as adding, subtracting, and multiplying, the code was contained within the data type. Throughout all versions of the pipeline an assumption to not overwrite data is present, this involves the creation of multiple 2D array of complex objects. This version is referred to as *UsingComplex*.

### 7.4.2 Version 2

Instead of using a Complex data type, the second version of the pipeline uses two double data types to represent a complex number. This involves having two separate 2D arrays, one named *real* and the other *imaginary* (imag for short). In order to implement the grid in such a way, all of the complex operations have to occur slightly differently with no functions being stored in a data type. This pipeline was named *Using2Grids*.

### 7.4.3 Version 3

For the final version of the pipeline implemented multithreading for the gridder and the Fourier Transform stages. This was to improve performance and compare against previous version. This version required the most development time as it is the most complex. Most of this complexity come from concurrent update issue with the gridder. The transform was easier to implement as each thread deals with a separate array at a time. These arrays are spilt evenly amongst the threads and there is no overlap between the threads. This version is referred to as *UsingMultipleThreads*. It also uses the double 2D array method (used by *Using2Grids*) as it was found to be more performant.

## 7.5 Results of Testing

All of the versions of the pipeline produced the same output. This was expected as they carry out exactly the same operations just in different ways. Using a perfect image output that was produced when the data was generated we can compare how different the two are. The relative difference between the two can be calculated using the same python script used to render the images. The difference between the produced image and the perfect image is 1.091%.

To compare runtimes, each pipeline ran through the entire process 100 times on the same system and under the same conditions. PC and software specs can be found in Appendix B.

|  |  |  |  |
| --- | --- | --- | --- |
|  | *UsingComplex* | *Using2Grids* | *UsingMultipleThreads* |
| Mean (nanoseconds) | 709,178,183.1 | 568,478,821.6 | 471,953,759.3 |
| Standard Deviation (nanoseconds) | 19,570,930.35 | 24,486,395.06 | 38,638,593.29 |

Figure 11. Table of runtimes

The gridder was also tested independently, with this data set only 223,000 visibilities were used when a typical pipeline would use closer to 3,000,000 visibilities. To test the gridders thoroughly, they were given sets of visibilities of increasing size. The data sets given began with 100,000 and went up to 1,000,000, this will show the differences in how the runtimes scale with size. Figure 10 shows the results of the testing and the raw data can be found in Appendix C.

Figure 12. Graph shows the increasing runtime (nanoseconds) it takes for the three gridders to process data of increasing size.

8.0 Discussion of Results

## 8.1 Output Comparison

The image produced by the pipeline is very close to that of the perfect image. It cannot be exactly same because the perfect image was produced under perfect gridding conditions. The gridders produced for the pipeline use techniques suitable for most situations.

The 1.091% is well within the acceptable parameters for the pipeline, this can be seen visually as well, both images have the same characteristics in the location of the bright point and shape of the pattern made by the Fourier Transform. One difference that can be seen in Figure 7 when compared to Figure 6 is that the image has been rotated after the convolution has been applied, this is to match the configuration of the perfect image and a simple transformation to apply.

When the deconvolution is applied to the perfect image an almost identical output is created. this is due to as much of the convolution applied by the gridder being removed, most of the difference in image comes from the difference in gridders.

## 8.2 Deconvolution Output

As seen in the CLEAN images in Figure 10, there is a small point of light located just to the side of the very bright point. This is likely an artifact of too many iterations of the Högbom CLEAN taking place, the brightest point has slowly decreased as the iterations take place and lesser point have began to be added to the CLEAN image.

Therefore, iterations of different sizes should be run, and the outputs compared. This was left in this report to show a potential side effect from the deconvolution process.

## 8.3 Development of Versions

The time to develop the code was by far the longest part of this research project. At each part pseudocode and a large amount of trial and error occurred until the correct processes were being carried out.

For the first version *UsingComplex,* took a large amount of time to figure out the correct implementation of the algorithms. It was made slightly easier by using the complex data type as the code was much easier to read. This allowed for the versions to be developed faster as it was a good base to work off of and make more complex.

*Using2Grids* improved upon the previous version by dropping the data type and storing the values in two arrays. This iteration required changes at every point in code as the operations were completely different. This change was effective though as seen in the improvements in runtime. The code itself becomes slightly harder to read as it is difficult to follow where the values are being accessed from.

The final iteration produced, *UsingMultiThreaded,* took much longer to develop then *Using2Grids.* The introduction of using threads made this version much more complex then the other two. Only 4 threads were used for this version as there was found to be diminishing returns when using more threads.

The versions as a whole were a large amount of work to produce, but once images were begin produced and the improvements were effective, the development was very successful.

## 8.4 Concurrent Updates issue

To improve the performance of the gridder, multiple threads were used to deal with an equal amount of visibilities. Four threads were used, each getting a quarter of the 223,000 visibilities. However, this caused an issue when adding values to the grid. If two threads had visibilities located at nearby points, there is an issue where both threads try to update the value at the same time with neither account for the others value. This issue is covered, and some solutions provided in the book (Arpaci-Dusseau & Arpaci-Dusseau, 2018).

A solution considered to solve this issue was assigning a lock variable to each grid point, to ensure only one thread has access at any one time. This approach was not adopted as there will have been significant overhead with threads waiting on each other, instead a different solution was chosen. Each thread contains its own 2D grids, therefore each visibility is added independently from each other. Once each thread has finished, all 4 grids are combined, giving the final grid.

## 8.5 Analysis of Runtimes

The runtimes shown in Figure 11 show each version having faster performance then the last. To ensure this was the case three statistical tests were carried out. Paired t-tests were used to test the mean runtime as all pipelines used the same data. The following code was ran using RStudio.

> t.test(UsingComplex, Using2Grids, paired = TRUE)

Paired t-test

data: UsingComplex and Using2Grids

t = 27.165, df = 19, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

129858744 151539979

sample estimates:

mean of the differences

140699361

> t.test(Using2Grids, UsingMultipleThreads, paired = TRUE)

Paired t-test

data: Using2Grids and UsingMultipleThreads

t = 9.6711, df = 19, p-value = 8.995e-09

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

75635012 117415113

sample estimates:

mean of the differences

96525062

Figure 13. Output from t-tests

For the tests we use a null hypothesis that the means are equal, given the p-values from the tests, the results are statistically significant and so we can say that each pipeline has a faster runtime then the previous version.

For the case of *Using2Grids* over *UsingComplex*, the improvements come from not have to create and destroy the complex data type. Each time a new array is created the pipeline has to cerate 10242 objects to store the values whereas using the double primitive this doesn’t have to occur.

*UsingMultipleThreads* has the advantage of splitting the work load for the grinder and the transform up among multiple threads whereas *UsingComplex* and *Using2Grids* deal with each process iteratively one at a time.

## 8.6 Gridder Runtime Tests

The tests carried out in Figure 12 show how the runtimes of the different versions compare when given increasing amounts of visibilities.

When given smaller amounts, the gridders all perform similarly. *UsingComplex* and *Using2Grids* have similar times as a small amount of complex object are created and destroyed. *UsingMultipleThreads* has a similar performance as the time taken to created threads and start them is relatively large to the time to process the small amounts.

As the amount of visibilities increase, the spread between the lines increase. The order of the pipelines follow that of the runtimes of the entire pipeline process.

*UsingMultipleThreads* is by far the fastest and the difference only increases with more visibilities. With the work load being split between multiple threads, its runtime doesn’t increase at the same rate as the other two gridders. Also, the time taken to instantiate and start the threads becomes negligible compared to the time of processing the data. The time spent combining the threads after they have completed their load is also negligible as it is just an addition of 4 numbers for each grid point and doesn’t change with more visibilities being added in.

9.0 Conclusion

## 9.1 Synthesis Imaging

The literature review led to knowledge being gathered on the processes involved in developing an imaging pipeline. This knowledge was applied with a Design Science methodology to create and improve code. This process was a success as

## 9.2 Gridder

This project led to the successful development of software capable of placing visibility data on a grid. Through the images output from the pipeline it is known that the gridder carries out the correct processes. The gridder ended up being the complicated code to implement, however, because of the optimisations it ended up being where the most runtime was saved over the older versions.

## 9.3 Fourier Transform

The transform implemented was capable of performing an FFT (Fast Fourier Transform).

## 9.4 Deconvolution

## 9.5 Performance Improvements

10.0 References

Arpaci-Dusseau, R. H., & Arpaci-Dusseau, A. C. (2018). *Operating Systems: Three Easy Pieces.* Arpaci-Dusseau Books, LLC.

Brigham, R. O. (1988). *The Fast Fourier Transform and its Applications.* Prentice-Hall.

Brouw, W. N. (1975). Aperture Synthesis. In C. De Jager, & H. Nieuwenhuijzen, *Image Processing Techinques in Astronomy* (pp. 301-307). Dordrecht: Springer.

Clark, B. G. (1980). An efficient implementation of the algorithm 'CLEAN'. *Astronomy and Astrophysics*, 377-378.

Cooley, J., & Tukey, J. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 297-301.

Cornwell, T., & Bridle, A. (1996). *Deconvolution Tutorial*. Retrieved from National Radio: https://www.cv.nrao.edu/~abridle/deconvol/deconvol.html

Högbom, J. (1974). *Astronomy and Astrophysics Supplement Series*, 417.

Hogg, D. E., MacDonald, G. H., Conway, R. G., & Wade, C. M. (1969). Synthesis of Brightness Distribution in Radio Sources. *Astronomical Journal*, 1206-1213.

Oracle. (1993). *Class Double*. Retrieved from Javadocs: https://docs.oracle.com/javase/7/docs/api/java/lang/Double.html

O'Sullivan, J. D. (1985). A Fast Sinc Function Gridding Algorithm for Fourier Inversion in Computer Tomography. *IEEE Transactions on Medical Imaging*, 200-207.

Rau, U. (2012, Sept 24). *Imaging and Deconvolution.* Retrieved from Australia Telescope National Facility: https://www.atnf.csiro.au/research/radio-school/2012/lectures/tue/RVU\_ImagingDeconvolution.pdf

Ryle, M., & Hewish, A. (1960). The synthesis of large radio telescopes. *Monthly Notices of the Royal Astronomical Society, Vol. 120*, 220-230.

Schwab, F. R. (1984). Relaxing the isoplanatism assumption in self-calibration; applications to low-frequency radio interferometry. *Astronomical Journal*, 1076-1081.

Skilling, J., & Bryan, R. K. (1984). Maximum Entropy Image Reconstruction. *Monthly Notices of the Royal Astronomical Society*, 111-124.

The Event Horizon Telescope Collaboration et al. (2018). First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole. *The Astrophysical Journal Letters*, L1.

11.0 Appendix

## Appendix A

The full ‘CLEAN’ image. It shows the main bright point found near the center of the image.

A close up of a computer

Description generated with high confidence

## Appendix B

The specifications for the computer used are as follows.

* Processor: Intel(R) Core(TM) i5-4690K CPU @ 3.50GHz, 3501 Mhz, 4 Core(s), 4 Logical Processor(s)
* Installed Physical Memory (RAM): 8.00 GB
* OS: Microsoft Windows 10 Pro
* Version: 10.0.17134 Build 17134
* IDE: Eclipse Java 2018-12

## Appendix C

|  |  |  |  |
| --- | --- | --- | --- |
| Test Number | UsingComplex (nanoseconds) | Using2Grids (nanoseconds) | UsingMultipleThreads (nanoseconds) |
|  |  |  |  |
| 1 | 721572276 | 595707985 | 471013685 |
| 2 | 707869718 | 598323862 | 549086228 |
| 3 | 737278067 | 564050352 | 529862565 |
| 4 | 702845398 | 576725692 | 473142139 |
| 5 | 695590810 | 553444657 | 492900328 |
| 6 | 736351787 | 552710596 | 481913999 |
| 7 | 709737198 | 553394920 | 472064892 |
| 8 | 685787637 | 553895508 | 477841998 |
| 9 | 687502981 | 558225850 | 466101420 |
| 10 | 707462460 | 602770354 | 444948252 |
| 11 | 741156094 | 644431864 | 437897000 |
| 12 | 738374038 | 570869009 | 576250265 |
| 13 | 733327775 | 569788252 | 439863662 |
| 14 | 704690643 | 557532749 | 444330050 |
| 15 | 695773960 | 550730184 | 439328257 |
| 16 | 693642581 | 558114673 | 453925497 |
| 17 | 717103548 | 556702141 | 451920508 |
| 18 | 699835136 | 545330788 | 445945041 |
| 19 | 686543933 | 549577455 | 442779424 |
| 20 | 681117621 | 557249541 | 447959976 |
| Mean | 709178183.1 | 568478821.6 | 471953759.3 |
| Standard Deviation | 19570930.35 | 24486395.06 | 38638593.29 |