INSTITUT NATIONAL DES SCIENCES APPLIQUÉES RENNES

Engineering Cycle Internship

4^{th} year internship

presented by

Fabio Eid MOROOKA

Engineering Student of INSA Rennes

Speciality EII

School year 2018 - 2019

DASK description of a deconvolution algorithm in radio astronomy

Location of the Internship

Laboratory of Signals and Systems (L2S)

Internship tutors

François ORIEUX, Nicolas GAC

Pedagogical Correspondent INSA Rennes

Jean-François NEZAN

Internship period: 20/05/2019- 25/09/2019

Résumé

Le but de ce document est expliquer le travail effectué pendant le stage. Il présente d'abord le projet Square Kilometer Array (SKA), puis il présente le Laboratoire des Signaux et Systèmes (L2S) et sa relation avec lui. On y trouve également une brève explication du processus d'interférométrie radio astronomie. Il explique ensuite les défis du projet SKA. Ensuite il présente la bibliothèque DASK et comment elle peut être utilisée pour résoudre certains des défis que SKA propose. Le but principal de cette étude est de faire une première étude de DASK, une bibliothèque python qui a pour objectif de faire du calcul parallèle et du calcul haute performance (HPC). Ensuite, une étude d'un code qui fait la reconstruction de l'image par la méthode du filtre de Wiener sera faite pour ensuite l'adapter à l'utilisation de la bibliothèque DASK. Quelques études de DASK seront faites en analysant la performance de ce code dans une machine locale et dans une machine distribuée. Ce code sera également exécuté dans un HPC, puis des études sur ses performances seront effectuées lors de la réalisation du calcul dans le cluster. Enfin, un tutoriel sera créé pour les futurs étudiants et chercheurs qui souhaitent en savoir plus sur dask. La dernière partie de ce rapport est la conclusion de l'étude réalisée, discutant ainsi de la bibliothèque DASK elle-même et ensuite de la viabilité de l'utilisation de la bibliothèque DASK dans le contexte de SKA.

Mots-clés : DASK, Calcul Haute Performance, Programmation parallèle, déconvolution, square kilometer array.

Abstract

This document explains the work done during the internship. Firstly it presents the Square Kilometer Array (SKA) project, then it presents the Laboratory of Signals and Systems and its relationship with it. It has also a briefly explanation of the radio interferometry process. Then it provides an explanation of the challenges in the SKA project. It is presented then the DASK library and how it may be used to solve some of the challenges SKA proposes. The main goal of this study is to make a first study of DASK, a python library that has the objetive to do parallel computing and High-Performance Computing (HPC). Then a study of a code that make the image reconstruction by the Wiener Filter method will be done to then adapt it to the use of DASK library. Next some studies of DASK will be done by analysing the performance of this code in a local machine and in a distributed machine. This code will also be executed in a HPC, and then some studies of its performance will be done when making the computing in the cluster. Finally a tutorial will be created to future students and researchers who wants to learn more about dask. The last part of this report is the conclusion of the study made, so discussing about the DASK library itself and then the viability of using the DASK library in the context of SKA.

Key-words : DASK, High-Performance Computing, Parallel programming, deconvolution, Square Kilometer Array.

INSA Rennes
20 Avenue des Buttes de Coësmes
CS 70839 - 35708 Rennes Cedex 7
Tél. +33 (0) 2 23 23 82 00 - Fax +33 (0) 2 23 23 83 96
www.insa-rennes.fr

Table of Contents

List of Figures . . . 4
Acronyms . . . 5
Introduction . . . 6
I The Beginning : SKA project and The L2S . . . 7
I.1 SKA project . . . 7
I.2 The L2S . . . 8
II The Dark ages : The challenges of SKA . . . 9
II.1 Image reconstruction . . . 9
II.2 Inverse Problems . . . 9
III The first light : Studies and solutions . . . 10
III.1 Numpy x DASK . . . 10
III.2 Dask Tutorial - Misson in Rennes . . . 12
IV The universe : The future of SKA and final points . . . 13
IV.1 Perspectives . . . 13
IV.2 Personal and Professional Experience . . . 13
IV.3 Final considerations . . . 14
IV.4 Acknowledgments . . . 14
Bibliography . . . 15

List of Figures

Figure I-1 SKA project 7
Figure III-1 Graph executing time 10
Figure III-2 Graph of the computing time 11
Figure III-3 PBS file example 12

Acronyms

Abbreviation	Explanation
CNRS	Centre National de la Recherche Scientifique
CPU	Central Processing Unit
ENS	École normale supérieure
FFT	Fast Fourier Transform
FPGA	Field-Programmable Gate Array
GPGPU	General-Purpose computing on Graphics Processing Units
GPI	Groupe Problèmes Inverses
HPC	High-Performance Computing
IETR	Institut d'Electronique et de Télécommunications de Rennes
IFFT	Inverse Fast Fourier Transform
INS2I	Institut des sciences de l'information et de leurs interactions
INSIS	Institut des sciences de l'ingénierie et des systèmes
ISPO	Intenational SKA Project Office
ISSC	International Square Kilometre Array Steering Committee
LOFAR	Low Frequency Array
LTWG	Large Telescope Working Group
MPI	Message Passing Interface
PBS	Portable Batch System
PhD	Doctor of Philosophy
PSF	Point Spread Function
SAR	Synthetic Aperture Radar
SKA	Square Kilometer Array
SNR	Signal-to-noise ratio
UK	United Kingdom
UMR	Unité Mixte de Recherche
UPS	Université Paris-Sud
URSI	International Union of Radio Science
USA	United States of America
VLA	Very Large Array

Introduction

For the past few decades, scientists, companies and universities around the world have been working seeking to discover more about the universe we live in. So every year radio telescopes are getting larger and larger trying to acquire more data, achieving then higher resolutions. However the biggest questions of the universe will not be answer only by huge telescopes. Square Kilometre Array (SKA) is an magnificent international scientific project aiming to build the largest radio telescope ever designed (more efficient than the Very Large Array (VLA) and faster than the Low Frequency Array (LOFAR) and will be able to achieve results never seen before. Sky image reconstruction at this scale is a huge challenge to this project, because of the enormous quantity of data that will be processed and treated it. So it is essential to have programs that will process all data in real-time and the most efficient possible, by doing as much parallelism as possible. This reconstruction process is done by solving a large-scale of inverse problems.

In order to achieve the image reconstruction given the challenges of data and for consequence, memory, it is necessary to create real-time programs that will be executed in parallel and most efficient possible. These algorithms solve the inverse problem of the image reconstructions by the Wiener Filter method and the Huber deconvolution. The programs were made considering that they will be executed in a computing cluster (High-Performance Computing). Therefore the main goal of this internship is to explore parallel computing of different image reconstruction programs using the Python's library DASK [5] and its interaction in a computing cluster. Furthermore, this internship has also the objective to create a tutorial of DASK [12] with the purpose to introduce the DASK library to future students, interns or researchers who want to learn DASK.

This report covers the work done during my internship. In the first section, the laboratory of signal and systems and its relationship with SKA project is presented. It also aims to briefly explain the historical evolution of SKA project. The second part gives a general overview of the existent algorithms developed to make the reconstruction of images and its performance before improving it, and the library python DASK as well. The next section reveals the modifications made in the algorithms seeking the improvement of its performance. It is also discussed its performance and its applications in different hardware. Finally, the last section discusses the study developed during the internship and its relevance to the future of SKA project. A plan is proposed to elaborate another solution for the challenges studied.

I The Beginning : SKA project and The L2S

In this section the SKA project will be presented and the L2S. Then, the relationship between inverse problems and the internship project.

I.1 SKA project

The SKA project is an international project which objective is to build the world's largest radio telescope, with more than a square kilometre (on million square kilometres) of collection area. This project is an unique instrument that represents a huge step for engineering and research & development.

The project consists in build thousands of dishes and up to a million low-frequency antennas with the purpose to monitor the sky thoroughly and much faster than the actual systems. Its unique configuration will generate images with a much better resolution than the Hubble Space Telescope.

FIGURE I-1 – SKA project

The Radio Interferometers Principle

The radio interferometry in astronomy is an array of antennas or telescopes that are connected to form a single telescope, providing a higher image resolution of astronomical objects such as stars, galaxies and nebulas. This technique allows to capture higher resolution images without constructing gigantic telescopes.

The different antennas of the array detects signals from the astronomical sources in the sky with different delays. These signals are properly treated (amplified, digitized and delayed) to be correctly correlated two by two to maximize the signal-to-noise ratio (SNR). Each pair of antenna generates a single position of the source. The collection of all correlated signals from different pair of antennas with a different baseline (distance between two antennas) is called Complex Visibility. These visibilities in the plane of the interferometer are equivalent to the 2D Fast Fourier Transform (FFT) of the sky image. That way, the more baselines there are in the interferometer, more information the interferometer captures from the source. After the correlation, the generated visibilities are treated to obtain the image.

History of SKA

In September 1993 started a history of one of the biggest projects in the world. The International Union of Radio Science (URSI) created the Large Telescope Working Group (LTWG), an international group with the purpose to develop the next generation of radio telescope. Several meetings of this working group lead to a forum where it was mobilised a scientific community to cooperate in achieving this goal. So in 1997 it was signed by eight institutions from six countries (Australia, Canada, China, India, the Netherlands, and the USA) the first memorandum of agreement to cooperate in a technology study program with the objective to build the very large radio telescope. [9] and [10]

Others memorandums were signed during SKA's history : in 2000 was established the International Square Kilometre Array Steering Committee (ISSC) composed by eleven countries (the first six countries and Germany, Italy, Poland, Sweden and the UK).

This memorandum was followed by the memorandum of Agreement to Collaborate in the Development of the SKA project in 2005 which has been extended until 2007. This agreement also established the International SKA Project Office (ISPO). Three proposals for the expansion of the ISPO were sent to the ISSC, then in the end of 2007, after several discussions it was signed a memorandum that established the University of Manchester as the host organisation for the Project Office. In the next year another memorandum was signed to create the SKA Science and Engineering Committee (SSEC) to substitute the ISSC. Its objective is to provide the primary forum to discuss and decide the scientific and technical matters for SKA. [9] and [10]

SKA France is a national organization which has the objective to prepare the SKA project in France, jointly by the CNRS, Paris Observatory, Côte d'Azur Observatory, Bordeaux University and Orléans University [9]. The project SKALLAS is an national project with the support of SKA France that aimms to propose an Algorithm Architecture matching approach to size HPC systems with strong requirements in terms of throughput, cost, computing precision and energy efficiency. It has four collaborators : The L2S, the IETR, the Paris Observatory and the Côte d'Azur Observatory. The company ATOS-BULL is a partner of this project since they have an expertise on parallel programming and it is a worldwide company in HPC.

In the end of 2011 to centralise the leadership and to formalise relationships between international partners, it was created the SKA Organisation, a not-for-profit company with the purpose to led the SKA project. After 2007 the office of the SKA organisation changed to the Alan Turing building in Manchester in 2008, and then to new building at Jodrell Bank Observatory in Cheshire, UK in 2012.

The project

SKA will be the largest radio telescope array ever constructed with over one square kilometre. The telescope will be arranged in multiple spiral arm configuration. Each dish will extend the central core, creating a long baseline interferometer array.

Computers will combine the signals captured by each receiver to synthesize the signal as the equivalent of a single dish measuring the distance between the two scopes. Thereby the project will emulate a telescope with a size equal to the distance between the telescopes in the array. So rather than building one gigantic dish it will be constructed a system that can act as one gigantic telescope or multiple smaller telescopes. This dishes and antennas will be constructed in Africa and in Australia. The next process consist in make the reconstruction of the sky image from the visibilities. Finally the reconstructed image will pass by a calibration process that will enhance its quality. This entire process is inspired from LOFAR.

I.2 The L2S

The Laboratory of Signals and Systems (L2S) is a Joint Research Center (UMR8506) of the National Centre for Scientific Research(CNRS) hosted by CentraleSupelec and Paris-Sud University (UPS). It is mainly linked to the Institute for Information Sciences and their Interactions (INS2I) and the Institute of Engineering and Systems Sciences (INSIS). The laboratory is organized in three research divisions : The Signals and Statistics, The Systems and Control and The Networks and Telecoms. Each center is responsible for solving problems by doing research in its field.

The Signals and Statistics division has two groups : The Inverse Problems group and the Modeling and Estimation group. The first makes the interface between the physics, statistics and signal or image processing. Some of the problems that the group works are microwave and 3D tomography imaging, Synthetic Aperture Radar imaging and infrared astronomy imaging. Meanwhile the second works on time series analysis, point processes and non Gaussian processes, sparse decomposition for source separation and multi-sensor signal processing.

The inverse problem team of the laboratory is the responsible for research and producing the most efficient algorithm to be used in the image reconstruction problem of SKA. The team has expertise in solving scientific problems that involves inverse problems as the SKA project image reconstruction.

II The Dark ages : The challenges of SKA

In this section it will be presented with some details the challenges of SKA project and how the inverse group problems of L2S is responsible for solving this issue. First of all, it will be presented the idea of image reconstruction and then how this is a problem in SKA project, then it will show how the inverse problem group is involved in solving this problem.

II.1 Image reconstruction

Image reconstruction in astronomy has the objective to recover the original sky image from the data collected by the interferometric telescope. Radio interferometer telescope is a solution to obtain a higher resolution without constructing a large size telescope. However this technique makes complex computation. The process to construct the sky image consists in two parts : imaging and deconvolution.

To first step of imaging consists in making some Fourier Transform to generate the dirty image from the measured visibilities. Next the deconvolution step uses this dirty image generated before and reconstruct the original sky image.

It has been developed a lot of algorithms to solve inverse problems. However they are not directed to large-scale problems. The project SKA has plenty of challenges, mainly in signal processing and high performance computation. Particularly for making the image reconstruction the main challenge is the huge amount data that is difficult to work without using a lot of memory and that will have to be processed in many computers.

II.2 Inverse Problems

Recover the sky image from the convolved image (dirty) is represented as an inverse problem solved by doing a deconvolution. This problem is represented in the equation (1).

$$y_t = H_t * x_t + n_t \tag{1}$$

Where y_t is the dirty image, H_t is the deconvolution matrix, x_t is the sky image that the frequency domain and n_t is the noisy factor. However the equation (1) is an ill-posed inverse problem, studied and solved by many methods.

Wiener filter

One method to solve this inverse problem is the Wiener deconvolution. The method consist on finding a g_t that estimate x_t as the equation (2) [3] :

$$\hat{x}_t = g_t * y_t \tag{2}$$

where $\hat{x}_t$ is an estimate of x_t that minimizes the mean square error.

$$\hat{x} = argmin||y - Hx||^2 \tag{3}$$

Considering that the matrix H is a square and circulating matrix, the solution of the equation (3) is done by the equation (4) below.

$$\mathbf{G}(u,v) = \frac{\mathbf{H}(u,v)\mathbf{Y}(u,v)}{|\mathbf{H}(u,v)|^2\mathbf{Y}(u,v) + \lambda|\mathbf{L}(u,v)|^2}. \tag{4}$$

INSA Rennes
20 Avenue des Buttes de Coësmes
CS 70839 - 35708 Rennes Cedex 7
Tél. +33 (0) 2 23 23 82 00 - Fax +33 (0) 2 23 23 83 96
www.insa-rennes.fr

III The first light : Studies and solutions

In this section, it will be presented the main DASK studies made and its conclusion. It will be also showon the differences between a numpy python code and a DASK python code. Finally, it was made a study of codes in clusters. In this internship, it was used the Fusion Mesocenter of CentraleSupelec.

III.1 Numpy x DASK

The first part of the internship was to understand the code python that makes the wiener filter and the huber deconvolution, and then change them to use the library DASK instead of numpy. Since DASK is adaptive to numpy python users, almost all numpy functions are the same in DASK, so it was simply to change the libraries import in the beginning of the code and execute the code and compare the results of the numpy code and the DASK code.

Then some tests were made to explore more about the library DASK, this different tests were made in three type of Wiener filter codes. The first one with only a single image with the size of 2048x2048 pixels, then the second one was to work with larger images, and the last one consisted in working with more wavelength in the spectral cube (in this code it was used the same image replicate in each wavelength just to simulate the application of the Wiener filter in "different" images by wavelength). Furthermore some tests were also made using the huber deconvolution code, just to see the viability of a parallelization.

Simple profiling

The first profiling of the code was made by using the existent python libraries "time" and "cProfile". Both are simpler libraries that can be used in any python code. This first profiling had the objective to analyse the difference between a numpy python code and a DASK python code. For this test it was used the program that makes the Wiener filter in images of 2048x2048 pixels and also larger images. The time for processing all program was taken by each python library, the results for comparing the Wienier filter in different size of images are in the graphic 1 bellow.

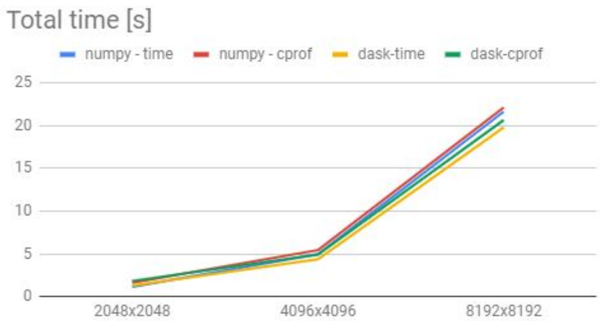

FIGURE III-1 – Graph executing time

This graph shows that for larger images there is an memory error in the Numpy codes, meanwhile DASK codes supports larger images (more data) since DASK evaluate computations in a low-memory footprint. Another thing it was possible to analyse when using the system monitor of ubuntu is the use of all CPU when executing a DASK code, meanwhile the numpy codes uses only one CPU.

After making the study in larger images, it was made the same study (codes time processing using existent python libraries), but for different wavelengths, and to analyse the difference in the way of making the python program, it was made 2 types of program for each python library (numpy and DASK). One code was made using a loop "for", so by doing the same image treatment one by one, and a second program was made but using the entire spectral cube as the input.

For this part of the study, the executing time of the program was separate in three parts for DASK codes : the time to load the file ".npy" to the program, the time to construct the computation graph, and then the time to compute the graph. The numpy codes have only the time for loading the file and the time to computing the result since it does not create a computational graph. For studying the influence of DASK in the codes, it was considered the computation time. These results are show in the graph 2 bellow.

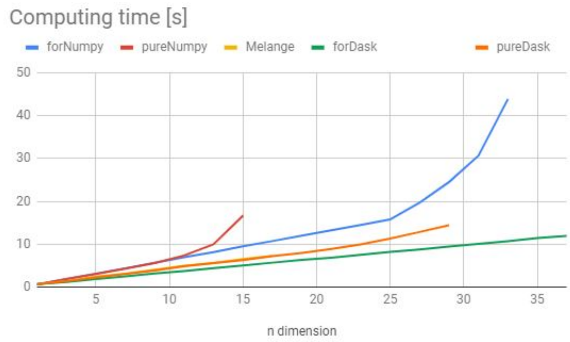

FIGURE III-2 – Graph of the computing time

When analysing this graph it is possible to see some interesting things : First, the DASK compute evaluation in a low-memory footprint maintains, so when comparing DASK and numpy codes that uses both a loop "for" it is possible to see that DASK codes executes for more wavelength (the same thing for codes that does not use the loop "for"). Another interesting thing to analyse is that when comparing both DASK codes (or even both numpy codes), the codes that uses a loop "for" are more scalable. Furthermore these loop codes run faster than the codes without the loop.

DASK profiling

After making the study comparing numpy and DASK libraries, it was made a deeply study about DASK library. The main study was in analysing DASK codes performance. DASK propose different tools for analyzing the performance of a code. The library DASK.diagnostics is the simpler tool for analysing the resources used during a code execution, the four python objects : Profiler, CacheProfiler, ResourceProfiler and Progress Bar allows the programmer know more about the functions that are been executed, the memory used during the execution and the resources used.

Another tool DASK proposes is the computational graph, been able to see this graph allow the programmer to know the relationship between how the code is structured and how it will be executed. This is also an opportunity to analyse the graph to see possibles places in the code that the programmer can change to make more parallelization in order to make the code more efficient.

Finally DASK proposes an interactive dashboard. This dashboard has some windows that shows all the information of the dask.diagnostics library (in the same time as the execution of a program), the information of the workers and more details of the computational graph.

Cluster Fusion

The last part of the internship consisted in studying the execution of DASK codes in a cluster, and also analysing its performance. The cluster used in this internship to execute the DASK codes was the fusion cluster of the Moulon Mésocentre. This is a cluster created by two institutions : CentraleSupélec and the École normale supérieure (ENS) Paris-Saclay. Fusion is a cluster composed by hundreds of nodes for parallel distributed memory computing, some server for parallel shared memory and sequential computing and some resources for general-purpose computing on graphics processing units (GPGPU) and 2 visualization nodes [13].

The first tests in the cluster was made via Portable Batch System (PBS) files. The PBS files contains the instructions for using the cluster to execute a code. An example of a PBS file is show in the image III-3 below.

```
#!/bin/bash

#PBS -l walltime=00:20:00
#PBS -l select=2:ncpus=8:mem=32gb
#PBS -M fabioeid.morooka@l2s.centralesupelec.fr
#PBS -m be
#PBS -N dask_seq
#PBS -j oe
#PBS -P decowska

# Load necessary modules
module load anaconda2/2019.03

# Activate anaconda environment
source activate myenv

# Move to directory where the job was submitted
cd $PBS_O_WORKDIR

# Run python script
WienerFilter.py
```

FIGURE III-3 – PBS file example

Some of the instructions set in this file is the resources that it will be used to execute this program, in this case the code "WienerFilter.py" will be executed using 2 blocs of resources (each one with 8 CPUs with 32GB of memory) in a maximum time of 20 minutes. The job will be entitled as "dask_seq" and one email will be sent to "fabioeid.morooka@l2s.centralesupelec.fr" in the beginning and in the end of the execution.

It was possible to execute DASK codes with the PBS file, but unfortunately it was not possible to follow the code execution, so making performance analysis was not easy. Then it was made a meeting with the team responsible for the cluster to talk about this issue.

After knowing more about the cluster, it was possible to make a solution for following the execution of the code using the resources of the cluster. The solution was to open a jupyter notebook in the cluster and use a shh tunnel to navigate in the jupyter notebook. This notebook is where the python code will be executed, but using the resources from the cluster since it was opened inside the cluster.

III.2 Dask Tutorial - Misson in Rennes

The main goal of this internship was not only to study python DASK, but also to make a tutorial guide [12] for future students and researchers who wants to start using DASK. This tutorial was made little by little since the beginning of the internship, because its goal it is not only to present DASK, but also shows everything from python from python installation until the execution of python codes in a cluster (using fusion as practical example).

This tutorial also contains some practical examples did during the internship, because to understand some DASK tools it must see some examples and its analysis. This tutorial was firstly done for the mission did in Rennes in the end of August. This mission had the objective to present DASK to the IETR laboratory and also to talk about the internship progress. Furthermore it was possible to see the efficiency of this tutorial, since the intern Axel Monot who was in IETR had the opportunity to learn dask from this guide made.

During this mission it was discussed the viability to use DASK in the SKA project with the ATOS BULL company represented by Nicolas Monier and Erwan Raffin and also the viability to use DASK in the PREESM software developed by the IETR laboratory. So it generate some insigths to the continuation of the internship.

Finally, a second version of the tutorial was made after finishing the internship, with the last studies made, mainly in the cluster fusion.

IV The universe : The future of SKA and final points

This last part is to conclude the internship, then it will be concluded by talking about the relationship between the obtained results and the perspectives in the beginning. Next it will show the personal and professional learning and finally the perspectives of SKA.

IV.1 Perspectives

The SKA project is just beginning, this internship was just an initial study to see the viability of using DASK. From this initial study, it is now possible to explore more about some aspects of this python library in the SKA context.

First of all, one study that can be made to see the dask's behavior is to decrease the size of the image and then raise the number of wavelengths to see the limit of the program. Moreover the code should be tested with different images in each wavelength to see if the graph and the time to compute change.

Another point that can be explored is the reuse of dask graphs. Making the reconstruction of an image by the Wiener filter method is always the same for different images, so if it was possible to generate the dask graph only one time and then reuse it to every other image reconstruction would decrease the time to execute the program. During the internship it was made a short study of the possibility to reuse the graph with Axel Monot, but this study lead to the conclusion that reusing it is complicate because dask graphes uses random tokens for each function and variable in the program. Nevertheless, making a deeply study to understand more about the construction of dask graphes and how to change it and reuse it might be interesting for SKA project.

Finally, for some dask's tools, the documentation in the internet is not extensive, sometimes there is just the introduction part of the tool in the main dask website [5], but it is not enough to understand deeply to really see its application in the project SKA, so there is the possibility to try to make contact with Matthew Rocklin, the responsible of DASK in anaconda and talk with him to understand this details.

IV.2 Personal and Professional Experience

Thanks to this internship, I was able to learn more about a topic that I was always interested in work : Astronomy. I had the opportunity to learn and develop my skills in python programming and HPC programming. Furthermore I also developed my soft skills as communication (both spoken and written) and autonomy and research abilities.

During this time I had the opportunity to work in collaboration with another laboratory in France, the IETR. This allowed me to interact with different people in both L2S and IETR laboratories. By doing this collaborative work I was able to improve my ability to adapt to different situations and therefore increase my network. One example of this collaborative work is the presentation of dask library at Rennes, by doing so I developed my french communications skills. I also had the opportunity to work with an intern at IETR, Axel Monot. We discussed about dask and exchange our knowledge of the SKA project. Finally by making the tutorial of dask python I was able to develop my french and my communication skills, since the tutorial is made for people who have already had an experience with python programming or not.

Since the beginning of this internship I have coded in python, so I was able to enhance my python skills acquired during my hole life, including the course in 4EII. I also had the experience to work with HPC by working with the Fusion cluster, so I was able to learn more about computer architectures and to work with the cluster team at CentraleSupelec.

Despite the individual project, I was always sharing and working with others Interns and PhD students in L2S, so I learned more about their research and projects. Moreover, working in a university laboratory was an opportunity to have more autonomy in my work, since I had a flexible working time.

IV.3 Final considerations

The project SKA is a complex project that will change how astronomers and researchers study the universe. They will have more data and more detail of what is happening in the universe we live in almost in real time or even what had happen in the origin of the universe. So SKA is a huge project that will revolutionize the astronomy and engineering field.

Dask is an amazing library created by anaconda that has a lot of tools and interactive code execution to see and analyse the execution of a program. But in the context of SKA, dask may not be the best solution for doing the image reconstruction, at least when using the Wiener filter method, for other methods/algorithms (like huber) that may be studied to see its application.

This was just an initial study of dask, applied in the SKA context. To really understand the details and how does dask works, it would be a nice opportunity to study dask in the DDFACET program [11] to see if dask can have a better performance in a more complex program.

IV.4 Acknowledgments

This work would have never been done if it was not for my family that has given me all support I need to study in France. I would like to thank my university from Brazil, the State University of Campinas (Unicamp) and my university in France, the National Institute of Applied Sciences (INSA) of Rennes for given me the opportunity to follow the double degree program in France.

Thanks to the Laboratory of Signals and Systems for receiving me in its facilities and for the opportunity to make this internship. Every employee in the laboratory were very receptive and thoughtful. I would also like to thanks the cluster fusion team for helping me to understand more about fusion.

I also thanks the IETR laboratory for giving me the opportunity to have a cooperative work during the internship, my tutor at INSA, Jean-François Nezan who was always available to talk and discuss the progress of the internship, Alex Monot who I could discuss and learn more not only about SKA, but also about DASK, and every teacher, PhD student who received me in Rennes.

I would like also to thank both my tutors, François ORIEUX and Nicolas GAC for receiving me in the Inverse Problems Group and for been there during all the internship to guide me and discuss the relation between my internship and the SKA project. I also thanks Thi thanh NGUYEN, Diakite DAOUDA, Ralph ABIRIZK and Chghaf MOHAMMED for helping me during my internship.

Finally, a special thanks to my friends who were always with me in my day life in Paris, that made the internship an incredible opportunity to not only have a professional experience, but also a personal one. Friends who were all over the world, in Brazil, Japan and Europe, and show me that no matter the (physical) distance, we can help each other.

Bibliography

[1] Rioja, Maria J., *Advanced Topic in Astrophysics, Lecture 3* [Slides]. Recovered 20 may 2019, in https://www.web.uwa.edu.au/__data/assets/pdf_file/0008/791126/rioja-icrar-lecture3_small.pdf.

[2] Smirnov, O., *The radio interferometric data challenge : From MeerKAT towards the SKA* [Slides]. Recovered 20 may 2019, in http://sites.ieee.org/sips2018/files/2018/11/meerkat.pdf.

[3] Giovannelli, J.-F., La méthode de Wiener-Hunt en déconvolution de signaux et d'images.

[4] Gary, D. E., *Physics 738, Radio Astronomy : Lecture #6*. Accessed 21 may 2019, in https://web.njit.edu/~gary/728/Lecture6.html.

[5] Dask Development Team, 2016, Dask : Library for dynamic task scheduling. Accessed 22 may 2019, in https://dask.org.

[6] Jupyter Notebook, Cours *Fundamentals of Radio Interferometry* du master NASSP. Accessed 23 may 2019, in https://github.com/ratt-ru/fundamentals_of_interferometry.

[7] Jupyter Notebook, Dask Tutorial. Accessed 23 may 2019, in https://github.com/dask/dask-tutorial.

[8] Site, Site personnel Matthew Rocklin. Accessed 26 may 2019, in https://matthewrocklin.com/.

[9] Site, Site du projet SKA. Accessed 26 may 2019, in https://astronomers.skatelescope.org/.

[10] Site, Livre blanc SKA. Recovered 29 may 2019, in https://arxiv.org/pdf/1712.06950.pdf

[11] Github, DDFACET program. Recovered 26 august 2019, in https://github.com/saopicc/DDFacet

[12] Morooka, Fabio Eid, *DASK tutorial*. Recovered 30 september 2019, in https://www.overleaf.com/read/zkmxphmfvqrw

[13] Site, Moulon Mesocentre, documentation of the cluster Fusion Accessed 10 september 2019, in http://mesocentre.centralesupelec.fr/