

# Detecting faint HI clumps with Tianlai as a proof for Intensity Mapping

A. Hamie,<sup>1,2</sup>

<sup>1</sup>*Paul Sabatier University, Toulouse, FRANCE (student)*

<sup>2</sup>*IJCLab, Orsay, FRANCE (intern)*

1 July 2022

## ABSTRACT

*Context.* As a category of 21 cm cosmology, the promising Intensity Mapping (IM) technique uses radio interferometry to constraint Dark Energy via Baryon Acoustic Oscillations (BAO) detection. IM is under development, five pathfinder interferometers are built to study its viability, resolve technical challenges, and determine the shape of future dedicated interferometers.

*Aims.* This work aims to prove the required sensitivity to detect faint and close HI clumps with the pathfinder Tianlai. These faint clumps are embedded in  $10^4$  brighter foreground and instruments noise, their detection represents a proof of concept for BAO's detection with future IM large interferometers.

*Methods.* We simulated the operation of Tianlai around the North Celestial Pole (NCP) using the map-making software JSkyMap. We developed a pipeline involving a sources finder, a multi-frequency foreground subtraction method, and post-processing applying the finder on large samples of observation strategies sky maps. The required sensitivity is proved by detections statistics.

*Results.* Using simulations of operations of the central seven-dish of Tianlai with a gaussian instruments noise having  $1\text{mK}$  standard deviation. We detected with an efficiency of 70 % clumps having an  $S/N=1.5$  and with a negligible false detection rate.

*Conclusion.* Detecting nearby faint HI clumps (mJy) requires a total exposure time of six months around the NCP region.

**Key words:** 21-cm cosmology, HI clumps, source finders, map-making artifacts

## CONTENTS

- 1 Introduction
- 2 Sky maps simulation
  - 2.1 Simple maps
  - 2.2 Map-making software maps
  - 2.3 Foregrounds subtraction
- 3 Proving the required sensitivity
  - 3.1 Unresolved sources detection algorithm
  - 3.2 Map-making software artifacts
  - 3.3 Post-processing
- 4 Results
- 5 Conclusions

## 1 INTRODUCTION

The universe's expansion has shifted from power law to an exponential function of cosmic time. While Dark Energy (DE) represents the  $\Lambda$ CDM interpretation of this acceleration, alternative theories use modifications of General Relativity. An observational cosmology method for constraining the expansion history consists of tracing Baryon Acoustic Oscillations (BAO) in the matter density field.

BAO represents the baryonic remnants of frozen oscillating overdensities in the baryons-photons fluid at photons decoupling time. At that epoch, they had the size of the sound horizon and later expand as a function of cosmic time. Therefore, they manifested as an abundance of a given scale in the statistical properties of cosmological surveys. Thus, tracing BAO through redshifts gives

indications of variations in the expansion rate. BAO are often traced by visible matter distribution in optical/IR redshift surveys. These surveys locate galaxies by their stars' light. Redshift measurements are performed by photometry or spectroscopy or both. Then, redshifts are converted to distances to extract statistical properties from the 3D galaxy distribution of surveyed volumes. Since the eighties, instruments and technics have been developed to improve precision, exposure time, number of analyzed objects per exposure, field of view, and maximal measured redshift. Thus, the volume below  $z = 2$  which includes the beginning of the acceleration epoch benefits from the cumulated data from many surveys resulting high precision measurements. A better constraining for DE or any other potential modified gravity theories during the pre-acceleration epoch requires at least the same level of precision for higher redshift. Unfortunately, above  $z = 2$ , galaxies become fainter and surveys often use quasars as direct tracers or map neutral hydrogen on their line of sight with the Ly- $\alpha$  line. The rarity of these objects deteriorates the extracted statistical properties. So, here is the need for new experiments that improve surveyed samples above  $z = 2$ . Currently, flagship optical/IR experiments such as Euclid, LSST, and DESI aim to increase surveyed volumes above  $z = 3$  and will produce an unprecedented amount of data. In parallel, other non-optical techniques are under development.

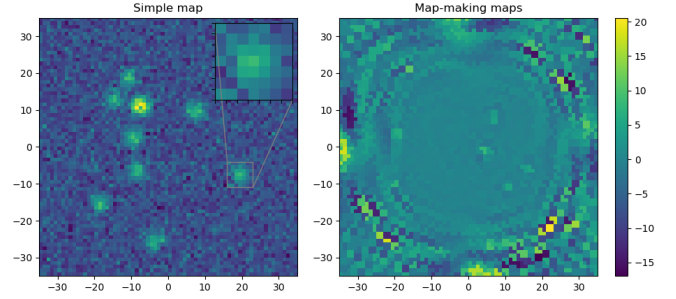
The development of new high finish mechanical and electrical technology in Radio Interferometry, particularly digitalization at Gigahertz frequencies, has emerged the Intensity Mapping (IM) technique. IM aims at surveying volumes up to  $z = 6$ . **Such data**

introduces a new tracer to available cross-correlations below  $z = 2$ . And a new opportunity on cosmological parameters estimation on higher redshifts. IM consists of mapping the intensity of the 21-cm spin-flip line of atomic hydrogen on the sky. Instead of detecting stars' light inside galaxies, IM maps their atomic hydrogen reservoir. It uses a limited angular resolution sufficient to measure the collective emission of galaxies that outline LSS without the need for resolving individual objects. In other words, IM uses sensitivity and resolution adjusted to cosmological LSS use. In addition, IM measures the 3D coordinates with a single instrument. The thin and isolated 21cm line reduces uncertainty on redshift measurement. Moreover, theoretically, 21 cm cosmology span redshift up to  $z = 150$ , future experiments aim to study the epoch of reionization and even dark ages [Pritchard & Loeb \(2012\)](#). These features and many others make IM a practical and cost-effective surveying technique. It will be proposed as a part of the post-optical cosmology era.

At present, IM is in the R&D phase and it is not yet been considered as a competitor for optical surveys. Five main experiments were built as pathfinders to future very large interferometers dedicated for IM [Cosmic Visions 21 cm Collaboration et al. \(2018\)](#). The first proof of concept was obtained from a non-dedicated instrument in 2010 using the GBT single-dish telescope around  $z \approx 1$  [Chang et al. \(2010\)](#). It detected the redshifted 21 cm emission of overlapping sources with the DEEP2 redshift survey. Pathfinders aim to build future interferometers that will operate in autocorrelation and at high redshift. IM utility is not limited to galaxy clustering, for instance, mapping neutral gas streaming around galaxies allows studying their evolution whether in the field or within clusters. SKA also will use IM for the lensing power spectrum construction at high redshift [Brown et al. \(2015\)](#). Although its different observation strategy, it still offers valuable experiences in terms of calibration and data management.

The main challenges of IM are subtracting foregrounds and getting the required sensitivity and signal-to-noise. The cosmological signal has an amplitude in the order of  $mK$ , it is embedded in instruments noise and foregrounds of 10 000 times brighter. While intrinsic sensitivity depends on instruments design, the signal-to-noise ratio is enhanced by averaging the signal on large integrating time. Increasing the integrating time with IM is complicated as it uses a transit surveying technique. The instrument is fixed and it observes the accessible zone of the sky which passes above it as the earth rotates. Thus, the time that the telescope spends daily in front of a given part of the sky is constant. In this context, the North Celestial Pole (NCP) gives the unique opportunity to point in the same direction for 24 hours. Another smarter strategy takes advantage of the redundant baselines in a radio antennas array. That is why future IM interferometers consist of thousands of dishes [Slosar et al. \(2019\)](#). And that is why it requires a proof of concept before its implementation. The work described in this article uses the simulation of the operation of the pathfinder Tianlai [Wu et al. \(2020\)](#) around the NCP region. It aims to prove the required sensitivity to detect faint sources ( $mK$  amplitude) with the IM regime. These sources are faints and close HI clumps. Their detection at low redshift is a preliminary stage for detecting the BAO with future interferometers at high redshift.

Proving the  $mK$  sensitivity required a three phases procedure. The first step consists of simulating sky maps that could be made simply by component addition or via map-making software. We used JSkyMap, a visibilities simulator, and map-making software



**Figure 1.** The left panel shows a synthetic map of the NCP region with an area of 300 degrees square. The sources have a perfect gaussian shape and a white gaussian noise is added to every pixel. The right panel shows the results of the map-making process. Gaussian noise is added to visibilities. Due to artifacts, on the map, noise is not gaussian, and sources are deformed.

for transit radio interferometer. At this stage, it's crucial to compare both maps type to evaluate the map-making artifacts and their effect on raw data. In parallel, we developed a template fitting foregrounds subtraction method based on the difference in spectral properties between signal components. The second phase consists of building an unresolved source detection algorithm based on sliding kernels convolution. Instead of using a predefined shape for the PSF, we built a flexible 3 kernel model that divides pixels of potential sources into 3 categories. This will allow a multi-scale study. The third phase uses a post-processing algorithm that evaluates a large sample of observation strategies maps. These strategies affect mainly the S/N ratio. The performance of the interferometer is evaluated using efficiency curves, false detection curves, and reconstructed sources fluxes.

This paper is structured as follows: after introducing the position of IM in observational cosmology and the purpose of our work, Section 2, describes the type of used maps and the foreground subtraction method, including a brief introduction to the map-making process. Section 3 describes our method. The detection algorithm and the statistical study of its efficiency. In addition to a visualization of the map-making effect on raw data. Section 4 discuss obtained results on sensitivity and detection statistics. Finally, 5 the conclusion put the results in its larger frame.

## 2 SKY MAPS SIMULATION

### 2.1 Simple maps

These maps contain unresolved sources and an instrument's noise. They are built with python as the following: a binary map is created with arbitrary distributed hot pixels. Those pixels are then convolved with a Gaussian PSF. After that, a Gaussian white noise map is added to the main map. We used these maps during the creation and optimization of the sources detection algorithm and to compare with maps resulting from map-making software in order to analyze its effects on raw data.

### 2.2 Map-making software maps

A single telescope received signal consists of intrinsic sky events convolved with its transfer function. In a two-dish radio interferome-

ter, raw received data is named visibility. It is defined as the temporal mean of the correlation of the two feeds (output voltages) of the dishes. The straight line that links these dishes is called the baseline. Similarly, an  $N$ -dishes interferometer consists of  $N \times (N - 1)$  visibilities. When all these dishes observe the same event in the sky, the total beam (transfer function) of the interferometer includes each baseline perspective. While visibilities corresponding to redundant baselines enhance the S/N, visibilities resulting from autocorrelations of each dish give signals offset. Thus, the interferometer configuration and its observation strategy dictate the quality of received data.

The first step when analyzing cosmological data is turning the raw data into manipulable structures. Therefore, visibilities are cleaned and sorted in time and frequency bins. Information can be extracted directly from visibilities variation, but often they are used to compute spherical sky maps. These maps can be studied locally through gnomonic projections. For cosmological use, after foreground removal, a power spectrum is reconstructed to extract cosmological parameters from statistical properties.

$$\mathcal{V}_{ij}(t) = \iint I(\vec{n}) L_{ij}(\vec{n}, t) d\vec{n} \quad (1)$$

Assuming expression 1 for the visibility  $\mathcal{V}_{ij}(t)$  seen by the pair of dishes  $i$  and  $j$ , where the vector  $\vec{n}$  is the pointing direction to the sky,  $L_{ij}(\vec{n}, t)$  represents the beam of the baseline and  $I(\vec{n})$  is the intrinsic sky intensity. Thus, data is modeled as in equation 2.

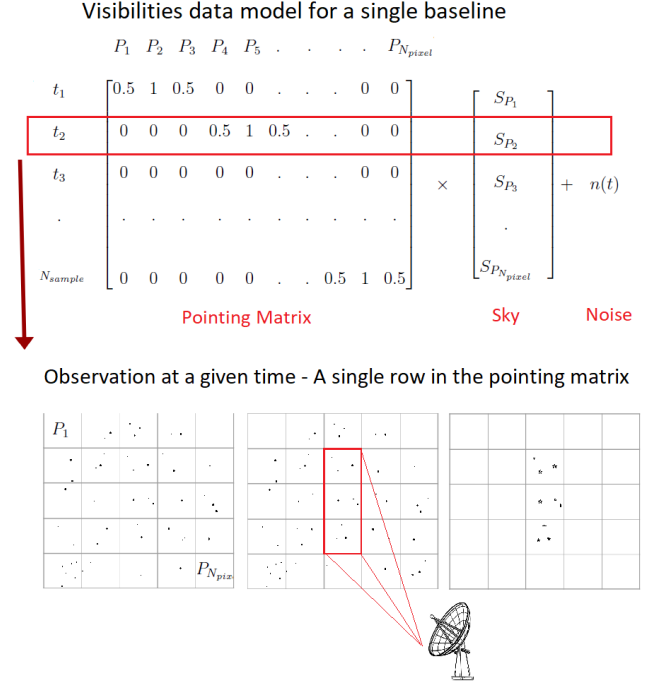
$$[\mathcal{V}_{ij}(t)] = L_{ij}(t) \times [I(\vec{n})] + [n_{ij}(t)] \quad (2)$$

The received visibilities vector  $[\mathcal{V}_{ij}(t)]$  of all baseline during all the time of the observation is equal to the convolution of the pointing matrix  $L_{ij}(t)$  and the sky intensity vector  $[I(\vec{n})]$  added to instruments noise vector  $[n_{ij}(t)]$ .  $L_{ij}(t)$  depend on beam and observation strategy. It consist of  $N_{pixel}$  column and  $N_{sample} \times N_{baseline}$  rows. Figure 2 shows the data structure for a single visibility. So, knowing visibilities and assuming a white Gaussian noise, the determination of the unknown sky  $I(\vec{n})$  is then the solution of a standard inverse linear problem. The difficulty here is the huge dimension of  $L_{ij}(t)$ . A typical interferometer having hundred of baselines using a sampling time of a few seconds and observing for a few hours can result in a matrix of up to  $10^5$  rows. Shifting to the spherical harmonic space broke up the problem into a set of independent linear systems, one for each spherical  $m$ -mode as in equation 3.

$$[\mathcal{V}]_m = L_m \times [I(\ell)]_m + [n]_m \quad (3)$$

$$[I(\ell)]_m = \mathcal{H}_m [\mathcal{V}]_m \quad (4)$$

Then, the solution is written as in equation 4 where  $\mathcal{H}_m$  is the noise weighted Moore-Penrose pseudo-inverse which is derived from known matrices by maximum likelihood Barata & Hussein (2012), Ansari et al. (2020). Finally, the sky intensity is computed by an inverse Spherical Harmonics Transform (SHT) on the estimated spherical modes coefficients  $[I(\ell)]_m$ . Shaw et al. (2014) An important aspect of the map-making process is transferring data from the interferometer pixel scheme into a grid of pixels on the surface of the celestial sphere compatible with computation and manipulation of sky maps. The new scheme should be representative and optimize information lost and at the same time facilitate spherical harmonics decomposition. For these reasons, and others related to programming we chose the HEALPix pixelization scheme Górski et al. (2005). In this context, the surface of the HEALPix pixel must be equal or inferior to the resolution of the interferometer which is the surface of the old pixel. (????????????????????????????????)



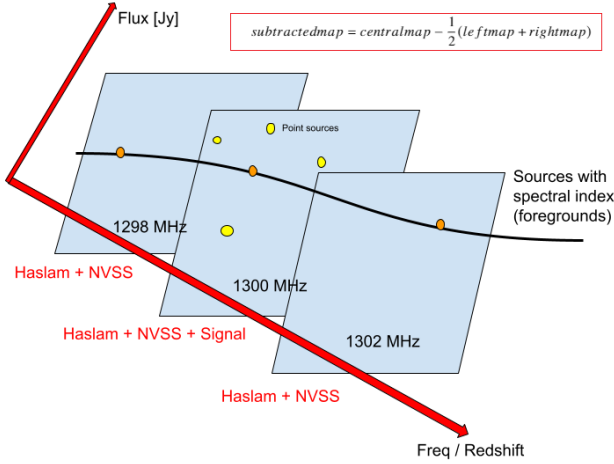
**Figure 2.** The field of view of the survey is seen as a collection of pixels having the size of interferometer angular resolution ( $\lambda/\text{longest baseline}$ ). The beam represents a set of pixels that can be seen with the interferometer at the same time. Each row in  $L_{ij}(t)$  describes the seen and unseen pixels by the interferometer beam at a given time. Unseen pixels are assigned with zero. Non-zero pixels inside the beam are weighted according to the beam shape. The product of this row with the sky vector gives zero or seen visibilities. As the earth rotates, over time, the transit interferometer survey the sky, more rows in  $L_{ij}(t)$  reveal more pixels, up to the point that the beam has drifted all pixels to give the full visibilities vector of the survey's field of view.

We used the JSkyMap visibility simulator and map reconstruction software package for interferometers operating in transit mode Zhang et al. (2016). We used it to produces full spherical maps with HEALPix pixelization. A gnomonic projection of the NCP region is made later. JSkyMap simulate visibilities from input catalogs or maps, it add gaussian noise to them, and them compute spherical maps. It give the ability to apply some filters in spherical harmonics space. Figure 1 (right panel) show simulated NCP map with JSkyMap.

### 2.3 Foregrounds subtraction

Foreground subtraction has crucial importance in IM, as the targeted mK signal is embedded in 10000 times brighter foreground. Compared to CMB for instance, the IM signal doesn't have privileged frequency where it domine the sky. To reach the goal of intensity mapping as surveying technique up to redshift  $z=6$  foregrounds must be subtracted through the cumulation of all spaned redshift.

In 2.2 we used single frequency data. The reality is that visibilities are stored in nearby frequency bins. Our basic subtraction method lay on the difference in spectral properties of different components of the signal. The foreground consists mainly of the diffuse galactic synchrotron and extragalactic radio sources



**Figure 3.** The three planes foreground subtraction method

that could be AGN or high SFR galaxies. These signals have a spectral correlation through redshifts that are their synchrotron spectral index. In contrast to IM signal that is punctual in frequency space.

The subtraction method could lay on an inverse of mixing matrix similarly to sky intensity extraction in 2.2. Galactic diffuse foreground has a smooth spectral correlation. This allows the subtraction by filtration of low frequencies that affect the useful signal too. We used a template fitting method that consist of assembling the multichannel foregrounds in a template and subtracting it from the original skymap. Figure 3 show the subtraction method that simulate 3 sky maps as the following:

- **Central Map:** This map results from map-making process. It uses simulated visibilities which include: Gaussian noise, extrapolated HASLAM 408 Mhz [Haslam et al. \(1982\)](#) map as galactic foreground, NRAO VLA Sky Survey (NVSS) catalogs [Condon et al. \(1998\)](#) as extragalactic foreground, and few nearby HI clumps added manually.
- **Two Sides Sky Maps:** These maps contain all central map components except HI clumps. They have frequencies equal to central map frequency  $\pm \epsilon$ , where  $\epsilon$  is a small deviation.

$$\text{subtractedmap} = \text{centralmap} - \frac{1}{2}(\text{leftmap} + \text{rightmap}) \quad (5)$$

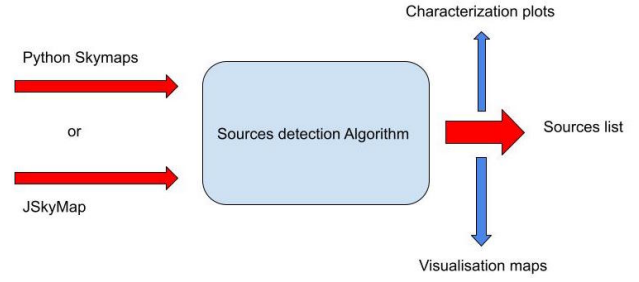
Foregrounds are subtracted according to equation 5. This is a simplified method. Averaging on side maps will be replaced with the synchrotron frequency distribution.

### 3 PROVING THE REQUIRED SENSITIVITY

#### 3.1 Unresolved sources detection algorithm

As shown in figure 4 takes skymaps as input to give sources catalog. The performance is tracked by many characterization plots and maps visualization at each phase.

**Input-** The algorithm takes spherical maps with HEALPix scheme pixelization in FITS format. Then an NCP gnomonic projection is made. **Background estimation-** Intensities of all pixels of the map are sorted in a histogram. The background is the median



**Figure 4.** This shows the algorithm functionality

intensity. The algorithm has the feature of  $\sigma$ -clipping that removes outliers from symmetric distributions during median computation. **Detection-** Usually, source detection is performed by convolving the map with the shape of the source in question. In the case of an unresolved source, this is the PSF. In the absence of predefined PSF we used a flexible set of 3 kernels to recognize it. It divides pixels of potential unresolved sources into 3 levels as shown in figure 5. the central kernel containing more than 70% of the flux density, the second contains the rest, and the third represents the local background. Kernels sizes are among algorithm parameters to obtain a multiscale approach. A given pixel is considered as a part of a possible source if it verifies condition 6. The local background represents the median of  $\text{ker3}$  and  $\sigma_{bg}$  correspond to its standard deviation. The threshold  $n$  is the parameter that drives the efficiency and the false detection rate of the algorithm. The output of this step is presented in figure 12 panel 1. Pixels of potential sources are located as agglomeration on the map. I mention that the detection uses the local background, the previously mentioned global background is used to set the S/N of simulated maps.

$$\text{mean}(\text{ker1}) > \text{background} + n \times \sigma_{bg} \quad (6)$$

**Clustering-** This step consists of extracting sources from pixels agglomeration obtained from the detection stage. Knowing that sources are all unresolved, we skip the differentiation between point and extended source. So, we define an aperture for potential sources around detected pixels. First, S/N of each pixel is computed, it is equal to the ratio of the  $\text{mean}(\text{ker1})$  and the standard deviation of  $\text{ker3}$ . Then detected pixels are sorted in decreasing S/N. Then in a hierarchical form, the algorithm targets the highest S/N pixel inside the aperture, searches for their 8 connected neighbors (among detected pixels). Then detached agglomeration inside the aperture is considered as one source, attached pixels outside the aperture are suppressed. This hierarchic operation took place up to the lowest S/N pixels. at each iteration used pixels are removed from the remaining list to not be used again. This step is shown in panel 2 of figure 12. **Barycenter and Photometry-** Figure 12 panel 3 and 4. Barycenters are computed to locate the sources in pixels coordinate and in RA dec system. Then the fluxes are computed by integrating all the intensity inside a given kernel consisting of  $N$  pixels. This kernel size varies between simple maps and map-making software maps. Further details in 3.2. Then the background is subtracted from this flux as in  $F = \sum_0^N I_i - N \times \text{localbg}$ . **Filtering-** This step consists of removing tiny sources and keeping sources only inside a given area in order to remove boundary effects (panel 5). **Exporting-** The final catalog contains RA dec coordinates and flux density in Jansky and others.



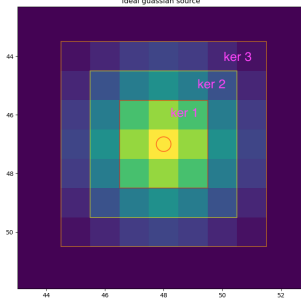


Figure 5. The three kernels model for unresolved sources

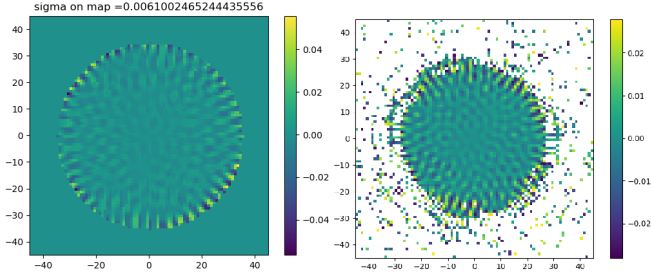


Figure 6. ...

### 3.2 Map-making software artifacts

The right panel in figure 1 shows an example of maps resulting from the map-making process. This results in deformations in the background as well as in sources. For the background, the transfer function of the map-making process transforms the Gaussian noise in visibilities into symmetric but not gaussian noise on the map. The pseudo-inverse method in spherical harmonics space increase the amplitude of the noise, and its effect on  $\sigma_{noise}$  is complicated to estimate because of the formed rings around the center of the map. These rings resulted from superpositions of the artifacts that feel every noisy tiny source in visibilities. This superposition is constructive in the radial direction and destructive elsewhere. This phenomenon is highly amplified if counting the autocorrelation visibilities in the survey. Those visibilities are very sensitive to the variation in the signal offset (background). Another amplifying factor is the diffuse sources that present low  $\ell$  feature, such as remnant of the galactic synchrotron that is weak around NCP. So, after rejecting autocorrelations and performing filtration of low  $\ell$  values, the background can be estimated after a geometrical slicing for the usefull zone in the map as shown in figure 6. The limitation here is that this zone have to be determined by the user for each variation of declination in the observation strategy. An automated solution will be an iterative sigma clipping that remove the distorted part. Figure ?? show a tentative. This deformation in background affect partially the concept fo 3 kernels detection as it use local background but estimating the global background will dictate the choice of used maps in the post-processing phase in 3.3.

Regarding sources, regardless of the true shape of the instrument PSF, sources do not have the perfect Gaussian shape. The map-making process increased their amplitude and created negative fluctuations in their temperature. An example is shown in figure ?? . These negative parts added to the negative fluctuations in

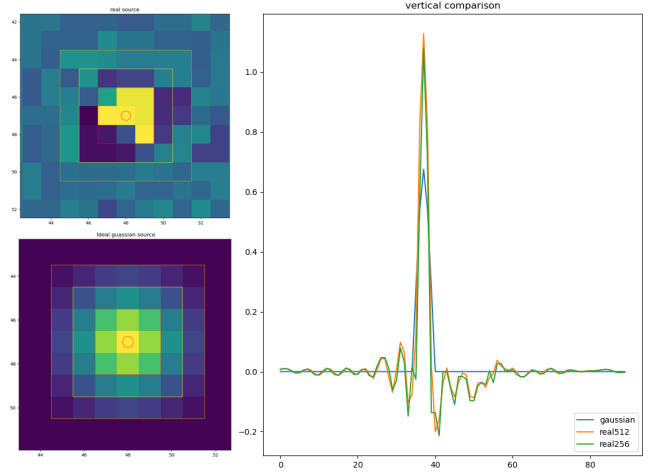


Figure 7. ...

the background have increased the spurious detection. I added supplementary conditions to condition 6 as the following:

$$mean(ker1) > mean(ker2) \quad (7)$$

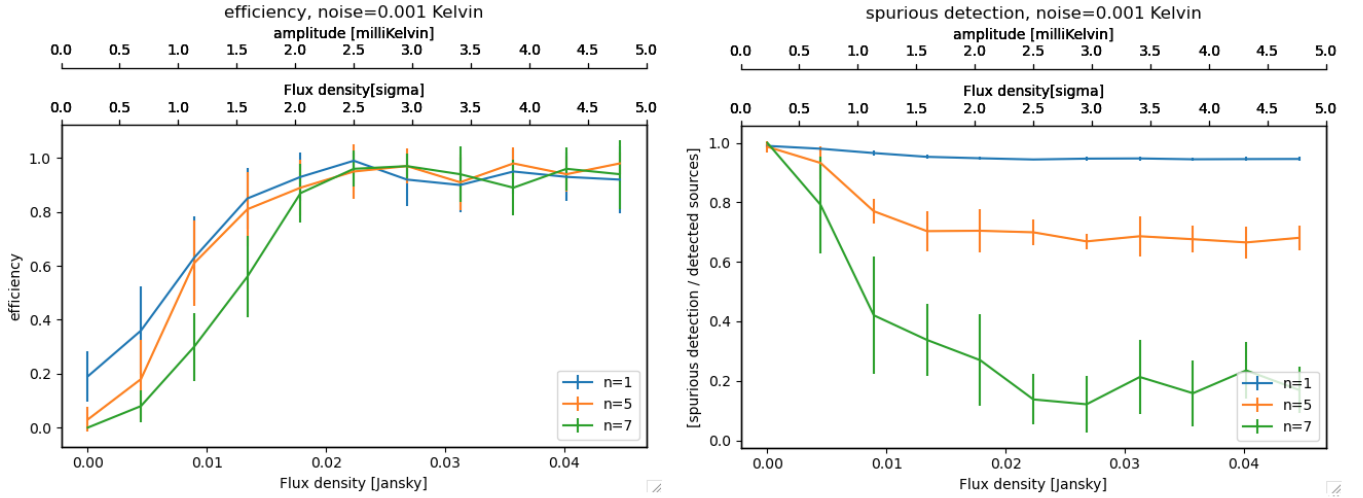
$$mean(ker1) > mean(ker3) \quad (8)$$

The flux density computation of detected sources is affected too. I used an aperture with the size of the central positive pixels instead of the aperture used with gaussian sources. Comparing the flux of original low amplitude and high sigma Gaussian sources to the flux of deformed high amplitude low sigma sources, the deformation doesn't conserve the flux.

### 3.3 Post-processing

This phase analyzes the detection efficiency of  $mK$ -amplitude sources in terms of their S/N. So, the work is to simulate a large sample of maps of different S/N. Then the detection algorithm is applied to this set. For each map, the detected sources are sorted to extract the efficiency (number of detected simulated sources/number of simulated sources), spurious detections, and comparing reconstructed detected flux and original simulated flux. Matching detected sources with sources from the catalog uses their separation in pixel for simple maps and separation in the angular distance for maps resulting from map-making. The matching uses flux comparison also to not count coincidences.

The maps sample is created as the following. F each value of instruments noise, we simulate  $N_{maps}$  maps having each one  $N_{sources}$  sources with equal flux-density. Fluxes values span S/N between 0 and 5. The foreground subtraction method requires repeating this for many frequencies. Furthermore, a better study of the relationship between efficiency and the false detection rate entails trying many values for the threshold  $n$ . As a result, we have  $N_{noise} \times N_{frequency} \times N_n \times N_{flux} \times N_{maps}$  map to study the detection statistics. Regarding flux density, input fluxes given to the map-making software are expressed in Jansky, resulting sources on the maps are manifested in kelvin. The detection algorithm integrates pixels belonging to a given source to compute flux density in kelvin. Their amplitudes are related to fluxes using the approximation 9.



**Figure 8.** **Top left:** Efficiency as function of sources fluxes density. **Top right:** ratio of spurious detection pixels to total number of pixels as function of sources density fluxes. **Bottom left:** reconstructed fluxes to generated fluxes

$\sigma_{source}$  correspond to the sigma of a perfect gaussian source that has the same dimension of detected deformed sources.

$$Flux[K] = 2 \times \pi \times Amplitude[K] \times \sigma_{source}^2 \quad (9)$$

The following equation convert fluxes in Kelvin to Jansky.  $\lambda$  is the redshift wavelength of 21cm emission.  $\Omega$  is the size in radian of map's pixel.(???)

$$T[K] = \frac{S[Jy] \times \lambda^2}{2K_B \times \Omega} \quad (10)$$

## 4 RESULTS

The smaller the threshold  $n$ , the more sample there is to analyze. This entails more detected sources, but more false detection too. For instance, if the false detection rate is very high, then there is a high possibility that all the sources detected by following the location criterion can be only coincidences. What will determine the right threshold is the reconstructed flux density of detected sources. Flux reconstruction is critical. It points to the deformation of skymaps by the map-making process. This includes sources and the background. For now, I won't describe these details.

Figure 8 show the characterization plots of the algorithm using synthetic maps. The **top left** plot shows the efficiency as a function of sources fluxes for many values of  $n$  and instrument's noise of 1 mK. The lower horizontal fluxes density axis is chosen according to the S/N ratio. A source with S/N=1 has an amplitude of 1 mK. Assuming a perfect Gaussian shape, its flux density in Kelvin is given by  $2 \times \pi \times Amplitude \times \sigma_{source}^2$ , then this flux density is converted to Jansky using the formula  $T[K] = \frac{S[Jy] \times \lambda^2}{2K_B \times \Omega}$ , where  $\Omega$  is the size in radians of one pixel of the rectangular map ( $\frac{1}{5} \times \frac{\pi}{180}$ )<sup>2</sup>. Each point and its error bars corresponds to 10 iterations of NCP skymaps. Each map contains 10 sources with the same flux. It has a size of  $90 \times 90$  pixels and a resolution of 12 arcmins. Excluding peripheric zones, the map corresponds to  $196 \text{ deg}^2$  (radius of 7 deg). The **top right** plot shows the false detection pixels normalized

to the number of pixels in the survey's field of view. The **bottom left** plot shows the relation between initial fluxes and reconstructed fluxes after sources detection. Although the efficiency curve of  $n = 7$  is slower than the other cases, we can see that the false detection rate is negligible, and the reconstructed fluxes are very coherent. I mention that fluctuations in efficiency values, after reaching the plateau results from the limitation of the algorithm when it faces overlapped sources. Figure 10 shows a scatter plot which describe the reconstructed flux in function of instruments noise.

Figure 11 presents the algorithm characterization plots using maps resulted from the map-making process. The **left** curve shows the efficiency using a map containing only sources and instruments noise. The **right** curve shows efficiency for maps containing foregrounds, sources, and noises. This Efficiency is measured after foreground subtraction. The main difference here is that for 1 mK of injected noise in visibilities I measured 6 mK on the map. Therefore, adopting the  $n = 7$  case on the right plot, the interferometer can detect sources with  $S/N = 1.6$  with an efficiency of 70 percent with a negligible spurious detection rate.

## 5 CONCLUSIONS

To prove the required sensitivity to detect faint HI clumps. This is a preliminary step to detect BAO with larger future interferometers. We simulated the operation of the 7 central dishes of Tianlai around the NCP region. As a result, we can detect clumps HI in the order of 1 mK. This corresponds to 10000 sec of integrating time. So a 2,7 hour per sky direction, and 42 days of observation for each declination. I worked with 4 declinations, therefore, around 6 months of observations.

## REFERENCES

- Ansari R., et al., 2020, *MNRAS*, **493**, 2965
- Barata J. C. A., Hussein M. S., 2012, *Brazilian Journal of Physics*, **42**, 146
- Brown M., et al., 2015, in *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*. p. 23 ([arXiv:1501.03828](https://arxiv.org/abs/1501.03828))

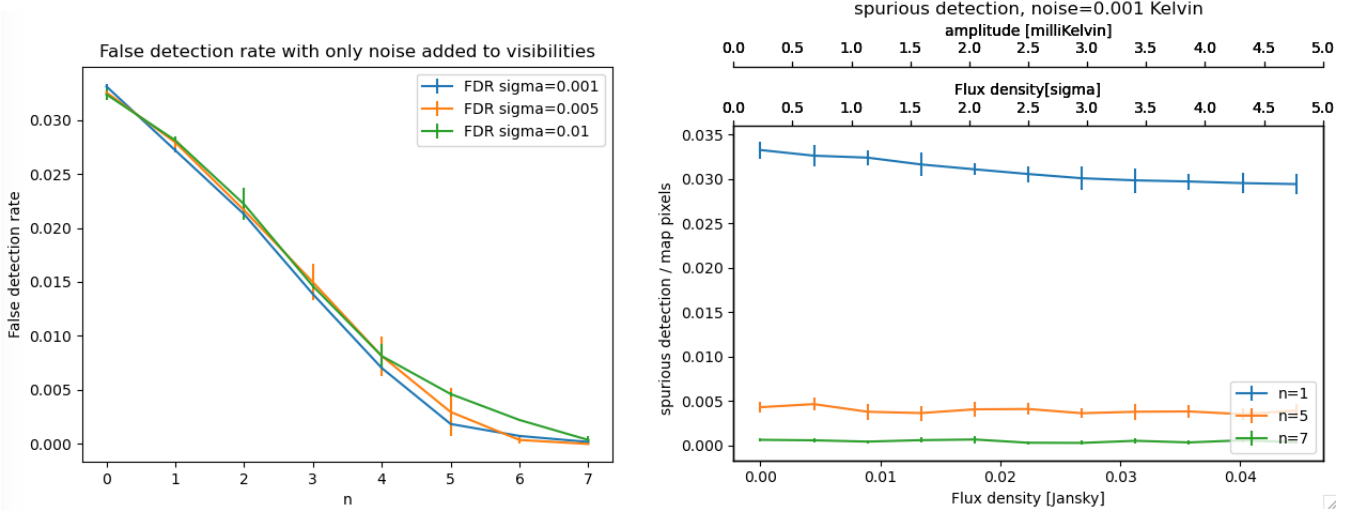


Figure 9. ....

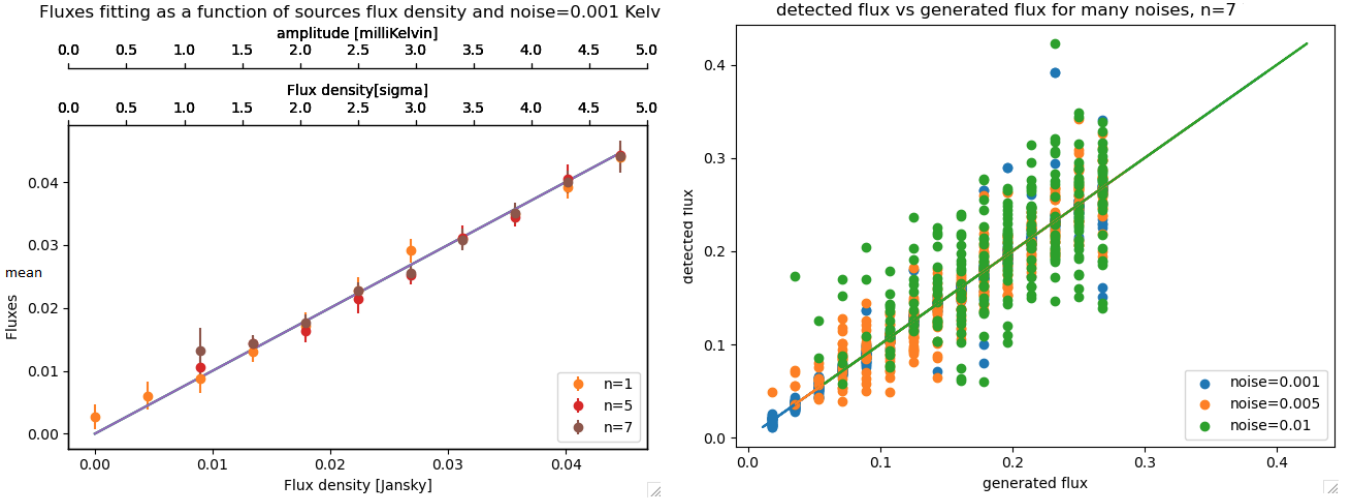
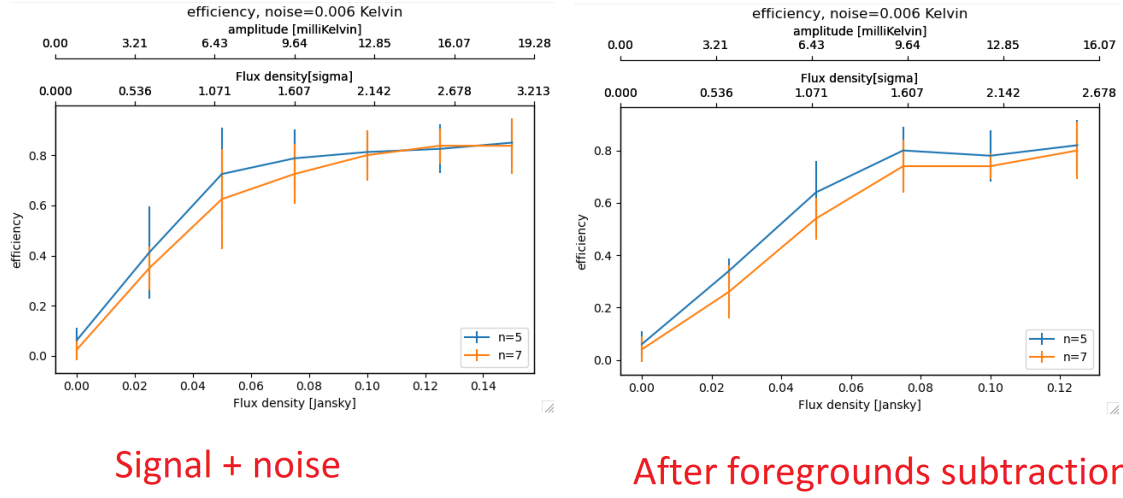
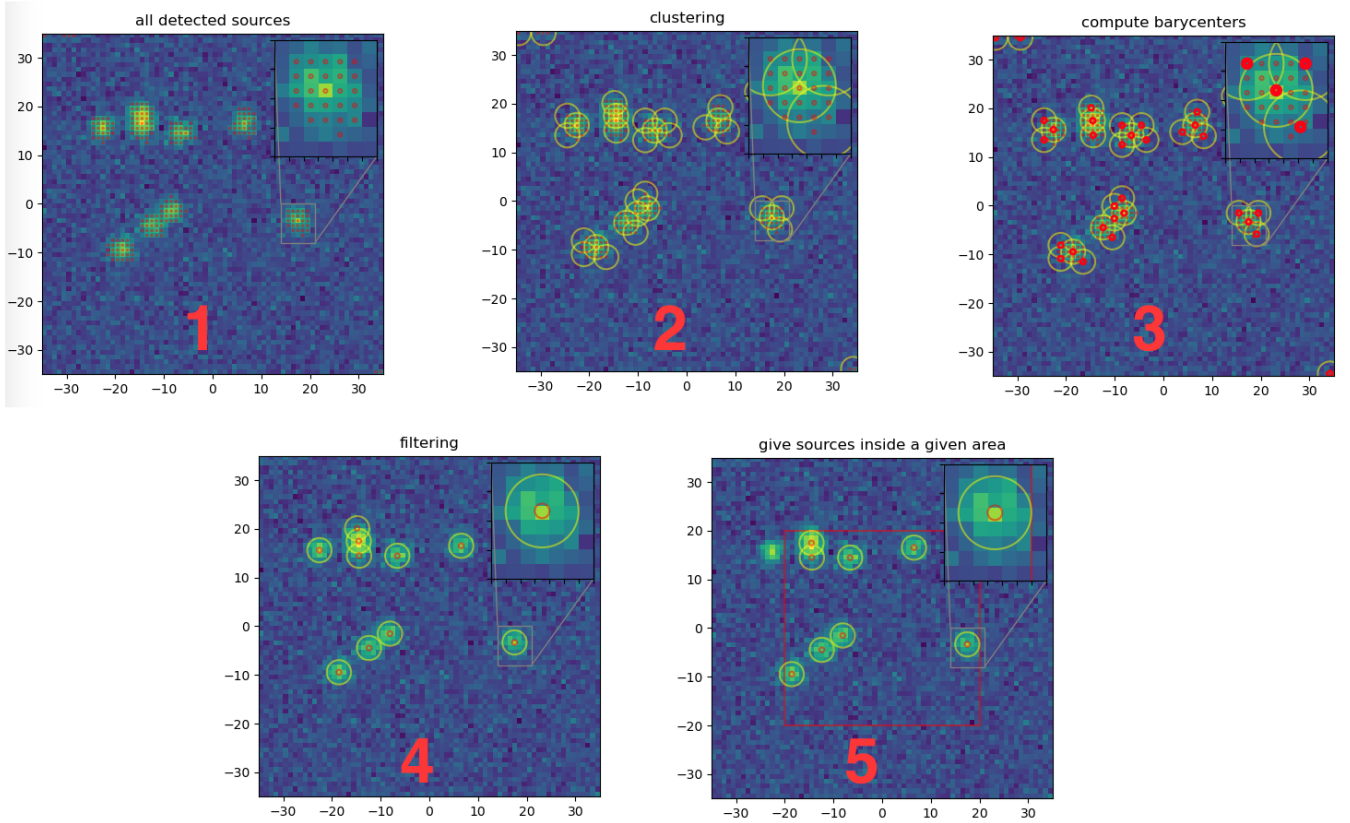


Figure 10. Left: A scatter plot shows the reconstructed fluxes as function of generated fluxes. Right: Efficiencies for many values of instruments noise.

- Chang T.-C., Pen U.-L., Bandura K., Peterson J. B., 2010, *Nature*, **466**, 463  
 Condon J. J., Cotton W. D., Greisen E. W., Yin Q. F., Perley R. A., Taylor G. B., Broderick J. J., 1998, *AJ*, **115**, 1693  
 Cosmic Visions 21 cm Collaboration et al., 2018, arXiv e-prints, p. [arXiv:1810.09572](https://arxiv.org/abs/1810.09572)  
 Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *ApJ*, **622**, 759  
 Haslam C. G. T., Salter C. J., Stoffel H., Wilson W. E., 1982, *A&AS*, **47**, 1  
 Pritchard J. R., Loeb A., 2012, *Reports on Progress in Physics*, **75**, 086901  
 Shaw J. R., Sigurdson K., Pen U.-L., Stebbins A., Sitwell M., 2014, *ApJ*, **781**, 57  
 Slosar A., et al., 2019, in *Bulletin of the American Astronomical Society*. p. 53 ([arXiv:1907.12559](https://arxiv.org/abs/1907.12559))  
 Wu F., et al., 2020, arXiv e-prints, p. [arXiv:2011.05946](https://arxiv.org/abs/2011.05946)  
 Zhang J., Ansari R., Chen X., Campagne J.-E., Magneville C., Wu F., 2016, *MNRAS*, **461**, 1950



**Figure 11.** **Left:** Efficiency as function of sources fluxes density for map-making maps containing only HI clumps and instruments noise. **Right:** Efficiency as function of sources fluxes density for map-making maps containing HI clumps, instruments noise, and foregrounds. Efficiency is measured after foreground subtraction.



**Figure 12.** These maps shows the different steps of the source detection algorithm. First step test if each pixel of the map verify the detection criterias, at the end of this step sources are represented as agglomerations of neighbours pixels. The second step consist of clustering these agglomeration and then compute barycenters. Finally, filter tiny sources, and exclude sources outside a given field of view.