university of groningen / faculty of mathematics and natural sciences / kapteyn astronomical institute

Bachelor Thesis

Astronomy

7th November, 2012

The 21-cm Power Spectrum Sensitivity of Current and Future Epoch of Reionization Arrays

Author:
R.C. Joseph

Supervisor:
Prof. Dr. L.V.E. Koopmans

University of Groningen

Kapteyn Astronomical Institute

front picture: The picture is a photograph of the LOFAR-superterp located in the Netherlands. The black tiles are the HBA antennas and grey "spots" are the LBA antennas. Courtesy to Astron.

Abstract

The Dark Ages of the Universe ended with the formation of the first structures. The formation of these first structures was accompanied by the heating and the subsequent reionization of the intergalactic medium. The Epoch of Reionization (EoR) is thought to hold the key to how and when the first galaxies formed. A promising probe to study this epoch is the redshifted 21-cm line of neutral hydrogen. In this thesis, we investigate the sensitivity of current and future low frequency radio telescopes to measure the redshifted 21-cm power spectrum during the Epoch of Reionization. In our comparison of current arrays we find for a bandwidth of 10 MHz, integration time of 1000 hr at redshift $z = 10$ and scale size $k = 0.1\,\mathrm{Mpc}^{-1}$, that LOFAR outperforms MWA and PAPER by an order of magnitude in power spectrum sensitivity. This comes mostly from LOFAR's larger collecting area. MWA and PAPER compensate their lack of collecting area by increasing their field of view (FoV) and making their arrays compact. This however shifts their sensitivity to smaller k-values (i.e. larger scale modes), which are more relevant for cosmology than reionization studies. We also find that the LOFAR-AARTFAAC extension can increase the sensitivity of LOFAR by a factor ~ 5 for $k < 0.1\,\mathrm{Mpc}^{-1}$, below redshift $z = 12$. This comes from the combination of the FoV of a single tile and the total collecting area of the LOFAR-Superterp, which contains 288 antenna tiles in 12 stations. The LOFAR-Superstation in Nancy, which will consist of 96 stations each containing 19 LBA dipoles, has half an order of magnitude more sensitivity than even the LOFAR-AARTFAAC system in LBA mode, making it one of the most promising instruments for very high redshift 21-cm EoR observations ($z > 15$) in the coming decade until the SKA comes online. We finally calculate the sensitivity of different SKA layouts, finding that compact arrays are the most sensitive, but that station size should be carefully considered since this constrains the range of measurable scale variations. Concentrating a large collecting area, e.g. $1\,\mathrm{km}^2$, in only few stations could even lead to less power spectrum sensitivity than current arrays, due to the small field of view and increased sample variance. We also find that increasing the number of antennas increases the sensitivity on all scales, as expected. But the maximum number of antennas is constrained by computational power, hence we need to balance collecting area and stations size within the limits of the correlator.

Contents

1 Introduction **3**
- 1.1 A short biography of the Universe . . . 4
- 1.2 Epoch of reionization . . . 4
- 1.3 Observing the epoch of reionization . . . 5

2 Radio Astronomy **7**
- 2.1 The Universe in brightness temperature . . . 7
- 2.2 Beginners guide to radio interferometry . . . 8
  - 2.2.1 The visibility . . . 8
  - 2.2.2 UV-plane . . . 10
  - 2.2.3 Fourier relations . . . 10
- 2.3 The 21-cm Power Spectrum . . . 11
- 2.4 Errors on the Powerspectum . . . 12
  - 2.4.1 System noise . . . 13
  - 2.4.2 Sample Variance . . . 14
  - 2.4.3 Total noise . . . 15
  - 2.4.4 Angular Averaged Power Spectrum . . . 15

3 The Code **16**
- 3.1 General assumptions . . . 16
- 3.2 Antenna distribution . . . 17
- 3.3 Baseline Distribution . . . 18
- 3.4 Theoretical Power Spectrum . . . 18

4 Results **20**
- 4.1 Comparison between Current Arrays . . . 20
- 4.2 The LOFAR-AARTFAAC and -Superstation Extensions . . . 23
  - 4.2.1 Reionization . . . 24
  - 4.2.2 Cosmic Dawn . . . 25
  - 4.2.3 Scaling relation . . . 26
- 4.3 SKA configurations . . . 28

5 Conclusion **30**
- 5.1 Summary . . . 30
- 5.2 Future Work . . . 31

6 Nederlandse Samenvatting **32**

7 Acknowledgements **33**

A Optimization Scheme **36**

B Parameter File **37**

C pyRadio **40**

Chapter 1

Introduction

Astronomers are no longer long bearded individuals who spend night after night looking through a copper telescope[1]. Pen and paper have been replaced by the CCD and the arm and eye have made way for auto guiders and tracking systems. Astronomy has changed over the centuries and so did the eyes of the astronomers. Astronomy is no longer limited to the visual spectrum, by going into space one can now detect high energy gamma rays, and by building enormous groups of antennas on earth we can observe low-energy radio waves.
Radio waves are the lowest frequencies electromagnetic waves can have and can become important when looking deep into the furthest "corners" of our Universe. Looking at these corners, light we receive is not the same as when it was emitted. Apart from propogation effects due to the matter between us and the source, the whole spectrum of the object is redshifted down to lower frequencies. This is due to the expansion of the Universe, and is named cosmological redshift. When looking at the early phases of the Universe, we see light (e.g. the CMB of HI 21-cm emission) which is redshifted towards the radio side of the electromagnetic spectrum. Radio observations can therefore play an important role in cosmology, which studies the universe as a whole: its origin, evolution and its ultimate fate. One of the evolutionary phases of the Universe predicted by cosmologists is called the Epoch of Reionization (EoR). This epoch is caused by the formation of the very first objects, and thus holds the key to how the structure we observe today is formed [Barkana and Loeb, 2001]. While radio arrays such as the Low Frequency Array (LOFAR)[2], the Precision Array for Probing the Epoch of Reionization (PAPER) [3] and the Murchison Widefield Array (MWA)[4] aim to be the first to observe this phase through redshifted 21-cm emission, a much more ambitious project called the Square Kilometer Array (SKA)[5] is under development.
One of the key science projects of SKA will be the epoch of reionization, and thus the question arises: what would be an optimal array configuration for this project? Studying the EoR can be done by observations of the hyperfine 21 cm transition of neutral hydrogen, a more precise formulation of the question behind this klein onderzoek is as follows: What is an optimal array configuration of SKA for 21 cm observations of the Epoch of Reionization?
The outline of the report is as follows: the remaining of Chapter 1 will give a short introduction on Reionization. Chapter 2 introduces some concepts from radio astronomy and the theory used in this research. Chapter 3 describes additional concepts necessary to implement the theory into a working code, which calculates the power spectrum sensitivity of different array configurations. Chapter 4 presents and discusses results from this research. Chapter 5 contains a summary of this report and a discussion for future work which can complement this research.

[1] The majority of astronomers no longer fits this discription, some might.
[2] www.lofar.org
[3] http://astro.berkeley.edu/ dbacker/eor/
[4] www.mwatelescope.org
[5] www.skatelescope.org

1.1 A short biography of the Universe

If one had the remote control of the Universe and pressed rewind, we would see the universe contract until it has shrunk into a point. This moment in time, when our whole universe compressed into a hot dense state, is called the Big Bang. Now press play to see the baby Universe expand into the version we see today.
Somewhere around $\sim 10^{-34}$ seconds after the Big Bang the expansion was exponential, a phase called *inflation*. Inflation caused the volume of the universe to increase dramatically, with a factor of $\sim 10^{26}$. The inflationary phase did not last very long, but even after this phase the universe continued to expand, although slower. Due to this expansion the temperature of the Universe decreased, giving rise to many processes which are studied by particle physicists today. Eventually these processes led to the formation of the particles which fill our Universe: bosons, leptons, neutrinos and more. The Universe is now about 10^{-3} seconds old and still cooling.
After the first second, the Universe reached the temperature and density required to combine protons and neutrons together to form form heavier atomic nuclei. This period of nuclear fusion is called *big bang nucleosynthesis*. Elements such as Helium and a small fraction of heavier elements such as lithium were formed in this period. However it was not until long before expansion cooled the Universe enough, to end nucleosynthesis and the Universe was left primarily filled with hydrogen and helium. Because of the large photon energy density, nuclei and electrons were not be able to form neutral atoms. High-energy photons would knock the electrons out of their bound states into a soup nuclei and electron. The free electrons kept colliding with photons, scattering, transferring energy from the electrons to the photons and back. This caused the Universe to be nearly in perfect thermodynamic equilibrium creating the Blackbody spectrum which can be observed today in every direction on the sky: the Cosmic Microwave Background (CMB). However the energy of this spectrum was trapped between the electrons until the expansion rate of the universe surpassed the scattering rate of the electrons. Some 350 000 years after the Big Bang the Universe cooled such that electrons could recombine with nuclei to form atoms. The photons, no longer hindered by electrons, could travel freely through the Universe while carrying information of its state at the time of last scattering. [Ryden, 2002]
The universe became almost completely neutral and dark at this stage, but it was far from uninteresting. Apart from the (sub)atomic physics that played its part during the first few seconds of the Universe, a process dominated by gravity also played a role on the larger scales. After the inflationary phase of the universe an imprint of the inflation process was left on the energy density of the Universe. The quantum fluctuations in the energy field that caused the Universe to expand by a factor of $\sim 10^{26}$, were enhanced and caused fluctuations in the mass density. This created overdense and underdense regions. When the Universe became cold enough, these regions of over- and under density grew by the accretion of matter under the influence of gravity. Until the density in the overdense regions was high enough to form the first objects such as stars.

1.2 Epoch of reionization

During the EoR, hydrogen which was neutral during the Dark Ages, was ionized again. This all started with the primordial fluctuations in density after the inflationary phase. Gravity causes overdense regions of dark matter to collapse to form even more dense regions. When the gas pressure lost its battle against gravity, due to cooling, this becomes a runaway process until the gas is so dense, that nuclear fusion starts and a star is born. Possibly collapse proceeded even further leading to the formation of a black hole, however this is still unclear. These first objects would not simple relight the Universe but, if massive enough, which they probably were,would emit ultraviolet photons. Photons in the UV and higher energies can ionize hydrogen atoms and this is why this epoch is not called the re-enlightenment. The main change was in the state of the gas, which changes from the neutral phase to the ionized phase.
The stars formed, were most likely clustered together in the first galaxies, which would emit the combined ionizing photon flux to ionize the gas between those galaxies. However, it is still unclear

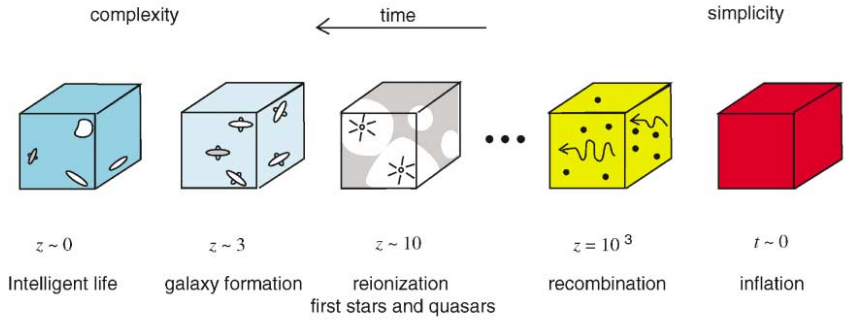

Figure 1.1: History of the universe [Barkana and Loeb, 2001]

how many of these ionizing photons would actually escape to ionize the intergalactic medium. Other (partial) explanations for the reionization process may be the formation of (mini)quasars. Since black holes are very efficient in converting mass into energy, they could emit even more ionizing photons. Also even more exotic theories have been developed such as the self-annihilation of dark matter particles.[Barkana and Loeb, 2001]
What processes caused reionization ultimately determined the structures we see around us today. The formation of the very first stars caused a change in the chemical composition of the Universe. Metals[6] were formed in these first stars, since the Big Bang Nucleosynthesis only gave us hydrogen, helium and a small fraction of heavier elements. These stars would later on release their metals when they ended as a supernova, after which it would be re-used in the formation of a new generation of galaxies, stars, and eventually planets and bachelor students.
Now we have collected motivation to observe the EoR using the redshifted 21-cm line as our probe. To summarize, the EoR holds the formation of the first stars. Their formation proceeded in an environment very different from today, since it was metal poor. Also the first clusters of stars lead to the formation of the first galaxies. And these in turn clustered into galaxy clusters and super clusters. In other words, the structures we observe all around us were seeded in this epoch, and the key to understanding the formation of structure must then also lie there. How reionization proceeded and how neutral hydrogen was distributed can also tell us something about the cosmology which dominates our universe today.

1.3 Observing the epoch of reionization

There are many scenarios which explain reionization and structure formation, so we need observations to lift the degeneracy between these scenarios. There have already been several indirect observations of the EoR which put constraints on the models. One observation is called the Lyman-α forest, an effect in the spectra of high redshift quasars. The Lyman-α forest is the collection of all absorption lines due to the neutral hydrogen between the quasar and the observer. Photons are continuously redshifted when they are hurdling towards us over cosmological distances. At a certain redshift their wavelength has been stretched to 121.6 nm. If this redshifted photon travels trough a cloud of neutral hydrogen, there is a high probability that it will be absorbed in a so called Lyman-α transition. This effect is called the Gunn-Peterson effect. Depending on the density of neutral hydrogen the absorption line will be deeper. Measuring the depth of the Lyman-α absorption gives information on the amount of hydrogen at a certain redshift in the direction of that specific quasar. The Lyman-α forest show that reionization took place before $z = 6.5$, however the state of the intergalactic medium at higher redshifts is unclear since there is

[6]Astronomy 101: Metals indicate every element heavier than Helium.

no detection of a quasars beyond redshift $z = 7.1$. [Zaroubi, 2012b]
Other observations come from the polarization of the CMB, due to scattering by electrons, the current temperature of the intergalactic medium and several other measurements. More details can be found in Zaroubi [2012b].
These observations are however not direct, and as such can only constrain general features of the EoR. If we want to measure the reionization process directly and in detail we need a probe which observes the intergalactic medium itself. We need some characteristic to observe neutral hydrogen. Observing neutral hydrogen in galaxies can be done by using the 21-cm transition of atomic hydrogen. The transition is caused by the spin-transition of the electron from parallel to anti-parallel. However when radiation reaches us, it will redshifted to longer wavelengths.

$$z = \frac{\lambda_{obsv} - \lambda_{emit}}{\lambda_{obsv}} = \frac{\nu_{emit} - \nu_{obsv}}{\nu_{obsv}} \tag{1.1}$$

Using the constraints given by Lyman-α forest, which shows reionization to take place beyond a redshift of$z = 6$, the 21-cm line has been redshifted to a wavelength of 1.5 m or 202 MHz. However this part of the radio spectrum is affected by radio-transmitters, RFI and the ionosphere. On top of that emission from the sky is dominated by synchrotron emission from electrons interacting with the magnetic field of the galaxy. These factors pose a big challenge for EoR projects and demand enormous sensitivity of radio arrays to measure this weak spin transition in a mix of galactic emission, ionospheric and instrumental distortions and radio broadcasts. To really image the hydrogen in the Universe we have to resort to an even more ambitious project such as the SKA. However this does not make current arrays useless for EoR observations. Instead of mapping the hydrogen distribution during the EoR, current arrays will employ the 21-cm power spectrum to observe global processes during reionization. This will be discussed in chapter 2.

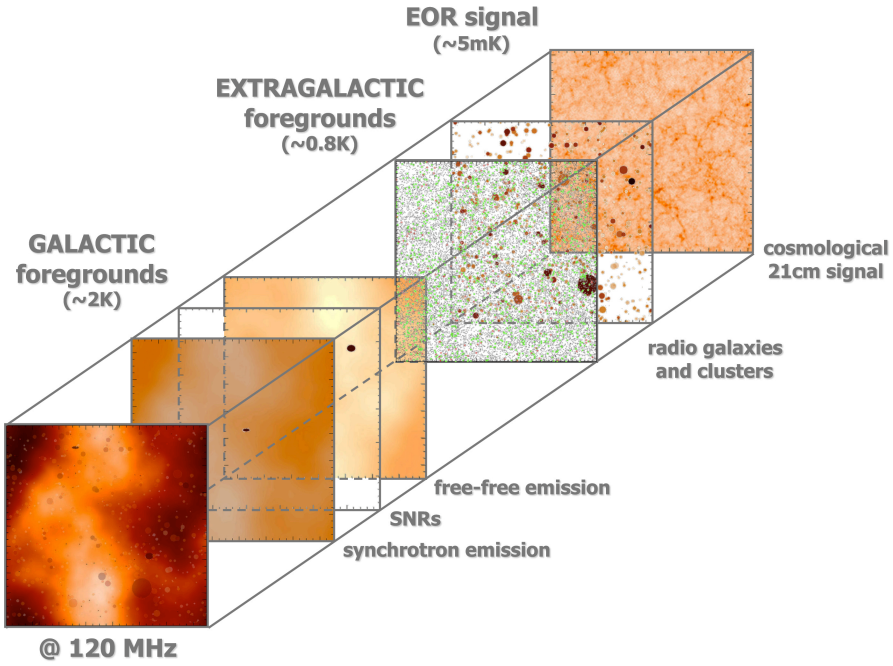

Figure 1.2: Reionization signal hidden in the foregrounds. Courtesy to V. Jelic.

Chapter 2

Radio Astronomy

To observe reionization of neutral hydrogen in the universe the 21-cm power spectrum is used, instead of direct imaging of hydrogen itself. This chapter discuss a few basic concepts behind radio interferometry and the 21-cm power spectrum itself. Using those basic concepts the errors on the 21-cm power spectrum will be calculated. The results will be used to find an optimal array configuration which reduces the errors for the desired EoR observations.

2.1 The Universe in brightness temperature

The emission of a photon at a wavelength of 21-cm is caused by a small transition in which the orbiting electron changes its spin from parallel to anti-parallel as depicted in figure 2.1. The amount of 21-cm radiation we receive from a patch of neutral hydrogen, depends on how many atoms are in the parallel state n_1 and how many in the anti-parallel state n_0. More atoms in state n_1 will lead to more 21-cm emission when they decay to the ground state n_0. The distribution of energy levels is given by the Boltzmann distribution.

$$\frac{n_1}{n_0} = 3\, e^{-T_{21}/T_{spin}} \tag{2.1}$$

Where $T_{21} = 68\,\mathrm{mK}$ is related to the energy corresponding to the 21-cm transition via $T_{21} = E_{21}/k_B$, k_B is the Boltzmann constant. T_{spin} is the temperature corresponding to a certain ratio of energy levels, and not specifically the temperature of the gas. This is because the n_1 level can also be populated in other ways than collisions, of which the rate is determined by the gas (kinetic) temperature. The factor 3 comes in because of the degeneracy of the excited state n_1. The spin temperature T_{spin}, depends on several energy sources, the details can be found in [Field, 1958]. In short it depends on the temperature of the CMB photons T_{CMB}, which can be absorbed. The kinetic temperature of the gas T_k, which determines the collisional excitations. And on the amount of Lyman-α photons, which we can assign a temperature T_α. Lyman-α excitations can lead to a decay to the n_1 state. Different ionizing sources will lead to different spin temperatures because they lead to a different kinetic temperature T_k and Lyman-α temperature T_α, the CMB temperature is globally the same. So the spin temperature is related to the physical processes which drive reionization.
The photon energy detected by radio arrays is low enough that we can assume $h\nu \ll kT$, i.e. the photon energy is much lower than the equilibrium temperature. Which means we can take the Rayleigh-Jeans limit of the Planck function, resulting into

$$I_\nu = \frac{2\nu^2}{c^2} k_B T. \tag{2.2}$$

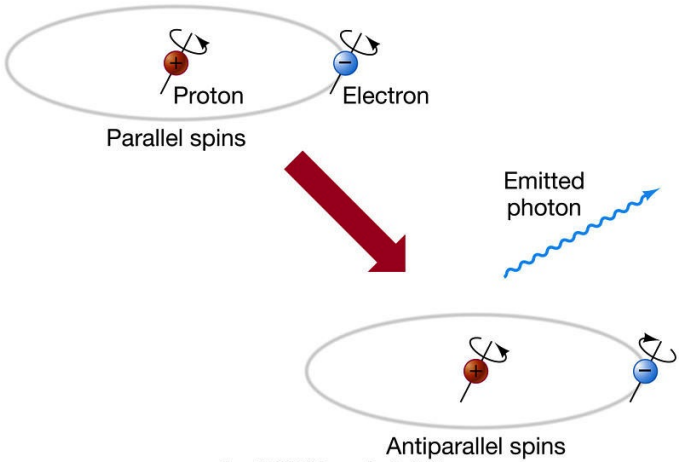

Figure 2.1: The 21-cm line transition. Courtesy to Pearson Prentice Hall.

As seen in equation (2.2) the brightness depends linearly on temperature, so we can relate our brightness I_ν directly to a temperature.. In the case of the 21-cm line we can relate it directly to the spin temperature. From now on we will use Kelvin as our unit of choice to express brightness, because this relates to physical processes behind the 21-cm emission.

2.2 Beginners guide to radio interferometry

A radio interferometer works quite differently than an optical telescope. An optical telescope is just a big photon bucket, and works in the same way as our eye does. It counts the number of photons from a given location and this is what we call the power. An interferometer is not just one (radio)telescope, it consists of several and together they observe the sky as if they were one radio telescope, with equivalent size. Interferometry refers to interference, which is the interaction between waves. So in order to do an interferometric measurement the wavelike nature of light should be captured. These signals can then be combined digitally to recreate the effect of a single imaging telescope.

2.2.1 The visibility

The purpose of an antenna receiver in an interferometer is to collect two characteristics of the incoming lightwave: amplitude and phase. When we assume the wave is sinusoidal[1], we have all the information we need to describe the incoming signal:

$$V(t) = E\,cos(\omega t + \phi). \tag{2.3}$$

We will now assume we have two antennas which form our interferometer and there is a point source infinitely far away emitting a signal (figure 2.2). One of the antennas will receive a delayed signal with respect to the other. If $\mathbf{b}$ is the vector from antenna 1 to antenna 2, and $\mathbf{s}$ is the vector pointing in the direction of the source, the signal has to travel an extra distance $\mathbf{b} \cdot \mathbf{s}$ to reach antenna 2. The vector $\mathbf{b}$ is our baseline vector, hereafter baseline.[2] So the time delay is $\tau = \frac{\mathbf{b}\cdot\mathbf{s}}{c}$. The different signals from the antennas are multiplied in a cross-correlator and averaged over an appropriate time-interval. The resulting signal is given by:

[1] This example is a simplification of the actual EM wave, which is far more complex and better discribed by a gaussian random field.

[2] Baselines are measured in wavelengths, the distance separation between two antennas divided by the wavelength of observation.

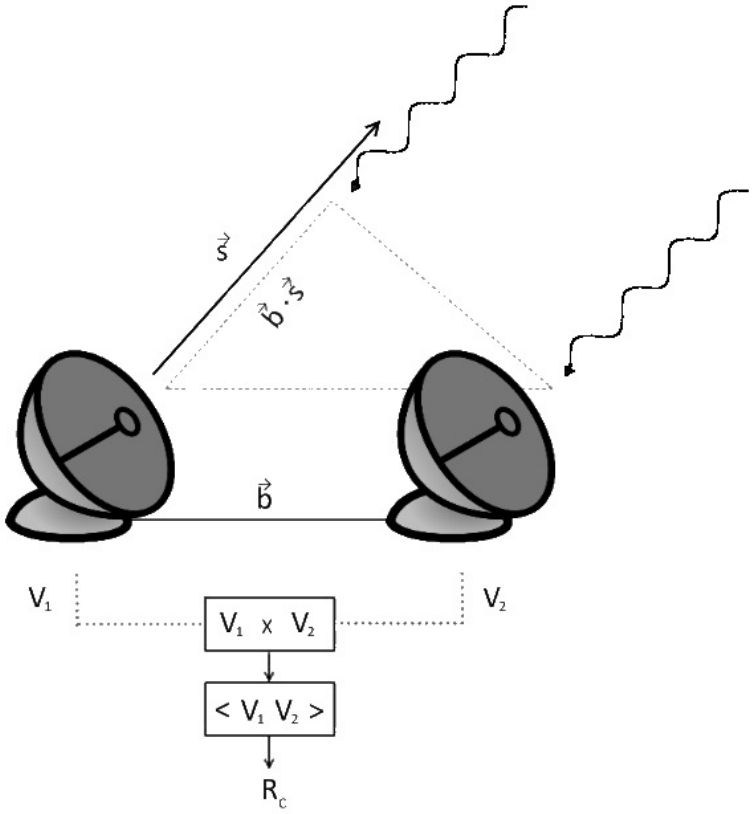

Figure 2.2: Schematic overview of an interferometer.

$$\begin{aligned} R_C &= \langle V_1 \cdot V_2 \rangle \\ &= \langle E^2 \, cos(\omega t) \, cos(\omega t - \omega \tau) \rangle \\ &= P \, cos(\omega \tau). \end{aligned} \tag{2.4}$$

This resulting signal depends only on the received power P, the baseline orientation and the source direction. Determining the location of sources on the sky does not depend on our pointing accuracy, but on our clock which measures the time delay τ. Aside from the normal delay due to antenna separation, we can also artificially shift one of the input signals with 90°. This will result in a sine as output instead of a cosine. If we then follow the same path through the correlator, the resulting signal per baseline is:

$$R_S = P \, sin(\omega \tau) \tag{2.5}$$

We can extract two components of the source signal from the baseline, an even part (the cosine) R_C and an odd part (the sine) R_S. This is the information our interferometer has given us about our source, together they form the "complex visibility" which is defined as [Perley, 2011]

$$V = R_C + iR_S. \tag{2.6}$$

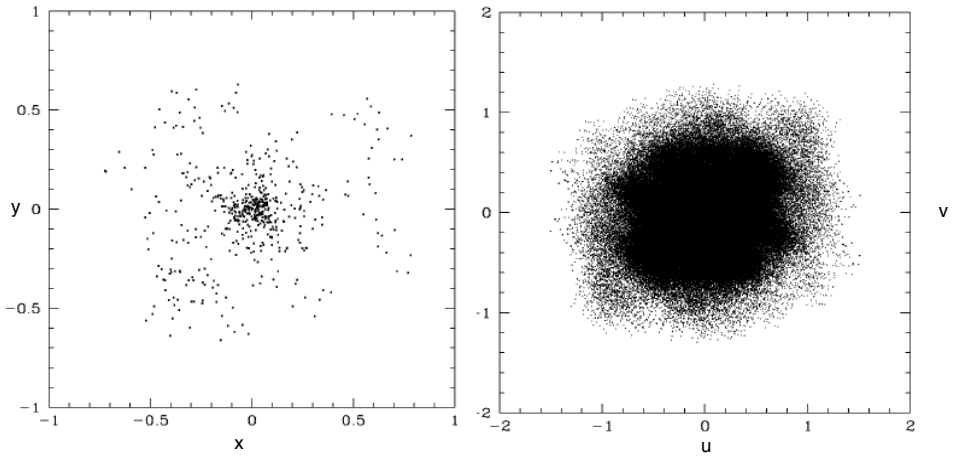

Figure 2.3: Hypothetical antenna lay-out (left) for MWA and corresponding baselines in the uv-plane (right). The number of baselines is given by $N_b = \frac{1}{2}N_a(N_a - 1)$, so high amount antennas leads to a very high amount of baselines. Every point in the uv-plane will collect data about our source. [Zaroubi, 2012a]

2.2.2 UV-plane

Instead of just one antenna pair, astronomical interferometers are built out of several antennas, N. Each one of these antennas can form a baseline with one of the other $N - 1$ and produce a visibility. We need to define a so called uv-plane, which collects all of the baseline vectors $\mathbf{b} = u\hat{\mathbf{i}} + v\hat{\mathbf{j}}$, present in the array[3]. Where we define the coordinates with respect to our source in direction $\mathbf{s}$. Every baseline will represent a point in this uv-plane, of which the size is defined by the size of the antenna. Each of these points measures a visibility, so for an array consisting out of several antennas we can define a visibility function: V(u,v). See figure 2.3 for an antenna lay-out and its corresponding uv-plane.

2.2.3 Fourier relations

Now we know what complex information an interferometer gives us about the source, but this has to be related to more natural real quantities. Normally we measure the brightness $I(l, m, \nu)$,where l, m are some coordinates on the sky, and ν indicates the observing frequency. The translation from visibility to brightness is given in the title of this subsection: Fourier transform. The intensity and the visibility are Fourier conjugates, related via equation (2.7).

$$I(l, m, \nu) = \iint V(u, v, \nu)e^{2\pi i(ul+vm)}\mathrm{d}l\mathrm{d}m \tag{2.7}$$

In order to get as much information as possible about the brightness distribution, we need to sample the visibility function as densely as we can. This is why radio-astronomers talk about uv-coverage: gathering a lot of visibilities at different coordinates in the uv-plane. This can be done in two ways:

- Using a large number of baselines, since every single baseline will be a point in the uv-plane. The problem with this strategy is the difficulty with cross-correlating a large amount of data, which takes a lot of computer power.

[3]If the baselines are extended such that we have to take into account the curvature of the earth, $\mathbf{b} = u\hat{\mathbf{i}}+v\hat{\mathbf{j}}+w\hat{\mathbf{k}}$ since all the antennas do not lie on the same plane.

- A computationally less expensive method is integration time. Since the earth rotates, the direction vector of the source on the sky **s** will change. Our baseline vectors are defined with respect to the direction of the source, in other words they also change. So by taking a long integration time, the baselines will move in the uv-plane, covering tracks.

The combination of both, a sufficient number of antennas and integration time, will lead to a good sampling of the uv-plane and thus a better measurement of the source brightness. [Perley, 2011]

2.3 The 21-cm Power Spectrum

Now we have a basic idea of how we can image the hydrogen in the Universe. This gives us the possibility of truly mapping reionization in 3D (tomography), since we can Fourier Transform from uv-coordinates to sky coordinates, 2 dimensions. But we can also measure the 21-cm line over different frequencies. This gives us information on distance (or time), when using the appropriate cosmological formulas to convert redshift to distance, the 3rd dimension. However one major setback comes from the foregrounds. As said before the 21-cm signals from reionization travel a vast distance before reaching the array and when it arrives it will be outmatched by signals from extragalactic sources, galactic synchrotron radiation and distorted by the ionosphere. This causes our signal-to-noise ratio to be rather low and therefore 3D-imaging of reionization will be difficult. A second approach is a statistical detection which employs the power spectrum which we will discuss now.

The radio array which we use to observe the EoR, has observed a small volume[4] of the Universe at the redshift corresponding to the observing frequency. We can calculate the width of the volume with simple trigonometry. This width is given by the Field of View (FoV) of each station, in other words the angular size, multiplied with the distance x to the source of emission. The depth of the volume corresponds to the bandwidth of the observation, since this can also be translated to distance [McQuinn et al., 2006]. Since the density of hydrogen and the spin temperature of neutral hydrogen will vary at different places in the volume, the brightness of the 21-cm signal $I(\mathbf{x})$ will fluctuate as well. These fluctuations can be described by the 21-cm power spectrum. The very short recipe for extraction of the 21-cm power spectrum out of a volume filled with fluctuating 21-cm signals is as follows.

- Take a 3D-Fourier transform of the volume, this would expand all these fluctuations in terms of waves with a wave vector $\mathbf{k}$. From now on we will refer to these wave vectors as Fourier modes.
- This has given us $\tilde{I}(\mathbf{k})$, the magnitude of the brightness fluctuations as a function of wave vector $\mathbf{k}$. In other words how much the brightness variation there is at different scales. The tilde indicates Fourier transformed brightness, which should remind us of the visibility: also a Fourier transform of the brightness but only in 2D.
- Then we take the absolute square $|\tilde{I}(\mathbf{k})|^2$.
- After which we take the average over shells at a radius $\|\mathbf{k}\|$

Resulting in the magnitude of the fluctuation in 21-cm brightness at a given scale length, which gives us information about the fluctuations in density and ionization fraction at different scales. Important to note: $k = 2\pi/\lambda$, small k correspond to large scales and large k correspond to small scales.[Lidz et al., 2007] [Morales and Wyithe, 2010]
Because the 21-cm power spectrum is hidden in the Fourier representation of the sky, we do not need to transform the visibility function back to brightness as a function of sky coordinates. However the interferometer only performed a 2D-Fourier transform over the spatial coordinates. So there is only one dimension that still needs to be transformed: frequency direction. The

[4]When looking at cosmological scales you may call this megaparsec sized volume small.

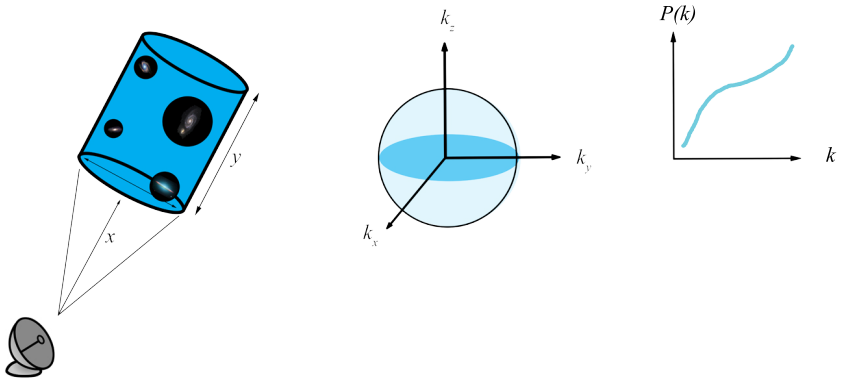

Figure 2.4: Very schematic overview of how the 21-cm power spectrum is extracted from a hypothetical volume filled with ionized bubbles.

transform will take place over the range of the bandwidth, since this correspond to the size of our observed volume. [Geil, 2011]

$$\tilde{I}(\mathbf{b},\eta) = \int V(\mathbf{b},\nu) e^{-2\pi i \nu \eta} \mathrm{d}\nu \tag{2.8}$$

The next step would be to relate our coordinates to k-space coordinates which correspond to scale sizes within the volume, after which which we can start the averaging. However while a Fourier transform of a hypothetical volume of the universe has a completely filled Fourier space corresponding to that volume, the Fourier representation given by the interferometer does not. An interferometer can only sample the Fourier transform of the observed volume partially. As shown in figure 2.2, the number and location of baselines determines how the uv-plane was filled. By using the rotation of the earth the baselines moved and we could fill the uv-plane even more. But it would be impossible to fill the uv-plane completely, we cannot place antenas closer than their diameter and we cannot place antennas infinitely far away. Because we cannot fill our uv-plane completely it is impossible to fill our k-space completely, which is just a coordinate transformation from u, v to k_x, k_y. This means the array has to be optimized in a way, that the interesting shells of $\mathbf{k}$ are filled with enough baselines to get an accurate determination of the 21-cm power spectrum.

We were forced to employ the 21-cm power spectrum due a low signal-to-noise ratio created by foregrounds. However in the process of extracting the power spectrum from from three-dimensional information we lost detailed information, leaving only global information. This could be a problem because different models may produce more or less the same 21-cm power spectrum. However the benefits from the power spectrum comes from the following. Because the 21-cm power spectrum is an average over a certain number of points within a shell at k, the error on that part of the power spectrum is reduced by $1/\sqrt{N_c}$, where N_c is the number of points measured in that shell.

2.4 Errors on the Powerspectum

To find an optimal array configuration we need to estimate the errors on the 21-cm power spectrum, and its dependence on design parameters of the array. This section will discuss the main equations on which the code is based. The derivation of the errors on the power spectrum will follow the formalism described in [McQuinn et al., 2006].

2.4.1 System noise

From antenna theory we can deduce the root mean square (r.m.s.) noise per visibility per antenna pair:

$$\Delta V^N = \frac{\lambda^2 T_{sys}}{A_{\text{eff}}\sqrt{\Delta\nu t_0}} \tag{2.9}$$

Where T_{sys} is the system temperature, i.e. the antenna temperature and temperatures of foregrounds. A_{eff} is the effective area of a single antenna, which will be discussed in chapter 4. And $\Delta\nu$ is the frequency resolution. This has to be transformed in the frequency direction to get into the same Fourier representation as the 21-cm power spectrum. If we assume $\Delta\nu \ll B$, we can replace the Fourier integral by a sum.

$$\tilde{I}^N(\mathbf{b},\eta) = \sum_{i=1}^{B/\Delta\nu} V^N(\mathbf{b},\nu)e^{2\pi i\nu_i\eta}\Delta\nu \tag{2.10}$$

To find the error on the power spectrum we follow more or less the same path of its extraction, by taking the square and taking the average over the bandwidth. Which gives us the average error on our power spectrum within the observed volume.

$$C_{ij}^N(\mathbf{b}_i,\mathbf{b}_j) = \langle \Delta\tilde{I}^N(\mathbf{b}_i,\eta)\Delta\tilde{I}^N(\mathbf{b}_j,\eta)^*\rangle \tag{2.11}$$

We can rewrite this using equation 2.10.

$$\begin{aligned} C_{ij}^N(\mathbf{b}_i,\mathbf{b}_j) &= \Big\langle \Big[\sum_{n=1}^{B/\Delta\nu} V^N(\mathbf{b}_i,\nu_n)e^{2\pi i\nu_n\eta)}\Delta\nu\Big]\Big[\sum_{m=1}^{B/\Delta\nu} V^N(\mathbf{b}_j,\nu_m)e^{2\pi i\nu_m\eta)}\Delta\nu\Big]^*\Big\rangle \\ &= \Big\langle \Big[\sum_{n=1}^{B/\Delta\nu} V^N(\mathbf{b}_i,\nu_n)e^{2\pi i\nu_n\eta)}\Big]\Big[\sum_{m=1}^{B/\Delta\nu} V^N(\mathbf{b}_j,\nu_m)e^{2\pi i\nu_m\eta)}\Big]^*\Big\rangle(\Delta\nu)^2 \\ &= \Big\langle V^N(\mathbf{b}_i,\nu_1)V^N(\mathbf{b}_j,\nu_1)^* + ... + V^N(\mathbf{b}_i,\nu_x)V^N(\mathbf{b}_j,\nu_x)^* \\ &\qquad + C(\mathbf{b}_i,\mathbf{b}_j,\nu_n,\nu_m)\Big\rangle(\Delta\nu)^2 \end{aligned} \tag{2.12}$$

Since the frequency resolution $\Delta\nu$ is constant, we can take it out of the sum and the averaging. Writing out the product of the sums gives the last line, where the complex exponential drops out for the same frequencies ν_n, leaving only the noise products for baselines i and j. $C(\mathbf{b}_i,\mathbf{b}_j,\nu_n,\nu_m)$ are cross terms between different baselines i and j and different frequencies ν_n and ν_m, i.e. they do contain complex exponentials.

$$\begin{aligned} C_{ij}^N(\mathbf{b}_i,\mathbf{b}_j) &= \frac{B}{\Delta\nu}(\Delta V^N)^2(\Delta\nu)^2\delta_{ij} + \Big\langle C(\mathbf{b}_i,\mathbf{b}_j,\nu_n,\nu_m)\Big\rangle(\Delta\nu)^2 \\ &= (\Delta V^N)^2 B\Delta\nu\delta_{ij} \end{aligned} \tag{2.13}$$

We reach the final line by noting that noise signals from different baselines are uncorrelated, the mean of their product is zero. While for correlated signals the mean of the product equals the r.m.s squared. The mean of the cross terms is also zero. Substituting our expression for the r.m.s. noise per visibility in equation (2.13) gives us

$$C^N(\mathbf{b}) = \Big(\frac{\lambda^2 BT_{sys}}{A_{\text{eff}}}\Big)^2\frac{1}{Bt_0}. \tag{2.14}$$

The kronecker delta is dropped because we consider a single baseline. However we are talking about an array consisting out of several antennas, all measuring a visibility for a certain $\mathbf{b}$. This visibility is only measured for a certain amount of time because our array is turning and its baseline coordinates change. The baseline vector $\mathbf{b}$ is related to our fourier modes $\mathbf{k}_\perp$, so we can translate

our noise per baseline, to a noise per Fourier mode as follows. Using the definition $\mathbf{k}_\perp = 2\pi\mathbf{b}/x$ and using the fact that visibility is measured for a time given by equation (2.15).

$$t_{\mathbf{k}} \approx \frac{A_{\text{eff}} t_0}{\lambda^2} n[x\|\mathbf{k}\| \sin(\theta)/2\pi] \tag{2.15}$$

Where λ is the wavelength corresponding to the observing frequency. The first term A_{eff}/λ^2 determines what area the antenna samples in the uv-plane. The number density of baselines $n(|\mathbf{b}|)$ comes in, because we want to know how many baselines sample this value of $\mathbf{k}$. Where we have assumed the array is circular symmetric, which gives the circular symmetric baseline distribution $n(b)$. But since we are interested in the noise per Fourier mode, u has been rewritten as a function of k. The angle θ is the angle between our line of sight (LOS), see figure and the direction of the Fourier mode $\mathbf{k}$, see figure 2.4. Our array can only measure projections of Fourier modes which fit the covered area in the uv-plane. In other words modes smaller than the shortest baseline and longer than the longest baseline cannot be measured. This boundary is defined by the baseline distribution $n(\|\mathbf{b}\|)$, which will be discussed in chapter 3. Using equation (2.15) we can rewrite the system noise per Fourier mode as

$$C^N(\mathbf{k}) = \left(\frac{\lambda^2 B T_{sys}}{A_{\text{eff}}}\right)^2 \frac{1}{B t_{\mathbf{k}}}. \tag{2.16}$$

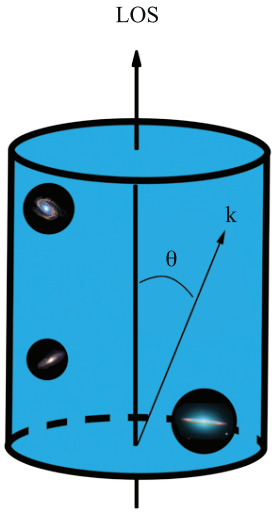

Figure 2.5: The angle between our line of sight (LOS) through our volume and Fourier modes k. This volume is in an earlier stage of reionization than the one depicted in figure 2.3.

2.4.2 Sample Variance

This research has focussed primarily on reducing the system noise, since observations are currently in the noise dominated region. Therefore the sample variance will be discussed in a short descriptive manner. The sample variance can be understood as follows. Since we are measuring a finite volume of space, defined by the FoV, and bandwidth. We can only sample Fourier modes a certain number of times, depending on how many can fit into the k-space volume. The large modes (i.e. small k-values) corresponding to scales on the same order of the survey volume, will be sampled only once or twice. While the smaller modes, corresponding to small scales, will be sampled much more often. So sample variance is not an actual noise like the system noise, but

rather an uncertainty set by the finite number of measurements (even if the S/N-ratio is high). The sample variance is given by:

$$C^{SV}(\mathbf{k}) = P_{21}(\mathbf{k})\frac{\lambda^2 B^2}{A_{\text{eff}} x^2 y} \tag{2.17}$$

Where x is the width of the volume and y is the depth of volume, corresponding to the bandwidth.

2.4.3 Total noise

We can sum the two error contributions, system noise and sample variance, to get the total error per observed point in k-space. Because the power spectrum is an average over the number of measured points at a certain scale k, we have to determine first what the number of point in that shell is. As mentioned before this depends on the number of baselines, because they determine the points which are sampled in the visibility function V(u,v). However the approach used in the derivation of the error assumed we have actually sampled the complete surface between the shortest and longest baseline. This is due to the assumption of a circular symmetric baseline distribution. This assumes there is some density of baselines at every point in the uv-plane, so every point within the baseline range is sampled. In other words the $N_b(N_b - 1)$ baselines have been smeared out over an area in the uv-plane.
Now we have to determine the number of points in some annulus with thickness Δk at k. We can only sample as much as fit inside the annulus, and this depends on the size of each point. This size of each point is determined by size of the observed Fourier volume and equals $(2\pi)^3/\mathcal{V}$, where $\mathcal{V} = x^2 y \lambda^2 / A_{\text{eff}}$ in real space.
The number of cells in a spherical annulus is given by equation

$$N_c(k,\theta) = 2\pi k^2 \sin(\theta)\Delta k \Delta\theta \frac{\mathcal{V}}{(2\pi)^3}. \tag{2.18}$$

The term $2\pi k^2 \Delta k$ determines the volume of the annulus, typical values for $\Delta k = 0.5k$. But we have to take into account that our observational volume is finite. The baseline distribution takes care of modes whose projection does not fit the baseline ranges. However modes much larger than the depth of our volume can have a projection on the uv-plane which fits inside the baseline distribution. This is possible when the angle between LOS and the mode k is small enough. To exclude these Fourier modes, the number of points N_c is set to zero when, the k no longer fits inside the volume: $2\pi/kcos(\theta) > y$. The total error on the power spectrum becomes,

$$\delta P_{21}(k,\theta) = \frac{1}{\sqrt{N_c}} \frac{A_{\text{eff}} x^2 y}{\lambda^2 B^2} [C^{SV}(k,\theta) + C^N(k,\theta)] \tag{2.19}$$

Where the factor $1/\sqrt{N_c}$ is introduced because the error is reduced by the number of cells which measured the power spectrum.

2.4.4 Angular Averaged Power Spectrum

Since the interest was in the fluctuations as a function of k-scale only, we still have to average the error over angle θ. Which is done using

$$\delta P_{21}(k) = \left\{ \sum_{\theta} \left[\frac{1}{\delta P_{21}(k,\theta)} \right]^2 \right\}^{-\frac{1}{2}} \tag{2.20}$$

Which results in our final estimation on the errors on the 21-cm power spectrum. Equation 2.20 will be used to determine the sensitivity of an array for power spectrum measurements, as a function of its design.

Chapter 3

The Code

This chapter outlines the implementation of theory into a code which calculates the sensitivity of an array given initial parameters, after which this can be used to optimize the array for power spectrum measurements. The code is written in the language Python[1]. The actual calculating work such as interpolations, integrations and optimization is done by functions within the module SciPy. Because Python is open source, the use of python makes it also possible for distribution among interested within the astronomical community. I will not discuss the program in great detail, for the program itself I refer to the appendix, but I will discuss several components of the program and their relation to theory, and why several assumptions have been made.

3.1 General assumptions

For the system temperature T_{sys} we assume we can neglect the temperature of the antennas and only consider the sky temperature, this can be easily modified by adding an antenna temperature component if required. The sky is dominated by synchrotron emission from the galaxy, whose brightness temperature is approximated by a power law as a function of frequency, given in equation (3.1)[2]. Although this varies along different lines of sight, being at its strongest near the Galactic center and weaker out of the Galactic plane. [Jelic et al., 2008]

$$T_{\mathrm{sky}} = 400 \left(\frac{\nu}{150\,\mathrm{MHz}}\right)^{-2.55} \mathrm{K} \tag{3.1}$$

For dipoles we cannot really talk about a physical area, when looking at figure 3.1 this becomes clear. However, we can define an effective area which is determined by the sensitivity of a dipole. The sensitivity of a dipole has some angular dependancy. This angular sensitivity pattern is called a beam, the size of this beam and wavelength of observation define the so called effective area of the dipole. For the effective area we assume the following function.

$$A_{\mathrm{eff}} = A_p \left(\frac{\nu}{120\,\mathrm{MHz}}\right)^{-2} \mathrm{m}^2 \tag{3.2}$$

Where A_p is the physical area of the dipole, which is related to the dimensions of the antenna. Equation (3.2) shows that the dipoles are less sensitive to higher frequencies.
We also assume a flat universe with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $\Omega_{\mathrm{k}} = 0$ and a Hubble constant of $H_0 = 70\,(\mathrm{km/s})/\mathrm{Mpc}$. These cosmological parameters are used to calculate the distances to the source of emission x and the depth of the observed volume y. Which are calculated using equation (3.3). [Hogg, 2007]

[1] http://www.python.org/
[2] This equations holds for frequencies below 200 MHz.

Figure 3.1: LOFAR HBA Dipole, courtesy to R. van den Brink.

$$D = \frac{c}{H_0} \int_{z_{\min}}^{z_{\max}} \left(\Omega_m \left(1+z\right)^3 + \Omega_k \left(1+z\right)^2 + \Omega_\Lambda\right)^{-\frac{1}{2}} \mathrm{d}z \tag{3.3}$$

3.2 Antenna distribution

The distributions of the antennas has great influence on the array sensitivity. Since the locations of the antennas with respect to each other determine what baseline lengths are present and as such what sensitivity the array has on different scales lengths. A small dense array will have antenna stations close to each other, i.e. short baselines, and thus sensitivity on large scales. While a large diffuse array will have long baselines and thus more sensitivity on small scales. A sensible choice for a function which spans several scales is a power law. So we assume the following antenna distribution, a core area which has a constant array density, and an outer area in which the density follows a power law as a function of radius. The core area will provide us with sensitivity on large scales, and the outer area will provide us with sensitivity on the smaller scales. This function is of course continuous, while a true array will have a discrete distribution, but this is a good approximation to first order when discussing arrays consisting of a large number of antennas. For this distribution the antenna density is given by [Geil, 2011]:

$$n(r) = \begin{cases} n_c & 0 \leq r \leq r_c \\ n_c \left(\frac{r_c}{r}\right)^p & r_c < r < r_{\max} \\ 0 & r > r_{\max} \end{cases} \tag{3.4}$$

Where n_c is a normalization constant such that the integral over the surface of the array results in the total number of placed antennas N.

$$N = \iint n_a(r) \sin(\phi) \mathrm{d}\phi \mathrm{d}r \tag{3.5}$$

The crucial parameters for this power law antenna distribution are: the core radius r_c,the outer radius $r_{\max}$ and the slope p. The outer radius determines the maximum baseline length. The slope will determine how fast the sensitivity decreases over k and how dense the core array will be. A steeper slope will lead to more antennas in the core area. To prevent overfilling of the core an upper limit has been placed where the total area of the antennas in the core $n_c\, A_{\text{eff}}$ cannot exceed the physical area of the core $\pi\, r_c^2$.

3.3 Baseline Distribution

For a given antenna distribution the corresponding baseline distribution is given by the convolution integral[Geil, 2011]

$$n_b(b,\nu) = C_b(\nu) \int_0^{r_{\max}} 2\pi r \; n_{\mathrm{a}}(r)\mathrm{d}r, \int_0^{2\pi} n_{\mathrm{a}}(\mathbf{r} - \lambda\mathbf{b})\mathrm{d}\phi \tag{3.6}$$

Where $C_b(\nu)$ is a frequency dependent normalization constant such that the integral over the baseline distribution equals the total number of baselines $N_a(N_a - 1) = \int N_b(b,\nu)d\mathbf{b}$. By taking the convolution integral, the result is again a continuous function. A secondary effect is the creation of baselines below d_a/λ, where d_a is the antenna diameter. These short baselines are physically impossible, since it is not possible to place two antennas closer than twice their antenna radius. This already implies that while a larger antenna station has more collecting area, it makes it impossible for the array to measure large scale fluctuations, because of the smaller FoV. In order to correct for these artifacts, the code removes all baselines below the limiting length. After which the distribution is renormalized to match the total number of baselines. The convolution is calculated numerically, therefore it can also handle all types of antenna distributions as long as they are circularly symmetric. Because the convolution is quite time consuming, the baseline distribution is calculated once per realization. The results are further used by one-dimensional interpolation in the **b** direction. An interpolation in the frequency direction is replaced by selecting the nearest frequency.

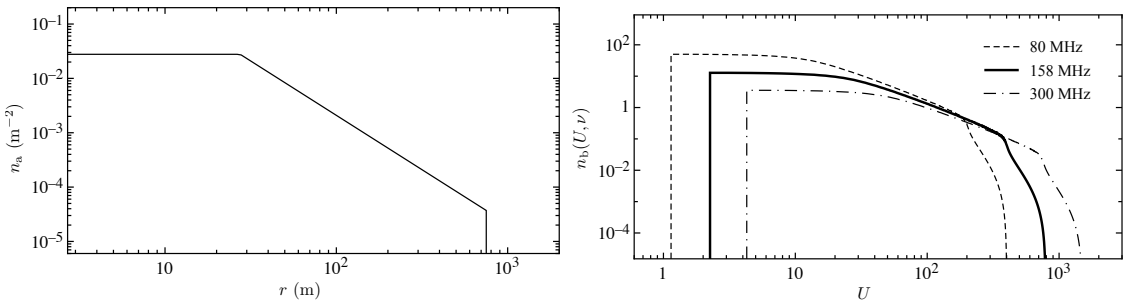

Figure 3.2: The antenna distribution (left) and the baseline distributions (right) for an MWA like instrument. [Geil, 2011]

3.4 Theoretical Power Spectrum

For calculation of the cosmic variance and optimization of the array configuration a theoretical power spectrum was provided by Prof. Dr. S. Zaroubi. The simulated power spectra are created using the reionization code 21cmFast [Mesinger et al., 2010], which produces 21 cm brightness boxes with dimensions of 400 Mpc. The power spectrum is extracted from these brightness temperature boxes with the aid of a FastFourierTransform routine in IDL.
The power spectra were provided for redshifts z= 12,11,10,9.5,9 and 8.5, and sensitivity calculations for intermediate redshifts is done by interpolation. Except for frequencies outside this range when the nearest frequency will be used. The range in scale lengths is $-1.5 \leq \log k \leq 0.8$. The minimum k is defined by the size of the box and the maximum k is defined by the resolution of the simulation. Other power spectra can be used in the code by simply replacing the input file containing the data. Figure 3.3 shows the power spectrum for several redshifts.

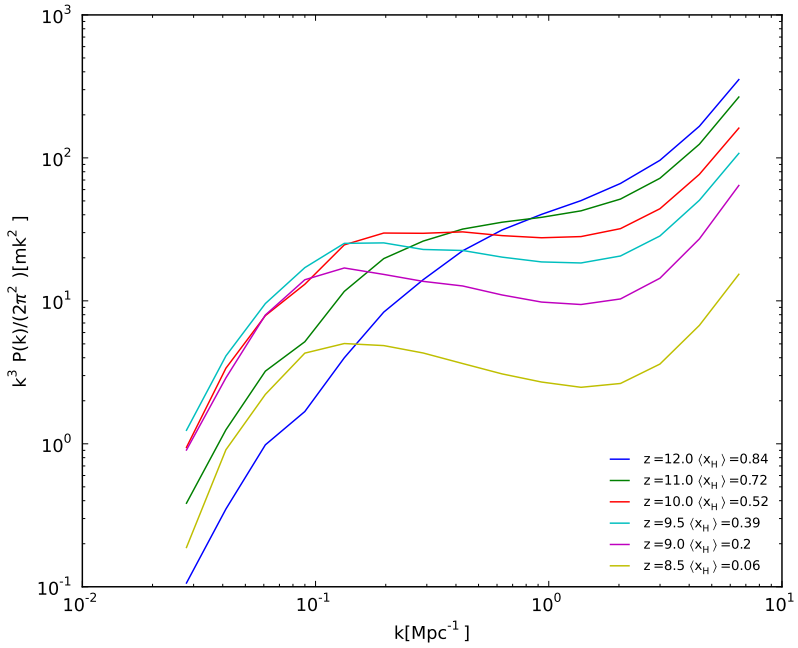

Figure 3.3: 21-cm power spectrum produced by 21cmFast, $\langle x_H \rangle$ indicates the mean fraction of neutral hydrogen at the corresponding redshift z.

Chapter 4

Results

After implementation into a code, which passed several test against literature [McQuinn et al., 2006], an analytic case and against a similar code developed by the group of Garrelt Mellema, the next step was to run calculations using parameters of current arrays. Over the years a number have been published on several arrays, but due to de- and rescopes or other factors, several changes have been made in the final array designs. To obtain a more updated state of affairs we first made a comparison between the arrays: PAPER, MWA and LOFAR. We also looked at implementation of two LOFAR-extensions for power spectrum measurements, LOFAR-AARTFAAC and the french LOFAR-Superstation (LSS). Calculations were also made for several SKA lay-outs. As future research we plan to use an optimization routine, to find the optimal SKA lay-out for power spectrum measurements. The results of these calculations will be presented in this chapter.

4.1 Comparison between Current Arrays

There are several arrays trying to measure the 21-cm power spectrum of reionization. But all arrays employ a different configuration strategy, hence each array will have a different k-regime at which its sensitivity is optimal. For the comparison the sensitivity of MWA, PAPER and LOFAR were calculated.

PAPER located in South Africa, is a radio array with a design focused on a high number of antenna stations, i.e. baselines. PAPER is build out of 128 station, each with a collecting area of $1.5^2\pi\ \mathrm{m}^2$. These stations are distributed uniformly in an area with a radius of 150 m, except in a central cavity with a radius of 10 m [Jacobs et al., 2011]. This creates a dense array, with a high number of baselines inside a small area of the uv-plane. The large FoV of each individual antenna gives PAPER instantaneous sensitivity to large scale fluctuations.

MWA located in Australia, follows a similar strategy as PAPER. MWA also has a large number of stations, 112 to be precise, creating a large number of baselines. Each antenna station has a physical collecting area of $14.5\ \mathrm{m}^2$, and are more spread out than the stations in the PAPER configuration. MWA consists of a central region with a uniform distribution of stations and a less dense outer region. This outer region has a power law distribution with an index of 2 [Beardsley et al., 2012]. The increase in collecting area gives MWA a higher sensitivity, within the ranges of the FoV. By placing stations in a more extended area, the MWA contains somewhat longer baselines and thus sensitivity on smaller scales.

LOFAR located in Europe (we only consider the core area) is built out of 48 stations, which is less than half of the stations of PAPER and MWA station. This is however compensated by the size of each station. Each station has a collecting area of $16^2\pi\ \mathrm{m}^2$, which is much larger than the station size of PAPER and MWA. LOFAR is also a more diffuse array then the before

mentioned arrays with its stations distributed uniformly in a central region and a power law, also with index 2, decreasing outer region. The central region, a.k.a. the "Superterp", has a radius of 150 m and the outer region stretches out until 1500 m. The array parameters can be found in table 4.1. The antenna lay-out and the corresponding baseline distribution of these arrays is displayed in figure 4.1.

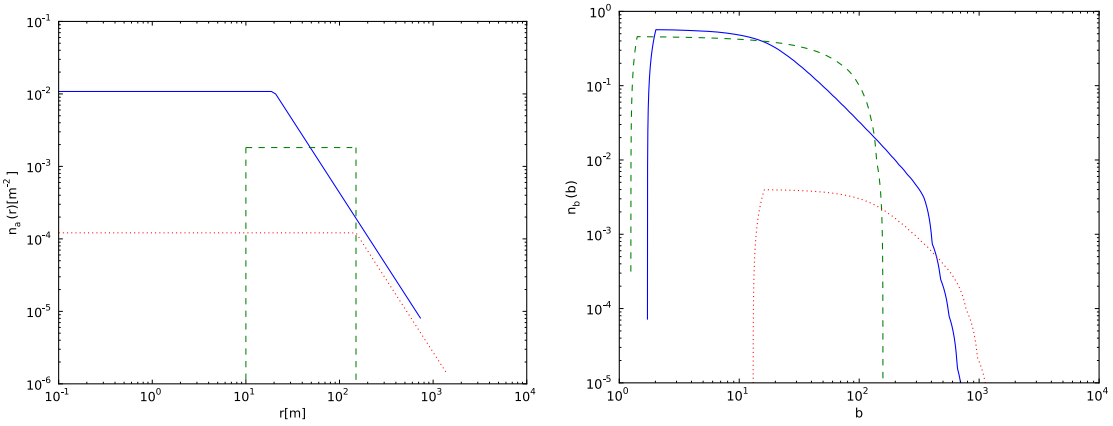

Figure 4.1: The antenna distribution (left) and the baseline distributions (right) at 150 MHz for LOFAR: red dotted line, PAPER: green dashed line and MWA: blue solid line.

Array	A_{p} (m^2)	N_{ant}	p	r_c (m)	r_{max} (m)
PAPER	7.1	128	0	150	-
MWA	14.5	112	2	25	750
LOFAR	804	48	2	150	1500

Table 4.1: Array parameters

Using the parameters for the arrays given in table 4.1 and a binning size of $\Delta k = 0.5k$, we calculated the power spectrum sensitivity at redshift $z = 8, 10$ and 12, for an observation time of 1000 hours and a bandwidth of 10 MHz. The results are shown in figure 4.2 and tabulated in table 4.2. The sample variance and system noise are tabulated separately because sample variance is dependent on the power spectrum model, while the system noise is universal.

From figure 4.2 we can conclude the following:

- LOFAR is the most sensitive array and PAPER is the least sensitive array at the relevant scales for the 21-cm power spectrum.
- MWA and PAPER are unlikely to measure the 21-cm power spectrum, while LOFAR can measure it partially until a redshift of $z = 10$.

Looking at figure 4.1, we see MWA and PAPER outmatch LOFAR in antenna and baseline density by orders of magnitude. We also note that PAPER and MWA contain much shorter baselines than LOFAR. This is due to LOFAR's station size, which limits the minimum baseline to $b_{\mathrm{min}} = d_{\mathrm{a}}/\lambda$ where d_{a} is the antenna diameter. Despite these orders of magnitude difference in baseline densities, LOFAR still has at least half a magnitude more sensitivity, see figure 4.2. This seems to

indicate that sensitivity lies in station size, since this is where LOFAR exceeds the other arrays. The other arrays have much more shorter baselines and larger FoV's, making them more interesting for cosmological studies rather than EoR observations. Since the more detectable part of the 21-cm power spectrum is around $k \sim 0.2\,\mathrm{Mpc}^{-1}$.
So it seems sensitivity is easily gained by increasing the collecting area of a station, rather than increase of the amount of stations. MWA and PAPER are most likely not able to measure the 21-cm power spectrum, while LOFAR can until a redshift of about z=10.

	$z = 8$		$z = 10$		$z = 12$	
Array	δP_{21}^{SV} $[10^{-3}\,\mathrm{mk}^2]$	δP_{21}^{N} $[\mathrm{mk}^2]$	δP_{21}^{SV} $[10^{-3}\,\mathrm{mk}^2]$	δP_{21}^{N} $[\mathrm{mk}^2]$	δP_{21}^{SV} $[10^{-3}\,\mathrm{mk}^2]$	δP_{21}^{N} $[\mathrm{mk}^2]$
PAPER	3.45	46.0	19.0	194	4.86	$1.10 \cdot 10^3$
MWA	4.94	38.1	27.2	176	6.97	$1.08 \cdot 10^3$
LOFAR	36.8	1.25	202	4.88	51.8	26.2

Table 4.2: Noise Calculation results at $k = 0.2\,\mathrm{Mpc}^{-1}$,$\delta P_{21}^{SV}$ is the sample variance component of the noise and δP_{21}^{N} is the system noise component. Power spectrum values at k = 0.2 are $k^3 P_{21}/2\pi^2 = 4.84$, 29.8, 8.58 mk^2, at respectively redshift z=8,10,12.

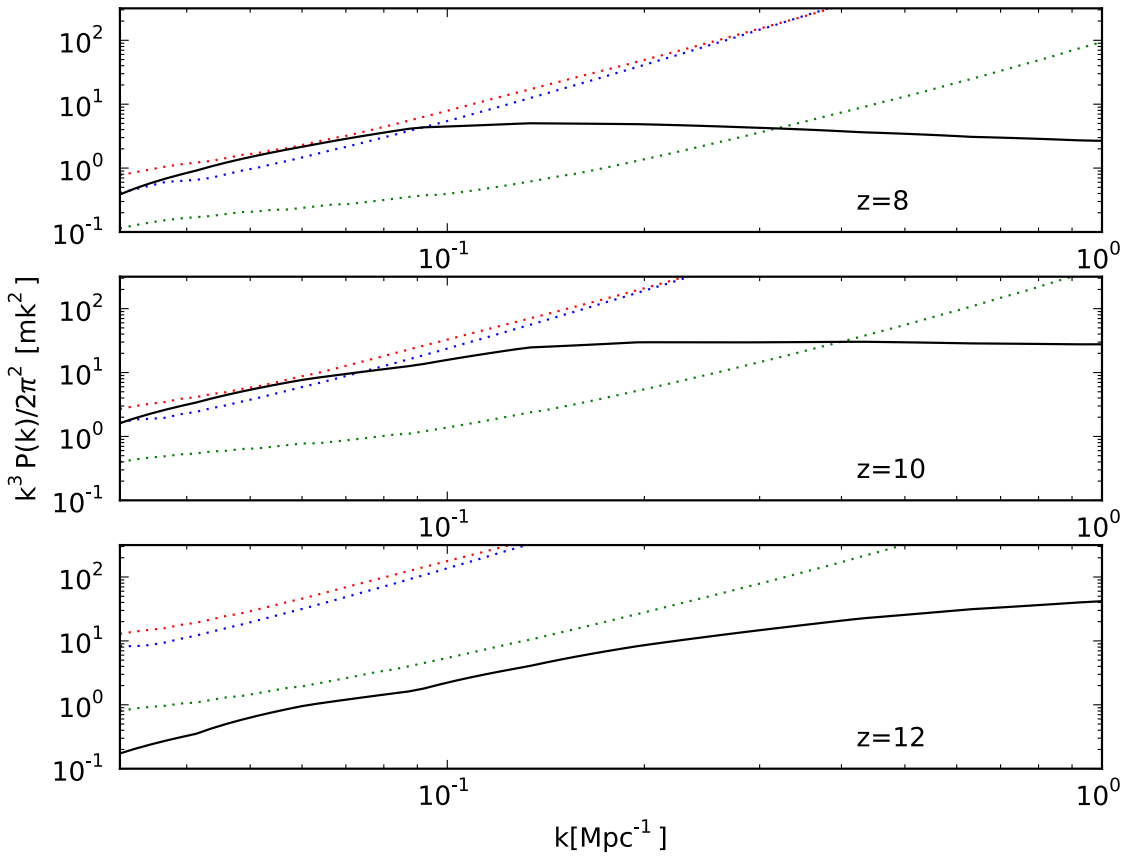

Figure 4.2: Comparison between MWA (blue dotted), PAPER (red dotted) and LOFAR (green dotted). The black solid line represents the power spectrum generated by 21-cmFAST.

4.2 The LOFAR-AARTFAAC and -Superstation Extensions

In the previous section we showed that all current arrays have marginal sensitivity to measure the 21-cm power spectrum. LOFAR has some sensitivity until a redshift of about $z = 10$, after which the 21-cm signal becomes too noisy. Even though construction of the LOFAR core has finished, the project is still evolving and undergoing exciting developments. We will consider two of these developments to investigate whether they can be used for 21-cm power spectrum measurements.

AARTFAAC stands for ASTRON Radio Transients Facility and Analysis Centre[1] and is an extension for the LOFAR-superterp. In standard operating mode a group of 48 HBA antennas or 48 LBA antennas operate as one antenna station. The superterp in Exloo contains 6 HBA-stations and 6 LBA-stations. The signals from each station are normally cross-correlated with another station producing the visibilities. The AARTFAAC system however cross-correlates the signals from all the individual antennas. The 288 LBA-stations or 288 HBA stations in the superterp can thus create a very dense array since the superterp has a radius of 150 m. [Prasad and Wijnholds, 2012]. The FoV increases because the station size has been reduced to the area of a single LBA or HBA receiver. This creates sensitivity on large scales. AARTFAAC was initially designed for large sky surveys in the search for transient sources, which requires the large FoV. The smaller station size might decrease the sensitivity. However compensation might come from the large baselines density, since the "station" number has increased by a factor of 24 thus a factor of ~ 2304 increase in number of baselines.

Superstation the LOFAR-Superstation is a project in Nancy. The superstation is planned to consist of a group of 96 LBA stations placed in an area with a radius of 175 m, which is comparable to that of the superterp. Although the number of individual elements is much lower than the 288 antennas of the LOFAR-AARTFAAC system, the collecting area of these antennas is much larger than the collecting area of a single antenna in the superterp, $300\,\mathrm{m}^2$ to be precise. The LOFAR-Superstation favors collecting area , while the LOFAR-AARTFAAC system favors a large number of baselines and wider FoV. However, the results from the previous section imply that collecting area is favored over number of baselines.

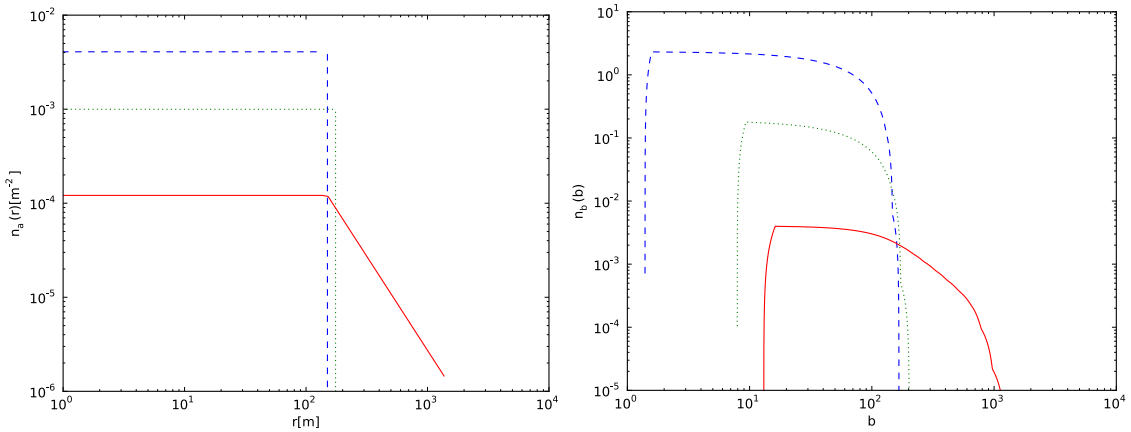

Figure 4.3: The antenna distribution (left) and the baseline distributions (right) at 150 MHz for LOFAR: red solid line, AARTFAAC: blue dashed line, Superstation: green dotted line.

[1] http://www.aartfaac.org/

Array	N_{ant}	$A_{\text{p}}\ [m^2]$	$r_{\text{c}}\ [m]$	p
Superstation	96	300	175	0
AARTFAAC	288	25	150	0

Table 4.3: Parameters for the AARTFAAC-system and Superstation project

4.2.1 Reionization

Using the parameters listed in table 4.3 and the parameters for LOFAR in table 4.1 we compared LOFAR and LOFAR-AARTFAAC in their capability of detecting the 21-cm power spectrum, at redshifts z=8,10 and 12 (LSS does not have HBA receivers). The results of this calculation are shown in figure 4.4 and detailed noise component values can be found in table 4.4. Looking at the results we can draw the following conclusions:

- LOFAR-AARTFAAC has a higher sensitivity than LOFAR on scales $k < 0.1\,\text{Mpc}^{-1}$
- LOFAR itself has a higher sensitivity than AARTFAAC beyond $k > 0.1\,\text{Mpc}^{-1}$.

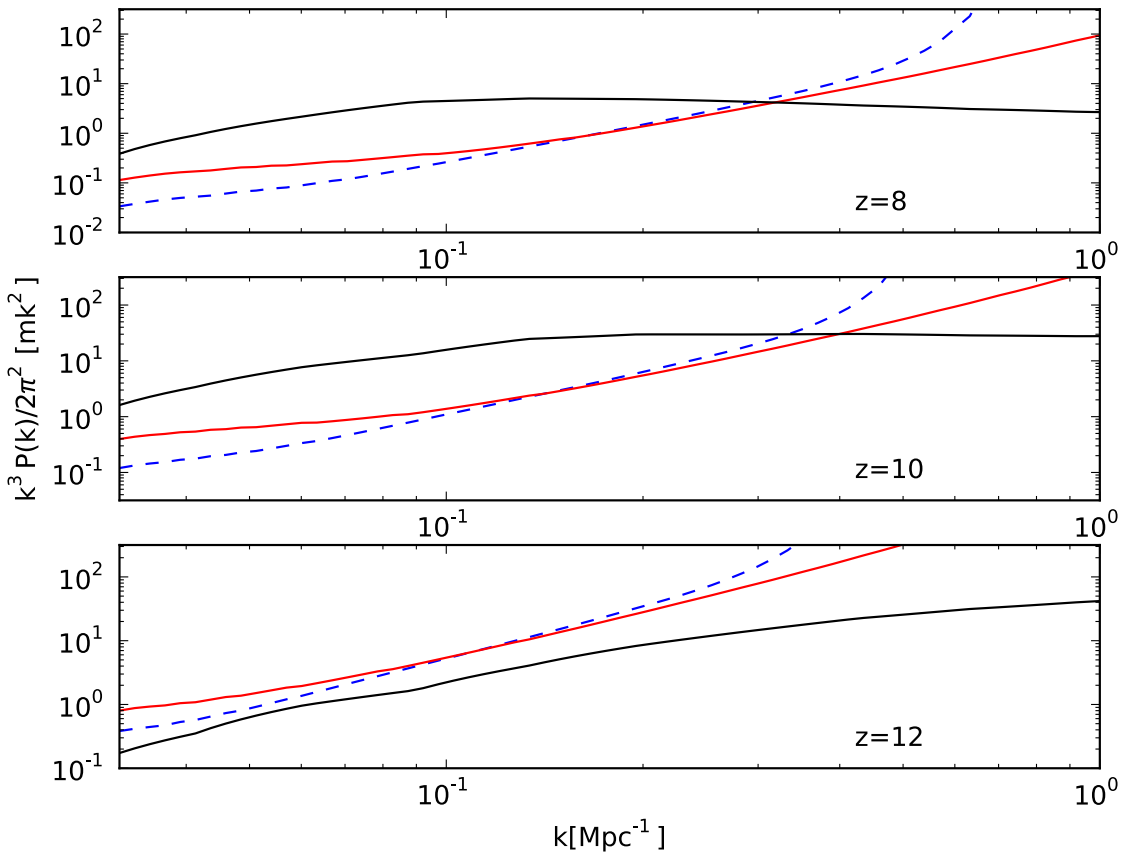

Figure 4.4: Comparison between LOFAR (red solid) and LOFAR-AARTFAAC (blue dashed). The black solid line represents the power spectrum generated by 21-cmFAST.

Array	$z = 8$		$z = 10$		$z = 12$	
	δP_{21}^{SV} $[10^{-3}\,\mathrm{mk}^2]$	δP_{21}^{N} $[\mathrm{mk}^2]$	δP_{21}^{SV} $[10^{-3}\,\mathrm{mk}^2]$	δP_{21}^{N} $[\mathrm{mk}^2]$	δP_{21}^{SV} $[10^{-3}\,\mathrm{mk}^2]$	δP_{21}^{N} $[\mathrm{mk}^2]$
LOFAR	36.8	1.25	202	4.88	51.8	26.2
AARTFAAC	6.48	1.37	35.6	5.74	9.15	32.1

Table 4.4: Noise Calculation results at $k = 0.2\,\mathrm{Mpc}^{-1}$,$\delta P_{21}^{SV}$ is the sample variance component of the noise and δP_{21}^{N} is the system noise component. Power spectrum values at k = 0.2 are $k^3 P_{21}/2\pi^2 = 4.84$, 29.8, $8.58\,\mathrm{mk}^2$, at respectively redshift z=8,10,12.

We can explain LOFAR's sensitivity on smaller scales (large k) with its larger baselines, LOFAR's antenna distribution stretches out until $r_{\max} = 1500$m. AARTFAAC's sensitivity on larger scales can be explained with the large FoV of each station. When we look at the numbers in table 4.4, we see:

- LOFAR has the lowest level of system noise, although it does not differ much from AARTFAAC.
- The AARTFAAC system has the lowest level of sample variance.

How strongly a noise component contributes is more or less determined by the effective area A_{eff} of each station within the array. Increasing the area lowers the system noise, see equation (2.2). However it also decreases the FoV, i.e. the size of the observed volume, therefore the sample variance increases because we are less able to measure all modes accurately. While it might seem tempting to maximize the effective area to decrease the system noise, this in turn increases the sample variance. This sample variance does not depend on integration time, so it is impossible to diminish this uncertainty by integrating longer. The choice of a certain station size, creates a lower limit for the measurable modes of k.

The results show that LOFAR and its extension AARTFAAC have promising sensitivity to measure the 21-cm power spectrum. The LOFAR-AARTFAAC system has increased sensitivity on scales below $k < 0.2\,\mathrm{Mpc}^{-1}$, using the AARTFAAC system with the "standard" LOFAR system would be an interesting application of the LOFAR-system in Holland.

4.2.2 Cosmic Dawn

In Visbal et al. [2012] a power spectrum at redshift $z = 20$ was presented which took into account the velocity difference between collapsing dark matter and baryons, this velocity difference would boost the power spectrum. This gives an interesting opportunity for the LOFAR-LBA system to potentially measure the power spectrum at much earlier times. The LOFAR-LBA system operates in the range of 10-70 MHz, the range required for measurements out to redshift $z = 20$.
To estimate which of the extensions would be more suitable for power spectrum measurements at $z = 20$ we compared LOFAR, AARTFAAC and the Superstation with an observation time of 1000 hours and a bandwidth of 10 MHz. The power spectrum at this redshift was created by taking a simple piecewise linear interpolation of the power spectrum displayed in Visbal et al. [2012]. The collecting area of an LBA antenna is given by $\lambda^2/3$. this is valid when the wavelength of observation is smaller than the distance separating each antenna. So $A_{\mathrm{p}} = 25/3$ around 60 MHz. The other parameters remain the same. The results of the comparison are displayed in figure 4.1 and tabulated in table 4.4.

The results in table 4.4 show that the AARTFAAC system is less limited by sample variance than the Superstation, which is due to the large FoV. This would enable LOFAR-AARTFAAC to measure the smallest k-modes, which is more interesting for Cosmology rather than for EoR-studies. The Superstation has more overall sensitivity, due to its large collecting area. This shows

that the *total collecting area* has more importance than the collecting area per station. Which means there is some optimal balance between station number and collecting area per station. The detection capabilities of the LOFAR-Superstation are very promising, since this exceed both LOFAR and the AARTFAAC system at these high redshifts. So given the power spectrum in Visbal et al. [2012] it might even be possible to detect the 21-cm power spectrum out to a redshift of $z = 20$ with the LOFAR-superstation in Nancy. This makes the Superstation the most promising instrument for Cosmic Dawn observations during the next decade, since this is the timescale over which the SKA should be build.

	$k = 0.03$		$k = 0.10$		$z = 0.20$	
Array	δP_{21}^{SV}	δP_{21}^{N}	δP_{21}^{SV}	δP_{21}^{N}	δP_{21}^{SV}	δP_{21}^{N}
LOFAR	2.29	167	1.17	$2.09 \cdot 10^2$	0.563	$12.4 \cdot 10^3$
AARTFAAC	0.404	102	0.212	$2.29 \cdot 10^2$	0.061	$4.16 \cdot 10^4$
Superstation	2.42	4.99	1.27	106	0.367	$1.08 \cdot 10^3$

Table 4.5: Noise calculation results for the LOFAR-AARTFAAC and -Superstation extensions. All results are in mk^2. The used 21-cm power spectrum values are $k^3\, P_{21}/2\pi^2 = 75,\ 284,\ 410\,\mathrm{mk}^2$, at respectively k = 0.03, 0.10, 0.20 Mpc^{-1}.

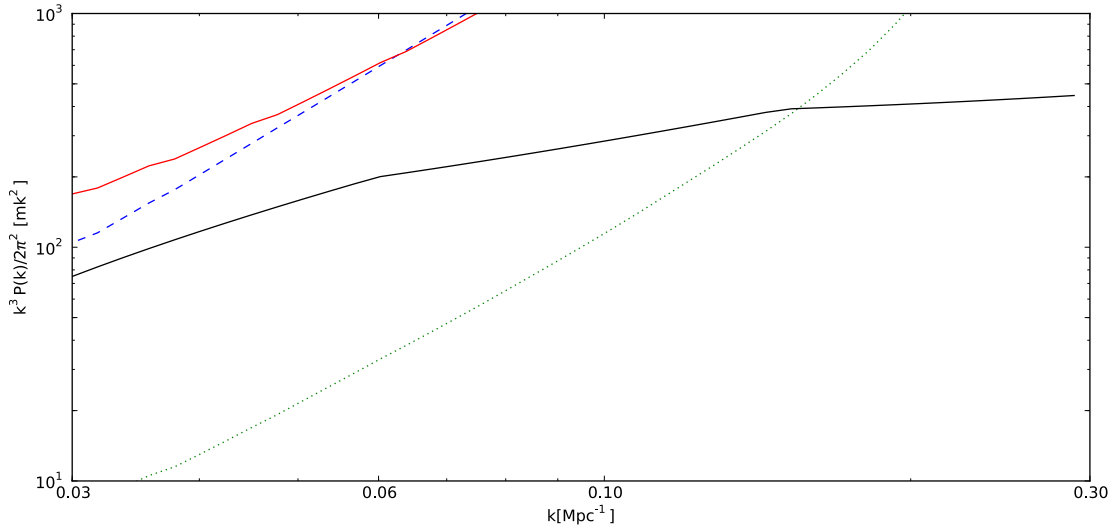

Figure 4.5: Comparison between LOFAR, LOFAR-AARTFAAC and the -Superstation extensions.The red solid line is the sensitivity of LOFAR's LBA system, the blue dashed line is the sensitivity of the superterp-LBA's using AARTFAAC, the green dotted line is the sensitivity of the superstation. The black solid line represents the crude reproduction of the predicted power spectrum

4.2.3 Scaling relation

To get a better understanding and quantification of the influence of design parameters on the sensitivity, we adopt a scaling relation[2] from Mellema et al. [2012], which is based on the derivation

[2]This equation is not valid for SKA, because SKA has high enough sensitvity that sample variance has to be included as well.

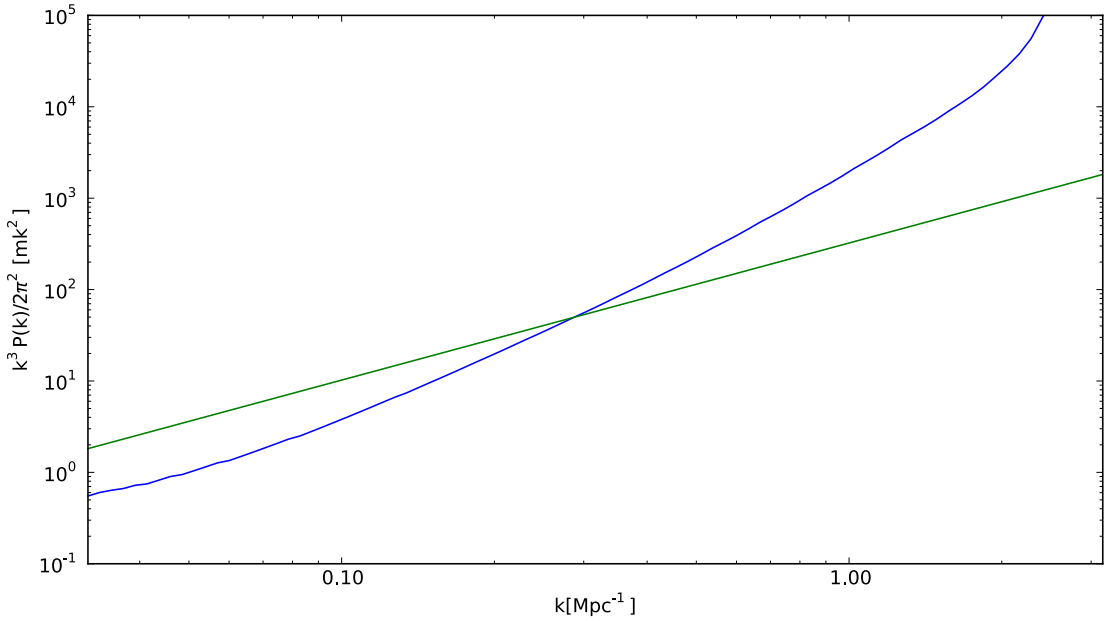

Figure 4.6: A comparison between the analytical scaling relation and a numerical calculation. The numerical results are plotted in blue, the analytical result is plotted in green. The LOFAR parameters as listed in table 4.1 were used in this calculation.

of McQuinn et al. [2006].

$$\delta P_{21}^N = \left(\frac{2}{\pi}\right) k^{\frac{3}{2}} [x^2 y \cdot \Omega_{\mathrm{FoV}}]^{\frac{1}{2}} \left(\frac{T_{\mathrm{sys}}}{\sqrt{B t_{\mathrm{int}}}}\right)^2 \left(\frac{A_{\mathrm{core}} A_{\mathrm{eff}}}{A_{\mathrm{coll}}^2}\right) \tag{4.1}$$

The scaling relation assumes the number density of baselines in the uv-plane is constant. This does not hold completely for real arrays, see figures 4.1 and 4.3. But the equation is valid in the central region of the uv-plane, since changes in the density of baselines are small here. A comparison between the scaling relation and a numerical calculation for LOFAR is depicted in figure 4.6. The numerical results diverge at larger k, since at larger k the density of baselines converges to zero, leading to an infinite noise. And at lower k the noise becomes dominated by sample variance, which is not taken into account for in the scaling relation. If we note that the FoV $\Omega_{\mathrm{FoV}} = \lambda^2 A_{\mathrm{eff}}$, we can rewrite this equation into one which only contains design parameters. For the details, read Mellema et al. [2012].

$$\delta P_{21}^N \propto \frac{A_{\mathrm{core}}}{N_{\mathrm{ant}}^2 A_{\mathrm{eff}}^{3/2}} \tag{4.2}$$

The new parameter A_{core} is the area of the core area of the array, i.e. $A_{\mathrm{core}} = \pi r_{\mathrm{c}}^2$. Equation 4.2 tells us that an increase of N_{ant} has the strongest effect on the sensitivity, after which the effective area per antenna A_{eff} has the strongest effect. Reduction of the core area, compactifying the array, also increases the sensitivity of the array. The results indicated that collecting area influences the sensitivity most dominantly, but this is merely due to the fact the effective area is more easily increased. The ratio between LOFAR and PAPER for the effective area was about 100, while the ratio for the number of antenna was about $\frac{2}{5}$. And for the AARTFAAC-Superstation comparison, the ratio between the collecting area per antenna is about 36, while AARTFAAC has only a factor of 3 more antennas.

So we can adopt equation 4.2, which tells us how the sensitivity of an array scales in the noise dominated region within an order of a magnitude accuracy. One form of this scaling relation shows that the sensitivity it not determined by only the effective antenna area, but by the ratio of the effective area and the core area. The number of antennas has the strongest influence on the the sensitivity, after which the effective area and the core area follow. However since it easier to increase the effective area of a station, than increasing the amount of stations, the sensitivity of an array can be increased dramatically by increasing the station size and compactifying the array.

4.3 SKA configurations

So far our analysis of current arrays. Our next goal is to find an optimal SKA lay-out, which fits the requirements for EoR science requirements. The already constructed arrays showed us that there is considerable amount of sensitivity to gain in the size of a station.We will now look at the influence of different design parameters on the sensitivity. Our next calculation consisted of varying the following parameters: the number of antennas N_{ant}, the total collecting area $N_{\mathrm{ant}} \cdot A_{\mathrm{p}}$, the slope of the antenna distribution and the outer radius of the array. This was done at a fixed redshift of z=10 or 130 MHz, an observation time t_0 of 1000 hours and core radius r_{c} of 1 km. The results are shown in figure 4.3.

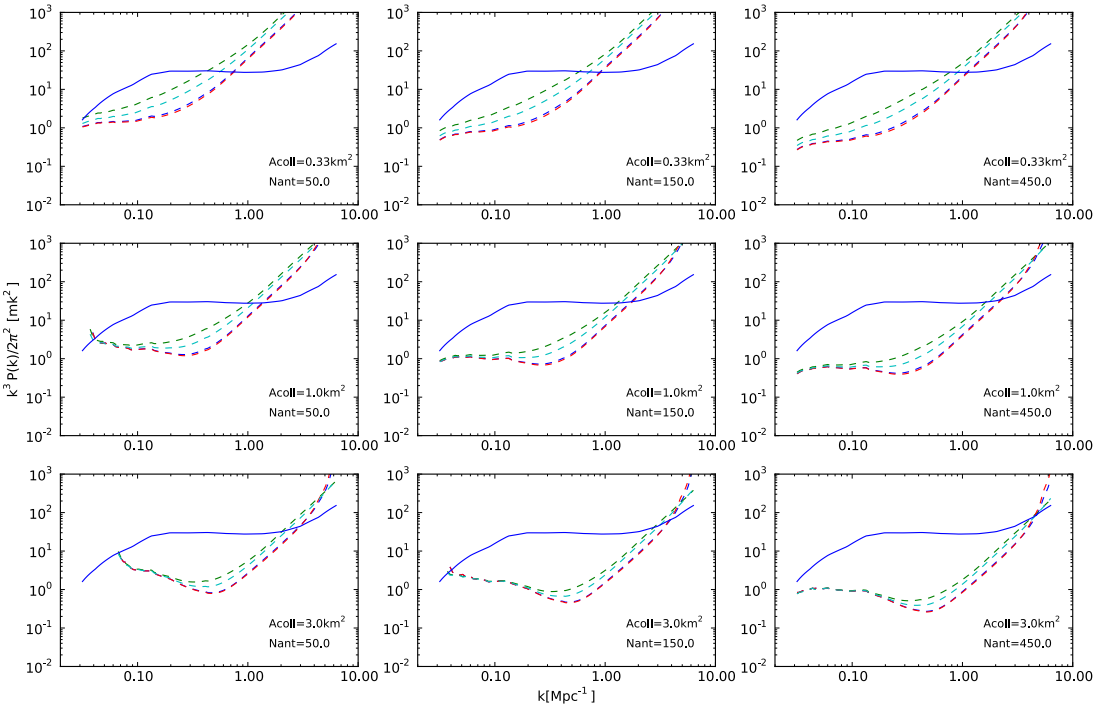

Figure 4.7: The results for different SKA lay-outs at redshift $z = 10$, note the red and blue dashed lines. The blue dashed line: $p = 1$, $r_{\mathrm{max}} = 2\,\mathrm{km}$, the green dashed line: $p = 1$, $r_{\mathrm{max}} = 5\,\mathrm{km}$, the red dashed line: $p = 2$, $r_{\mathrm{max}} = 2\,\mathrm{km}$ and the magenta dashed line: $p = 2$, $r_{\mathrm{max}} = 5\,\mathrm{km}$. The solid blue line represents the 21-cm power spectrum.

From figure 4.7 we can conclude the following:

- Compact arrays, i.e. in this case $r_{\mathrm{max}} = 2\,\mathrm{km}$, are the most sensitive configurations

Array	$A_{\rm p}$ (m^2)	$N_{\rm ant}$	p	r_c (m)	$r_{\rm max}$ (m)
SKA	$0.33, 1.0, 3.0 \cdot 10^6/N_{\rm ant}$	50, 150, 450	1, 2	1000	2000, 5000

Table 4.6: SKA parameters

- Arrays with a dense core, $p = 2$, are the more sensitive than configuration with $p = 1$.
- Increasing the total collecting area, at fixed antenna number, increases the sensitivity on a decreasing range of measurable k-scales.
- Sample variance limits all configurations, with different slopes p and outer radius r_{max} at fixed $A_{\rm eff}$ and $N_{\rm ant}$, to the same scale.
- Increasing the number of antennae increases the sensitivity on all scales.

So overall the most densest configurations lead to the highest sensitivity, as predicted by scaling relation (2.2). Decreasing $r_{\rm max}$ and increasing the slope p increase the antenna density in the core area. Increasing the slope, adds antennae to the core of the array while taking antennae from the outer regions, both are a way of making the array more compact. A more shallow slope ensures there is a certain number of long baselines present in the array, but the results indicate that the presence of long baselines do not affect the sensitivity on smaller scales (large k). Figure 4.7 shows that increasing the area of a station for a fixed collecting area increases the overall sensitivity, at the cost of making it impossible to measure large scale variations in the 21-cm signal. These large scale variations do not fit in the small FoV of large stations. So in order to observe the small k-scales we must limit the station size. Although increasing the antenna number $N_{\rm ant}$ increases the sensitivity on all scales, the main problem with increased antenna number is correlation cost. Increasing the antennas increases the data output and this requires more computing power and data storage, which is an issue of technological feasibility.

So we find an ideal SKA-lay out for power spectrum measurements would follow these guidelines. It should be compact; a dense core has more sensitivity than a diffuse core. This can be achieved by increasing the slope or a decrease in core radius. Constrain station size since this limits sample variance, since system noise can be compensated for with integration time, while sample variance does not depend on the observation itself. Most favorable would be the increase of the number of stations, but this will be limited by the available computational power.

Chapter 5

Conclusion

5.1 Summary

The main goals of this research were to properly compare current arrays EoR arrays and to find an optimal configuration of the Square Kilometer Array (SKA) for observations of the 21-cm power spectrum of Reionization. To find this optimal configuration we looked at the sensitivity of a few precursors and pathfinders in the HBA regime; LOFAR, LOFAR-AARTFAAC, the LOFAR-Superstation, MWA and PAPER. These sensitivities were calculated using a numerical adaptation of the equations derived in McQuinn et al. [2006] complemented by a 21-cm power spectrum generated with the reionization code 21cmFast. We also compared the sensitivity of the LOFAR-AARTFAAC and LOFAR-Superstation extensions at the LBA regime. We compared our numerical results with an analytical scaling relation to quantify the influence of design parameters on the sensitivity. Finally we calculated the sensitivities of different SKA lay-outs, which we used to discuss an optimal SKA lay-out The results of these calculations are summarized here. This chapter will also discuss some future work which can be performed to complement this research.

- LOFAR is currently the array which is most sensitive to the 21-cm power spectrum. PAPER and MWA have sensitivity on larger scales, which are less relevant for Reionization. However this sensitivity is relevant for cosmological studies.
- LOFAR and the LOFAR-AARTFAAC system have promising sensitivity to measure the 21-cm power spectrum out to redshift $z = 10$, with single beam observations.
- The LOFAR-Superstation is the most promising instrument we have for the next decade to measure the 21-cm power spectrum out to redshift $z = 20$.
- The sensitivity of an array can be described with a scaling relation: $\delta P_{21}^N \propto A_{\rm core}/(N_{\rm ant}^2 A_{\rm eff}^{3/2})$. However this equation is only valid within the central region of the uv-plane and when the S/N < 1. When S/N$\sim$1 the sample variance has to be included or a numerical calculation has to be made.
- An SKA lay-out would ideally follow the following guidelines. Compactify the array, adding more longer baselines does not enhance power spectrum sensitivity significantly. Constrain station size, since larger stations have small FoV's which increases the sample variance, which can not be diminished by increasing the observation time. A large number of antennae increases the sensitivity, as expected, however the number of antennas will be constrained by the available computing power.

The result of this research is a code which calculates baseline density of a radio array given several input parameters. This baseline density is used to calculate the sensitivity to the 21-cm power

spectrum. The code can handle various array configuration, only to be constrained by circular symmetry. The free parameters are: station area A_{p}, antenna number N_{ant}, slope of antenna distribution p, core radius r_{c}, outer radius r_{max} and an additional parameter inner radius r_{min}. Which allows for ring like arrays such as PAPER. However there was not enough time to exactly calculate which values these parameters should take for the SKA to fit the requirements of the EoR key science project.

5.2 Future Work

To fulfill the ultimate goal of this research an optimization routine, see appendix A, should be implemented in the code to actually calculate parameter values which fit the requirements of the EoR key science project of SKA. Not only should such a code optimize for a sensitivity such to measure the difference in the power spectrum at different redshifts, it should also take into account economical effects such as increasing the amount of antennas. In other words apart from science demands it should take into account technical feasibility.

The code itself could also become more user friendly, by building a general users interface around the existing functions. However creating user friendly program also limits the codes flexibility, since all possible demands of the user have be taken in account for. A user unfriendly code, the calculating part of the code itself, has more flexibility if the users knows how to use it. However for basic calculations given initial parameters, the user friendly version should suffice.

The code could also be extended to calculate the imaging capabilities of radio arrays. For this only the system noise should be considered.

Chapter 6

Nederlandse Samenvatting

Radio golven bieden niet alleen de mogelijkheid om naar het nabije waterstof gas te kijken in onze eigen melkweg, maar ook om naar het waterstofgas in het vroege heelal. Het waarnemen van het waterstofgas in het vroege heelal biedt namelijk de mogelijk om de vorming van de eerste structuren, sterren en sterrenstelsels, waar te nemen. Deze periode wordt "the Epoch of Reionization" (EoR) genoemd. En over het ontstaan van de structuren tijdens deze periode is nog veel onduidelijk. Observaties van het vroege heelal zouden echter duidelijkheid kunnen bieden aan wat er zich heeft afgespeeld. LOFAR en andere radiotelescopen zoals PAPER en MWA zijn in staat om deze radiogolven uit het vroege heelal waar te nemen om zo het waterstofgas waar te nemen. Het probleem op dit moment is echter dat deze radiotelescopen niet gevoelig genoeg zijn om gedetailleerde afbeeldingen van het vroege heelal te maken. De 21-cm straling die wordt uitgezonden door het waterstofgas zit namelijk verscholen in een overweldigende hoeveelheid radiostraling van de Melkweg zelf. Dus in plaats van het direct observeren van het waterstofgas is er een andere strategie ontworpen om toch uitspraken te kunnen doen over het vroege heelal. De huidige strategie is om te kijken naar verschillen in de hoeveelheid 21-cm emissie, op verschillende schaal groottes. Varieert de hoeveelheid neutraal waterstof op kleine afstanden of op grote afstanden? Maar ook met deze observatiestrategie zal het een uitdaging voor de huidige en toekomstige telescopen om het vroege heelal waar te nemen.
Een toekomstig telescoop project is de Square Kilometre Array (SKA) die onder andere ook op zoek zal gaan naar het waterstofgas in het vroege heelal. In dit onderzoek zal worden gezocht naar de richtlijnen waaraan het ontwerp van de SKA zal moeten voldoen om het waterstof te kunnen meten. De gevoeligheid van huidige telescopen zal ook worden vergeleken om uit te vinden welke al dusdanig gevoelig zijn om het waterstofgas te observeren.
Uit dit onderzoek blijkt dat de radiotelescopen MWA en PAPER niet gevoelig genoeg zijn om uitspraken te kunnen over de vorming van de eerste structuren tijdens de EoR. LOFAR heeft enige gevoeligheid tot een zeker moment terug (roodverschuiving $z = 10$) in het heelal. Ontwikkelingen aan LOFAR; het LOFAR-AARTFAAC systeem en het LOFAR-Superstation bieden interessante mogelijkheden om deze structuren nu al te meten. Het Nederlandse gedeelte van LOFAR zou in combinatie met het AARTFAAC-systeem verschillen in 21-cm emissie kunnen meten. Maar het LOFAR-superstation zal de komende tien jaar waarschijnlijk het meest gevoelige instrument zijn dat we hebben om het vroege heelal te kunnen observeren. Het bouwen van de SKA zal namelijk tenminste 10 jaar in beslag nemen.
Richtlijnen voor een SKA-ontwerp zijn als volgt: compact, een dusdanig groot aantal radioantennes, waarvan de oppervlakte niet te groot mag zijn. Een te groot aantal stations zal echter te veel computerkracht vergen om de data te kunnen verwerken. Er moet dus nog naar een optimale balans tussen het antenne oppervlak en het aantal stations gezocht worden.

Chapter 7

Acknowledgements

In this final chapter I would like to thank a few people who have helped and supported me through this project. First I would like thank my supervisor Leon Koopmans. When I came to Leon I had in mind to do research on galaxies and dark matter. What I didn't expect was the offer to do research on design parameters of the SKA. With no experience in the instrumental side of astronomy I just had to take this project and I was glad I could do this with him. Leon always had always plenty of ideas to keep me busy, but more importantly he took the time to explain many concepts behind radio interferometry which were quite new to me. His neverending enthusiasm when I came by with new results was always very motivating. I also enjoyed thinking about the use of the LOFAR-extensions which are relevant for present day observations. I would also like to thank Saleem Zaroubi, not only for the power spectrum he gave to me. But also for his capita course about the EoR and taking the time to explain what the 21-cm power spectrum actually is. His enthusiasm made writing the first part of my thesis much more easier. For the programming part I would like to thank Martin Vogelaar and Omar Choudhury, my conversations with them always helped me to improve my code. Furthermore I would like to thank Froukje Gjaltema for helping me with my struggle with Adobe Photoshop and her feedback on my report, her help made it much more easier creating some illustrative figures. And last but not least I want to thank my fellow astronomy students for keeping me company during my research. I had much fun sharing room 134 and many lunches with you.

Bibliography

Rennan Barkana and Abraham Loeb. In the beginning: The first sources of light and the reionization of the universe. *Phys.Rept.*, 349:125–238, 2001. doi: 10.1016/S0370-1573(01)00019-9.

A. P. Beardsley, B. J. Hazelton, M. F. Morales, W. Arcus, D. Barnes, G. Bernardi, J. D. Bowman, F. H. Briggs, J. D. Bunton, R. J. Cappallo, B. E. Corey, A. Deshpande, L. deSouza, D. Emrich, B. M. Gaensler, R. Goeke, L. J. Greenhill, D. Herne, J. N. Hewitt, M. Johnston-Hollitt, D. L. Kaplan, J. C. Kasper, B. B. Kincaid, R. Koenig, E. Kratzenberg, C. J. Lonsdale, M. J. Lynch, S. R. McWhirter, D. A. Mitchell, E. Morgan, D. Oberoi, S. M. Ord, J. Pathikulangara, T. Prabu, R. A. Remillard, A. E. E. Rogers, A. Roshi, J. E. Salah, R. J. Sault, N. Udaya Shankar, K. S. Srivani, J. Stevens, R. Subrahmanyan, S. J. Tingay, R. B. Wayth, M. Waterson, R. L. Webster, A. R. Whitney, A. Williams, C. L. Williams, and J. S. B. Wyithe. The eor sensitivity of the 128 antenna murchison widefield array. April 2012.

G. B. Field. Excitation of the hydrogen 21-cm line. *Proceedings of the IRE*, 46(1):240–250, 1958.

Paul M. Geil. *Studying the epoch of hydrogen reionisation in redshifted 21-cm radiation.* PhD thesis, University of Melbourne, 2011.

David W. Hogg. Distance measures in cosmology. 2007.

Daniel C. Jacobs, James E. Aguirre, Aaron R. Parsons, Jonathan C. Pober, Richard F. Bradley, Chris L. Carilli, Nicole E. Gugliucci, Jason R. Manley, Carel van der Merwe, David F. Moore, and Chaitali R. Parashare. New 145-mhz source measurements by paper in the southern sky. 2011,:ApJ,734,L34, May 2011.

Vibor Jelic, Saleem Zaroubi, Panagiotis Labropoulos, Rajat M. Thomas, Gianni Bernardi, et al. Foreground simulations for the lofar - epoch of reionization experiment. *Mon.Not.Roy.Astron.Soc.*, 389:1319–1335, 2008. doi: 10.1111/j.1365-2966.2008.13634.x.

Adam Lidz, Oliver Zahn, Matthew McQuinn, Matias Zaldarriaga, and Lars Hernquist. Detecting the rise and fall of 21 cm fluctuations with the murchison widefield array. *Astrophys. J.*, 680:962-974,:2008, November 2007.

Matthew McQuinn, Oliver Zahn, Matias Zaldarriaga, Lars Hernquist, and Steven R. Furlanetto. Cosmological parameter estimation using 21 cm radiation from the epoch of reionization. *Astrophys.J.*, 653:815–830, 2006. doi: 10.1086/505167.

Garrelt Mellema, Leon Koopmans, Filipe Abdalla, Gianni Bernardi, Benedetta Ciardi, Soobash Daiboo, Ger de Bruyn, Kanan K. Datta, Heino Falcke, Andrea Ferrara, Ilian T. Iliev, Fabio Iocco, Vibor Jelic, Hannes Jensen, Ronniy Joseph, Hans-Rainer Kloeckner, Panos Labroupoulos, Avery Meiksin, Andrei Mesinger, Andre Offringa, V. N. Pandey, Jonathan R. Pritchard, Mario G. Santos, Dominik J. Schwarz, Benoit Semelin, Harish Vedantham, Sarod

Yatawatta, and Saleem Zaroubi. Reionization and the cosmic dawn with the square kilometre array. October 2012.

Andrei Mesinger, Steven Furlanetto, and Renyue Cen. 21cmfast: A fast, semi-numerical simulation of the high-redshift 21-cm signal. March 2010.

Miguel F. Morales and J. Stuart B. Wyithe. Reionization and cosmology with 21 cm fluctuations. *Ann.Rev.Astron.Astrophys.*, 48:127–171, 2010. doi: 10.1146/annurev-astro-081309-130936.

Rick Perley. Fundamentals of radio interferometry, 2011. URL `http://www.aoc.nrao.edu/events/synthesis/2012/lectures/SISS-13-Intro.pdf`.

Peeyush Prasad and Stefan J. Wijnholds. Aartfaac: Towards a 24x7, all-sky monitor for lofar. May 2012.

Barbara Ryden. *Introduction to Cosmology*. Pearson Education, 2002.

Eli Visbal, Rennan Barkana, Anastasia Fialkov, Dmitriy Tseliakhovich, and Christopher M. Hirata. The signature of the first stars in atomic hydrogen at redshift 20. *Nature*, advance online publication:–, 06 2012.

Saleem Zaroubi. Eor redshifted 21-cm experiments, 2012a. URL `http://www.astro.rug.nl/~saleem/EoRCourse/EoR%20Course%203.pdf`.

Saleem Zaroubi. The epoch of reionization. June 2012b.

Appendix A

Optimization Scheme

This appendix will describe an attempt to calculate detailed design parameters for an optimal SKA lay-out which fits the EoR science requirements. However, there was not enough time to successfully implement the optimization.
The optimization routine was chosen from the available functions in the SciPy libraries. The fmin_l_bfgs_b method was chosen for its speed, since the calculation of the errors on the power spectrum are already quite time consuming. The routine makes use of the limited-memory Broyden-Fletcher-Goldfarb-Shanno (BFGS) method,which tries to calculate the hessian of the function which is to be optimized. This gives the algorithm information about the shape of the function, i.e. the extrema which we are after. The fmin_l_bfgs_b also allows for the input of bounds, which reduces the chance on runaway maximization of parameters such as the amount of antennas or their physical area.
For the optimization a target function is required, which contains the demands for the EoR project. The target function of choice is a 3-sigma certainty between the power spectrum at redshift z=8 and z=12. The difference between those power spectra must be three times as large as the error on the power spectrum. Possible issues could arise because the 3-sigma certainty is demanded on all scales, within the range of the given power spectrum. Which could to a decrease in S/N at certain scales to gain minimal sensitivity at other scales. Overall the configuration will resemble the target function the most, but in practice this configuration is less useful than a configuration which does have a peak sensitivity at a certain k. Or on the other hand the function will maximize sensitivity on smaller k which are easier measurable because most baselines are present at these scales.

Appendix B

Parameter File

```
#!/usr/bin/env python
"""
******************User Information*********************
This file contains the controlling parameter for pyRadio.py,
The parameters control which mode has been chosen:

Calculation of Errors on the  21 cm Power spectrum
        - Calculate Total noise = Sample Variance + System Noise
        - Calculate only system noise
        - Calculate only sample variance

Calculation of S/N using total noise

Comparison between different array configurations by calculating several arrays in one run

Optimization of a single array configuration.

******************Calculation Mode*********************
Set optimize_key to "no".
Then choose whether you want to calculate the errors and signal-to-noise.
If only signal-to-noise:        set snr_key to "yes" and errornoise_key to "no".
If BOTH S/N and noise:          set snr_key to "yes" and savenoise_key  to "yes"
If only errors on the power spectrum, set snr_key to "no":
        - for total noise: set tnoise_key to yes and set savenoise_key to
          yes
        - if also separate components necessary set savesys_key and/or
          savesv_key to "yes"
        - for system noise ONLY         set everything to "no" except
          snoise_key and savesys_key
        - for sample variance ONLY      set everything to "no" except
          vnoise_key and savesv_key
"""


file_names                              = ["SKA.param"]
#enter the filenames between "" and separated by commas.
#Everything before .param will be used to name output files.
#By leaving [""], the code will look for all .param files in parent directory

#Calculation type
optimize_key                    = "yes"
snr_key                                 = "no"
tnoise_key                              = "no"
snoise_key                              = "no"
vnoise_key                              = "no"

#optimize_key:  optimization "yes" or calculation "no"
```

```
#snr_key:                calculate signal to noise profile and errors on P_21
#tnoise_key:     calculate the total noise on power
#                                 spectrum (Limited between power spectrum range)
#Note: Can save components and total noise separately
#snoise_key:     ONLY calculate system noise, can
#                                 take any range in k and frequency

#vnoise_key:     ONLY calculate sample variance,
#                                 limited by power spectrum range

#Output and saving for further use
savesnr_key                                = "no"
savenoise_key                      = "no"
savesys_key                                = "no"
savesv_key                                 = "no"

#savesnr_key:    save signal to noise profile to ..._snr.txt
#savenoise_key: save total noise to ..._tnoise.txt
#savesys_key:    save system noise component separately to ..._system.txt
#savesv_key:     save sample variance separately          to ..._svar.txt

#Binning size for spherical averaging
deltak                                     = 0.5

#Calculation ranges
#Upper and lower limit for frequency, calculation will be done for intervals separated by
#the bandwidth. Unit: MHz
#Power spectrum in PowerSpecData2.txt limits are 120 < nu <150 MHz
lowernu = 120.
uppernu = 150.

#Upper and lower limit for k, in log10[h Mpc^-1]
#Power spectrum in PowerSpecData2.txt limits are -1.5 < log[k] < 0.8
lowerk  = -1.5
upperk  = 0.8



#Choose which parameters should be optimized
#flags free = 1 fixed =0
Nantflag  = 0
r_minflag = 0.
r_cflag   = 0
r_maxflag = 1
powerflag = 1
Areaflag  = 1
t0flag    = 0

#bounds for every parameter, if no upper- lowerbound enter None
#lowerbound
Nantmin  = 20.
r_minmin = 0.
r_cmin   = 5.
r_maxmin = 1500
powermin = 0
Areamin  = 20
t0min    = 700.

#upperbound
Nantmax  = 120.
r_minmax = 0.
r_cmax   = 1000.
r_maxmax = 10000.
powermax = 5
Areamax  = 10000
t0max    = 1200.
```

```
#Cosmological parameters
#Hubble parameters
h         = 0.7
#mass density parameter
omega_m = 0.3
#vacuum density parameter
omega_l = 0.7
#curvature density parameter
omega_k = 1.-omega_m - omega_l


#Constants of nature
#Speed of light vacuum (m/s)
c         = 2.9979*10.**8.
#Hubble constant (km/s)/Mpc
H0        = 100.*h
#frequency 21 cm at z=0   (Mhz)
nu21      = c/0.21*10.**-6.
#Megaparsec (m)
Mpc       = 3.26*365.*24.*3600.*c
#Hubble Distance (Mpc)
Dh        = c/H0*10.**-3.



#The amount of points used for calculation
dbase                   = 100
drad                    = 100
dphi                    = 100
dk                      = 100
```

Appendix C

pyRadio

```
#!/usr/bin/env python
import numpy as n
from scipy.interpolate import interp1d
from scipy.integrate import quad
from scipy.optimize import fmin_l_bfgs_b as fminb
from matplotlib import pyplot as plt
from matplotlib import rcParams
import os
import time
from parameters import *

#Start Clock for time keeping
start = time.clock()
#setting Latex Math-text for lay-out
rcParams['mathtext.default']='regular'

#ignore zero division warnings
n.seterr(divide='ignore')

#Create an array for log file
logfile = []

#systeem temperatuur definieren
def Tsys(nu):
  Tsys = 400.*(nu/(150.))**-2.55
  return Tsys


#Effective telescope area
def Aeff(nu,Area1):
  Aeff1 = Area1*(nu/(120.))**-2.
  return Aeff1


#Redshift
def redshift(nu):
#Redshift formula
  z = nu21/nu -1.
  return z


#inverse redshift
def frequency(z):
  f = nu21/(1.+z)
  return f
```

```
#Comoving Distances
def D(z_min,z_max):
#function for integration
  E = lambda z: (omega_m*(1.+z)**3. + omega_k*(1.+z)**2. + omega_l)**-0.5
#integrating the function
  A = quad(E,z_min, z_max)
#Selecting integrand from output and multiply with hubble distance
  D = Dh*A[0]
  return D


#Normalisation for Antenna distribution
def Normal(Nant0,D_a0,r_c0,r_max0,power0):
#select the right normalisation
#when the slope has power=2
  if power0 == 2.:
    intb = -D_a0**2.*n.pi + r_c0**2.*n.pi + 2.*r_c0**2.*n.pi*n.log(r_max0/r_c0)
    N = Nant0/intb
#check whether this configuration is physically possible: antenna filling
    if N*D_a0**2.*n.pi*(r_c0**2.-D_a0**2.)>1.1*n.pi*r_c0**2.:
      return "impossible"
    else:
      return N
  else:
#for any other power
#Integral over the antenna power law
    intb = -((r_max0**-power0*(2.*r_c0**power0*r_max0**2. - 2.*D_a0**2.*r_max0**power0 \
     + D_a0**2.*r_max0**power0*power0 - r_c0**2.*r_max0**power0*power0)*n.pi) \
     /(-2.+ power0))
#corresponding normalisation factor
    N = Nant0/intb
#check whether this configuration is physically possible: antenna filling
    if N*D_a0**2.*n.pi*(r_c0**2.-D_a0**2.)>1.1*n.pi*r_c0**2.:
      return "impossible"
    else:
      return N


#Antenna Distribution
def N_a(r1,Nant1,D_a1,r_c1,r_max1,power1):
#specify a range for integration over r
#creating n.pian empty array for number densities
  antenna1 = n.zeros(len(r1))
#selecting correct normalisation
  Norm1 = Normal(Nant1,D_a1,r_c1,r_max1,power1)
#loon.ping over the element of input
  for i1 in range(len(r1)):
    if D_a1<= r1[i1] <=r_c1:
      antenna1[i1] = Norm1
    elif r_c1< r1[i1] <r_max1:
      antenna1[i1] = Norm1*(r_c1/r1[i1])**power1
    else:
      antenna1[i1] = 0.
  return antenna1



#Function which tabulates convolution results
def Table_b(Nant2,r_min2,r_c2,r_max2,power2,Area2,nu2):
#specify a range for integration over r
  r2 = n.linspace(r_min2,r_max2,dbase)
  #r2   = n.logspace(-2.*abs(n.log10(r_min2)),2*n.log10(r_max2),dbase)
  #specify a range for baseline U
  bas2 = n.logspace(-2.*abs(n.log10(n.sqrt(Area2))),2*n.log10(r_max2),dbase)
#specify a range for antenna convolution over phi
  phi2   = n.linspace(0.,2.*n.pi,dphi)
#Create empty array for unnormalised results
```

```
  baseline2 = n.zeros((len(nu2),dbase))
#Create empty array for normalisation values
  normalis2 = n.zeros(len(nu2))
#check wheter configuration is physically possible
  check  = Normal(Nant2,r_min2,r_c2,r_max2,power2)
  if check == "impossible":
    return "impossible"
  else:
    for i2 in range(len(nu2)):
  #calculate wavelength corresponding to nu
      l2  = c/nu2[i2]*10.**-6.
  #calculate maximum number of baselines
      Nbase2  = 0.5*(Nant2**2. - Nant2)
  #create array for unnormalised baseline density
      unnorm2 = n.zeros(dbase)
      for j2 in range(dbase):
  #create array for integration over phi
        n_ur2 = n.zeros(len(r2))
        for o2 in range(len(r2)):
  #calculate |r-lambda*U| for a certain r and U for 0< phi<2Pi
          a2      = n.sqrt(r2[o2]**2.-2.*l2*bas2[j2]*r2[o2]*n.cos(phi2) \
                +l2**2.*bas2[j2]**2.)
  #calculate corresponding number of antennae
          b2      = N_a(a2,Nant2,r_min2,r_c2,r_max2,power2)
  #integrate over phi
          n_ur2[o2] = n.trapz(b2,phi2)
  #Calculating integrand for 0 < r < r_max
        c2  = 2.*n.pi*r2*N_a(r2,Nant2,r_min2,r_c2,r_max2,power2)*n_ur2
  #integrating from 0 to r_max
  #save Number of baseline for certain U in Table
  #save number of baselines for proper normalisation
        baseline2[i2,j2] = n.trapz(c2,r2)
  #Removing all values below the impossible baselinedistribution
  #define the minimum baseline
    minbas2 = n.sqrt(Area2)/l2
  #select all indices of baselines below this value
    indmin2 = n.where(bas2 < minbas2)
  #set density values to zero
    baseline2[i2,indmin2[0]] = 0.
  #calculate integrand for integration over u
    e2      = 2.*n.pi*bas2*baseline2[i2,:]
  #integrating over u
    f2      = n.trapz(e2,bas2)
  #calculate frequency dependent normalisation
    normalis2 = Nbase2/f2
    baseline2[i2,:]=normalis2*baseline2[i2,:]
  return baseline2

#Creating callable functions which interpolates row of Nbase for correct frequency
def N_b(nbase3,nu3,nuval3,u3,r_max3,Area3):
#Specify same range as in calculation the baselinedistribution
  bas3 = n.logspace(-2.*abs(n.log10(n.sqrt(Area3))),2*n.log10(r_max3),dbase)
#calculate difference of input freq. with frequency range
  diff3 = abs(nu3-nuval3)
#finding index at which difference is minimum
  ind3  = n.where(diff3 == min(diff3))
#calculate wavelength corresponding to frequency.
  l3  = c*10.**-6./nuval3
#using index to interpolate the desired row of nbase
  base3 = interp1d(bas3,nbase3[ind3[0][0],:])
#creating an empty array for output
  line3 = n.zeros(len(u3))
#Interpolation can only handle u-values in range, so need to check whether
#the u-values match range
  for i3 in range(len(u3)):
#checking whether required u1 is within interpolation range
    if -2.*abs(n.log10(n.sqrt(Area3)))< n.log10(u3[i3]) \
```

```
    and n.log10(u3[i3]) < 2.*n.log10(r_max3):
#saving result for output
      line3[i3] = base3(u3[i3])
#if baseline isn't within range
    else:
      line3[i3] = 0.
#return result for further calculation
  return line3


#Powerspectrum of 21-cm signal as a function of freq. and k
def P_21(k,nu):
#import the power spectrum
  data  = n.loadtxt("PowerSpecData2.txt")
#Extract redshift values from first row
  zrange  = data[0,1:data.shape[1]]
  krange  = data[1:data.shape[0],0]
#calculate redshift value corresponding to input freq.
  znu     = redshift(nu)
  intfreq  = n.zeros(len(krange))
  if zrange[0] > znu or znu > zrange[len(zrange)-1]:
    znu = zrange[n.where(zrange-znu==min(abs(zrange-znu)))[0][0]]
    print "Warning frequency does not fit power spectrum range"
    print "Redshift z="+str(znu)+" will be used."
#Create an interpolation of the power spectrum for a certain new frequency
  for i21 in range(len(krange)):
    intfreq[i21] = interp1d(zrange,data[i21+1,1:data.shape[1]])(znu)
#Interpolate this power spectrum over k-range, for desired values
  p_21  = interp1d(krange,intfreq)
  return p_21(k)


#Detector + Sample noise on the angular averaged power spectrum McQuinn [2006]
def Error(Nant5,r_min5,r_c5,r_max5,power5,Area5,B5,t05,k5,nu5):
#Calculate the spherically averaged error at a given frequency and mode k
#Create empty mathematica array for Noise values
  global savenoise_key
  global savesys_key
  global savesv_key
  tNoise5  = n.zeros((len(k5)*len(nu5),3))
  sNoise5  = n.zeros((len(k5)*len(nu5),3))
  vNoise5  = n.zeros((len(k5)*len(nu5),3))
#create empty python array for noise values
  PNoise5 = n.zeros((len(nu5),len(k5)))
#arbitrary counter for saving values
  count5  = 0
#use baseline table function to calculuate baselines
  nbase5 = Table_b(Nant5,r_min5,r_c5,r_max5,power5,Area5,nu5)
  if nbase5 == "impossible":
    print "This array configuration is physiccaly not possible: overfill in antenna core"
    logfile.append("Overfill for Nant="+str(Nant5)+" ,r_min="+str(r_min5)+ \
        " ,r_c="+str(r_c5)+" ,r_max="+str(r_max5)+" ,slope="+str(power5))
    logger = open("logfile.txt","w")
    for failure in logfile:
      logger.write("%s\n" % failure)
    logger.close()
    return "impossible"
  else:
#loon.ping over frequency and k range
    for i5 in range(len(nu5)):
  #minimum observing frequency (Mhz)
      nu_min5 = nu5[i5] - 0.5*B5
  #maximum observing frequency (Mhz)
      nu_max5 = nu5[i5] + 0.5*B5
  #minimum redshift
      z_min5  = redshift(nu_max5)
  #maximum redshift
```

```
    z_max5  = redshift(nu_min5)
#Real space depth (Mpc)
    y5      = D(z_min5,z_max5)
#calculate redshift
    z5      = redshift(nu5[i5])
#calculate distance to source (Mpc)
    x5      = D(0.,z5)
#calculate corresponding wavelength (m)
    l5      = c*10.**-6./nu5[i5]
#Real space volume
    V5      = x5**2.*y5*l5**2./Aeff(nu5[i5],Area5)
    for j5 in range(len(k5)):
#calculate modes of u
      u5  = x5*k5[j5]*n.sin(theta)/(2.*n.pi)
#calculate number densities baselines
      n5  = N_b(nbase5,nu5,nu5[i5],u5,r_max5,Area5)
#calculate observing times for mode k
      t_k5  = tk(nu5[i5],l5,n5,Area5,t05)
      if tnoise_key == "yes" or optimize_key== "yes":
  #calculate Noise covariances
      C_n5  = (l5**2.*B5*Tsys(nu5[i5])/Aeff(nu5[i5],Area5))**2.*(10.**6./(B5*t_k5))
  #calculate Sample Variance
      C_sv5      = P_21(k5[j5],nu5[i5])*l5**2.*B5**2./(Aeff(nu5[i5],Area5) \
            *x5**2.*y5)*(2.*n.pi**2./k5[j5]**3.)*10.**6.
  #Number of independent cells
      N_c5  = Ncell(k5[j5],theta,y5,V5)
  #Error in signal due to Noise
      ErrP5  = (Aeff(nu5[i5],Area5)*x5**2.*y5/(l5**2.*(B5*10.**6.)**2.)) \
            *((C_sv5+C_n5)/n.sqrt(N_c5))
      sErrP5  = (Aeff(nu5[i5],Area5)*x5**2.*y5/(l5**2.*(B5*10.**6.)**2.)) \
            *((C_n5)/n.sqrt(N_c5))
      vErrP5  = (Aeff(nu5[i5],Area5)*x5**2.*y5/(l5**2.*(B5*10.**6.)**2.)) \
            *((C_sv5)/n.sqrt(N_c5))
  #Averaging total noise over theta using composite trapezium for numerical integral
      a5  = ErrP5**-2.
      b5  = n.trapz(a5,theta)
    #averaging system noise only
      sa5  = sErrP5**-2.
      sb5  = n.trapz(sa5,theta)
    #averaging sample variance only
      va5  = vErrP5**-2.
      vb5  = n.trapz(va5,theta)
  #Spherically averaged total error
      sigma5     = 1./n.sqrt(b5)*10.**6.
    #spherically averaged system noise
      ssigma5     = 1./n.sqrt(sb5)*10.**6.
    #spherically averaged sample variance
      vsigma5     = 1./n.sqrt(vb5)*10.**6.
  #correcting to get mk^2 dimension
    #total noise
      plotvalue5 = sigma5*(k5[j5]**3./(2.*n.pi**2.))
    #system noise
      splotvalue5 = ssigma5*(k5[j5]**3./(2.*n.pi**2.))
    #sample variance
      vplotvalue5 = vsigma5*(k5[j5]**3./(2.*n.pi**2.))
  #Creating 2 dimensional mathematica array
  #k-entry for mathematica array
      tNoise5[count5,0] = k5[j5]
  #frequency entry for mathematica array
      tNoise5[count5,1] = nu5[i5]
  #result entry for mathematica array
      tNoise5[count5,2] = plotvalue5
    #system noise tabulation
      sNoise5[count5,0] = k5[j5]
      sNoise5[count5,1] = nu5[i5]
      sNoise5[count5,2] = splotvalue5
    #sample variance tabulation
```

```
      vNoise5[count5,0] = k5[j5]
      vNoise5[count5,1] = nu5[i5]
      vNoise5[count5,2] = vplotvalue5
      count5 +=1
#creating a 2 dimensional python array
      PNoise5[i5,j5]    = plotvalue5
    if snoise_key == "yes":
      savenoise_key  = "no"
      savesys_key    = "yes"
      savesv_key     = "no"
#calculate Noise covariances
      C_n5  = (l5**2.*B5*Tsys(nu5[i5])/Aeff(nu5[i5],Area5))**2.*(10.**6./(B5*t_k5))
#Number of independent cells
      N_c5  = Ncell(k5[j5],theta,y5,V5)
#Error in signal due to Noise
      sErrP5  = (Aeff(nu5[i5],Area5)*x5**2.*y5/(l5**2.*(B5*10.**6.)**2.)) \
            *((C_n5)/n.sqrt(N_c5))
  #averaging system noise only
      sa5  = sErrP5**-2.
      sb5  = n.trapz(sa5,theta)
  #spherically averaged system noise
      ssigma5     = 1./n.sqrt(sb5)*10.**6.
#correcting to get mk^2 dimension
      splotvalue5 = ssigma5*(k5[j5]**3./(2.*n.pi**2.))
#Creating 2 dimensional mathematica array
  #system noise tabulation
      sNoise5[count5,0] = k5[j5]
      sNoise5[count5,1] = nu5[i5]
      sNoise5[count5,2] = splotvalue5
      count5 +=1
    if vnoise_key == "yes":
      savenoise_key  = "no"
      savesys_key    = "no"
      savesv_key     = "yes"
#calculate Sample Variance
      C_sv5       = P_21(k5[j5],nu5[i5])*l5**2.*B5**2./(Aeff(nu5[i5],Area5) \
            *x5**2.*y5)*(2.*n.pi**2./k5[j5]**3.)*10.**6.
#Number of independent cells
      N_c5  = Ncell(k5[j5],theta,y5,V5)
      vErrP5  = (Aeff(nu5[i5],Area5)*x5**2.*y5/(l5**2.*(B5*10.**6.)**2.)) \
            *((C_sv5)/n.sqrt(N_c5))
  #averaging sample variance only
      va5  = vErrP5**-2.
      vb5  = n.trapz(va5,theta)
  #spherically averaged sample variance
      vsigma5     = 1./n.sqrt(vb5)*10.**6.
#correcting to get mk^2 dimension
  #sample variance
      vplotvalue5 = vsigma5*(k5[j5]**3./(2.*n.pi**2.))
#Creating 2 dimensional mathematica array
  #sample variance tabulation
      vNoise5[count5,0] = k5[j5]
      vNoise5[count5,1] = nu5[i5]
      vNoise5[count5,2] = vplotvalue5
      count5 +=1
#Create header containing used parameters
hn  = "Nant="+str(Nant5)
hm  = "\nr_min="+str(r_min5)
hrc  = "\nr_c="+str(r_c5)
hrx  = "\nr_max="+str(r_max5)
hp  = "\nslope="+str(power5)
ha  = "\nA_p="+str(Area5)
hb  = "\nB="+str(B5)
ht  = "\nt0="+str(t05)
hT  = ""
for ih in range(len(nu)):
  hT += "\nTsys("+str(nu[ih])+")="+str(Tsys(nu[ih]))
```

```
      header5 = hn+hm+hrc+hrx+hp+ha+hb+ht+hT
      if savenoise_key == "yes":
        save1t  = '_tnoise.txt'
        save2t  = array_name+save1t
  #save noise results for plotting in mathematica
        n.savetxt(save2t, tNoise5)
        fid = open( array_name+"_param.txt" , 'w' )
        fid.write( header5+'\n' )
        fid.close()
    if savesys_key == "yes":
      save1s  = '_system.txt'
      save2s  = array_name+save1s
      n.savetxt(save2s, sNoise5)
      fid = open( array_name+"_param.txt" , 'w' )
      fid.write( header5+'\n' )
      fid.close()
    if savesv_key == "yes":
      save1v  = '_svar.txt'
      save2v  = array_name+save1v
      n.savetxt(save2v, vNoise5)
      fid = open( array_name+"_param.txt" , 'w' )
      fid.write( header5+'\n' )
      fid.close()
  #return Python array containing total noise
    return PNoise5



#Average observing time mode k
def tk(nu2,l2,n2,Area2,t02):
  tk2 = t02*3600.*Aeff(nu2,Area2)/l2**2.*n2
  return tk2



def Ncell(k4,theta4,y4,V4):
  a4 = 2.*n.pi/(k4*n.cos(theta4))
#identifying whether input is array-like or a float
  if type(theta4) != type(n.zeros(1)):
#when mode fits in survey volume
    if a4 <= y4:
      Ncell4 = 2.*n.pi*k4**2.*n.sin(theta4)*deltak*k4*V4/(2.*n.pi)**3.
#when mode does not fit in survey volume
    else:
      Ncell4 = 0.
  else:
#for array entry
    Ncell4 = n.zeros(len(theta4))
    for i4 in range(len(theta4)):
#when mode fits in survey volume
      if a4[i4] <= y4:
        Ncell4[i4] = 2.*n.pi*k4**2.*n.sin(theta4[i4])*deltak*k4*V4/(2.*n.pi)**3.
#when mode does not fit in survey volume
      else:
        Ncell4[i4] = 0.
    return Ncell4



def SNR(Nant6,r_min6,r_c6,r_max6,power6,Area6,B6,t06,k6,nu6):
#create empty mathematica array for S/N results
  SNR6  = n.zeros((len(nu6)*len(k6),3))
#create empty python array for S/N results
  ratio6  = n.zeros((len(nu6),len(k6)))
#calculate the Noise + Sample Noise
  noise6  = Error(Nant6,r_min6,r_c6,r_max6,power6,Area6,B6,t06,k6,nu6)
#arbitrary counter for creating mathematica array
```

```
  count6  = 0
#calculating signal/noise for freq. and k
  for i6 in range(len(nu6)):
    for j6 in range(len(k6)):
#k-entry for mathematica array
      SNR6[count6,0]  = n.log10(k6[j6])
#frequency entry for mathematica array
      SNR6[count6,1]  = nu6[i6]
#calculate signal/noise
      snr6       = P_21(k6[j6],nu6[i6])/noise6[i6,j6]
#log10(signal/noise) entry for mathematica array for plotting
      SNR6[count6,2]  = n.log10(snr6)
#signal/noise entry for python array for further calculution
      ratio6[i6,j6]  = snr6
      count6 += 1
#Create header containing used parameters
  hn  = "Nant="+str(Nant6)
  hm  = "\nr_min="+str(r_min6)
  hrc  = "\nr_c="+str(r_c6)
  hrx  = "\nr_max="+str(r_max6)
  hp  = "\nslope="+str(power6)
  ha  = "\nA_p="+str(Area6)
  hb  = "\nB="+str(B6)
  ht  = "\nt0="+str(t06)
  hT  = ""
  for ih in range(len(nu)):
    hT += "\nTsys("+str(nu[ih])+")="+str(Tsys(nu[ih]))
  header6 = hn+hm+hrc+hrx+hp+ha+hb+ht+hT
  if savesnr_key == "yes":
    save1 = '_snr.txt'
    save2 = array_name+save1
#export mathematica array
    n.savetxt(save2,ratio6)
    fid = open( array_name+"_param.txt" , 'w' )
    fid.write( header6+'\n' )
    fid.close()
#return python array
  return ratio6



def TSNR(nu7,k7):
#create empty array for every k and frequency
  target = n.zeros((len(nu7),len(k7)))
#define target S/N for every k and freq. as 2
  target += 3.
  return target

#Function to be minimalized: the difference between SNR and TSNR
def Chi(Nant8,r_min8,r_c8,r_max8,power8,Area8,B8,t08,k8,nu8):
#calculate S/N according to free parameters per freq. and k
  #snr8  = SNR(Nant8,r_min8,r_c8,r_max8,power8,Area8,B8,t08,k8,nu8)
  p8_21 = P_21(k8,lowernu)
  p12_21= P_21(k8,uppernu)
  dp    = abs(p8_21-p12_21)
  error8= Error(Nant8,r_min8,r_c8,r_max8,power8,Area8,B8,t08,k8,nu8)
  snr8 = n.zeros((len(nu8),len(k8)))
  for i8 in range(len(nu)):
    snr8[i8,:] = dp/error8[i8,:]
#create target values
    tsnr8 = TSNR(nu8,k8)
#calculate differences squared
    a8   = (snr8-tsnr8)**2.
#create an empty array for chi per frequency
  #b5   = n.zeros(len(nu))
  #b5[i7] = n.trapz(a5[i7,:],k)
#if not interested in a particular frequency or k-range
```

```
  chi8 = n.sum(a8)
  return chi8


#communication function between optimalisation routine and Chi
def Communicate(x):
#create an empty array with length of number of parameters
  p0 = n.zeros(len(flag))
#arbitrary counter
  count9 = 0
  shoutput = []
#distributing variables according to input of Chi
  for i9 in range(len(flag)):
#if parameter is to be optimized
    if flag[i9] == 1:
#select parameter from optimalisation input
      p0[i9] = x[count9]
      shoutput.append(names[i9]+": "+str(x[count9]))
      count9 += 1
#if parameter is fixed select initial input
    else:
      p0[i9] = param[i9]
  #print shoutput
  print ""
  print "Nant  ="+str(p0[0])
  print "rmin  ="+str(p0[1])
  print "rc  ="+str(p0[2])
  print "rmax  ="+str(p0[3])
  print "slope  ="+str(p0[4])
  print "Area  ="+str(p0[5])
  print "B  ="+str(p0[6])
  print "t  ="+str(p0[7])
  result9 = Chi(p0[0],p0[1],p0[2],p0[3],p0[4],p0[5],p0[6],p0[7],k,nu)
  now9 = time.clock()
  print "Time elapsed since execution "+str(now9-start)+" seconds"
  return result9



#Optimalisation routine
def Optimal(opti,bound):
  optimum = fminb(Communicate,opti,approx_grad=1,bounds=(bound))
  count10  = 0
  output10 = []
  for i10 in range(len(flag)):
    if flag[i10] == 1:
      keys10 = str(names[i10])+": "+str(optimum[0][count10])
      output10.append(keys10)
      print keys10
      count10 += 1
    print optimum[2]
    results10 = open(array_name+"_optfile.txt","w")
    for key10 in output :
      results10.write("%s\n" % key10)
  results10.write("%s\n" % "value at minimum: "+str(optimum[2]))
  results10.close()


#Calculation range
#specifying a range for k
k    = n.logspace(lowerk,upperk,dk)
#define a range for theta for averaging over cylindrical averaging
theta    = n.linspace(0.,n.pi,20)


#Steps for communicating with optimalisation routine
#list containing names for a nice output
```

```
names    = ["Number of antennae","Minimum Radius","Core radius","Outer radius",
                  "Slope of antenna distribution", "Effective antenna area",
                  "Bandwidth","Integration time"]

#Combining all values for flags, bounds in a vector
flag      = n.array([Nantflag,r_minflag,r_cflag,r_maxflag,powerflag,Areaflag,0,t0flag])
Minimum   = n.array([Nantmin,r_minmin,r_cmin,r_maxmin,powermin,Areamin,0,t0min])
Maximum   = n.array([Nantmax,r_minmax,r_cmax,r_maxmax,powermax,Areamax,0,t0max])



#################################USER-interaction part#################################
#Uses information in parameters.py and param files to determine course of action#
######################################################################################



if file_names == [""]:
  if optimize_key == "yes":
    print "pyRadio cannot optimize for several array inputs, please set" +
     "optimize_key to no"
    print "Continuing noise or S/N calculations for all .param files."
  dir_names = os.listdir('.')
  tracker = 0
  for file in dir_names:
    if file.endswith(".param"):
      tracker +=1
      array_name = os.path.splitext(file)[0]
      param = n.loadtxt(file)
      if snr_key == "yes":
        tnoise_key  = "yes"
        savesnr_key  = "yes"
        print "Calculating Signal/Noise and Noise for "+file
        #specifying a range for central observing freq.
        nu    = n.linspace(lowernu,uppernu,(uppernu-lowernu)/param[6]+1)
        snr = SNR(param[0],param[1],param[2],param[3],param[4],param[5],param[6]
        ,param[7],k,nu)
      elif tnoise_key == "yes" or snoise_key == "yes" or vnoise_key == "yes":
        print "Calculating Noise for "+file
        #specifying a range for central observing freq.
        nu    = n.linspace(lowernu,uppernu,(uppernu-lowernu)/param[6]+1)
        error = Error(param[0],param[1],param[2],param[3],param[4],param[5]
        ,param[6]
        ,param[7],k,nu)
      else:
        print "No Calculation keys were set to yes, what are you expecting from me?"
        print "Please set a calculation key to yes, then I'll do some work for you"
      now1 = time.clock()
      print "Elapsed time since execution "+str(now1-start)+" seconds"
  if tracker == 0:
    print "No .param files were found, if you want calculations please",
     "enter you array parameters in a .param file"
else:
  if len(file_names) != 1 and optimize_key == "yes":
    print "pyRadio cannot optimize for multiple array inputs"
    print "Please some files in the file_names list"
    print "Thank You, Goodbye"
  elif len(file_names) == 1 and optimize_key == "yes":
    print "Starting optimization with initial parameters given in " + file_names[0]
    array_name = os.path.splitext(file_names[0])[0]
    param = n.loadtxt(file_names[0])
    nu    = n.linspace(lowernu,uppernu,(uppernu-lowernu)/param[6]+1)
#create list with optimalisation starting points
    opti      = []
#create list with corresponding optimalisation bounds
    bound     = []
```

```
    for i10 in range(len(flag)):
        if flag[i10] == 1:
          opti.append(param[i10])
          bound.append((Minimum[i10],Maximum[i10]))
    savenoise_key        = "no"
    savesys_key          = "no"
    savesv_key           = "no"
    Optimal(opti,bound)
  else:
    tracker = 0
    for file in file_names:
      if (file.endswith(".param")):
        tracker += 1
        array_name = os.path.splitext(file)[0]
        param = n.loadtxt(file)
        if snr_key == "yes":
          tnoise_key  = "yes"
          savesnr_key  = "yes"
          print "Calculating Signal/Noise and Noise for "+file
        #specifying a range for central observing freq.
          nu   = n.linspace(lowernu,uppernu,(uppernu-lowernu)/param[6]+1)
          snr = SNR(param[0],param[1],param[2],param[3],param[4],param[5]
          ,param[6],param[7],k,nu)
        elif tnoise_key == "yes" or snoise_key == "yes" or vnoise_key == "yes":
          print "Calculating Noise for "+file
        #specifying a range for central observing freq.
          nu   = n.linspace(lowernu,uppernu,(uppernu-lowernu)/param[6]+1)
          error = Error(param[0],param[1],param[2],param[3],param[4],param[5]
          ,param[6],param[7],k,nu)
        else:
          print "No Calculation keys were set to yes, what are you expecting from me?"
          print "Please set a calculation key to yes, then I'll do some work for you"
        now2 = time.clock()
        print "Time elapsed since execution "+str(now2-start)+" seconds"
    if tracker == 0.:
      print  "No .param files were found, if you want calculations",
      "please enter you array parameters in a .param file"
end = time.clock()
print "Total time elapsed: "+str(end-start)+" seconds"
```