Section A: Overview of the Research Project

1. Title of the research project: HI Intensity Mapping with MIGHTEE

2. Broad area of research (Engineering or Science): Science

3. Academic level of research project (Masters or Doctoral): Doctoral

4. Abstract of research project

Probing the neutral hydrogen content of the Universe is one of the primary goals in both galaxy evolution and cosmology. Neutral gas eventually turns into stars after moving through the molecular phase, playing a crucial role in galaxy formation and evolution across cosmic time. At the same time, HI is a great tracer of the underlying dark matter distribution and can thus be an exquisite observational tool for Cosmology. Usually this is done on large scales (using single dish in the case of MeerKAT). However, due to the compact core of MeerKAT, we can probe semi-linear scales with the interferometer. This project will use data from MIGHTEE to measure the HI power spectrum and the cross-power spectrum with galaxy surveys on this quasi-linear scales using the intensity mapping technique. This will allow a novel way to probe the HI mass function as well as to constrain cosmological parameters. Measurement of these statistical quantities will provide invaluable information to compare with current state of the art simulations.

5. Primary supervisor’s details

    a. Full name of primary supervisor: Prof. Mario Santos

    b. Primary supervisor’s email address: mgrsantos@uwc.ac.za

    c. University where primary supervisor is employed: University of the Western Cape

6. Co-supervisor/Research supervisor’s details (if relevant)

    a. Full name of co-supervisor: Dr. Matt Jarvis

    b. University where co-supervisor is employed: Oxford

Section B: Details of Research Project

1. Scientific merit

Intensity mapping of the neutral Hydrogen line (HI IM) has been proposed as an innovative technique to probe the large scale structure of the Universe and deliver precision constraints on cosmology (see https://arxiv.org/abs/1501.03989 for a review). It relies on observations of the sky intensity from the integrated 21cm line emission over a wide sky area. For a reasonably large 3D pixel in solid angle and frequency interval, we expect to have several HI galaxies in each pixel so that their combined emission will provide a strong signal. Fluctuations in the observed intensity of this redshifted HI emission will follow fluctuations in the underlying matter density as traced by the HI emitting galaxies, allowing the density field to be reconstructed on sufficiently large scales from intensity maps. Although with low angular resolution, it is well matched to the scales required for cosmology. Moreover, as we are probing a specific emission line (21cm) we have immediately a one to correspondence between observed frequency and redshift, which delivers very high redshift resolution. Such survey is much less time consuming than a spectroscopic galaxy survey, which requires a high signal to noise detection of each individual galaxy. This technique has been proposed as the main path to probe cosmology with SKA1-MID (https://arxiv.org/abs/1811.02743). Given that the SKA1-MID baselines are not small enough to probe the large cosmological scales, the plan is to do a survey using the auto-correlation data from each dish.

MeerKAT, with its 64 dishes, can already provide exquisite constrains on cosmology using this technique and prepare the stage for SKA1 (see https://arxiv.org/abs/1709.06099). There are ongoing efforts to test the method with pilot surveys with the ultimate goal of making a first detection of the HI power spectrum on large scales. This in turn should allow to set important constraints on Baryon Acoustic Oscillations and redshift space distortions, thus allowing us to probe the evolution of dark energy. With MeerKAT, using the single dish information, we are constrained to scales > 1 deg. At z~0.5 this translates to scales of about 30 Mpc (co-moving). However, MeerKAT has a reasonable dense core, allowing to also probe relevant cosmological scales using the interferometer data. This will provide a useful complementarity to the single dish data (see figure 1). It also has the advantage that we can use standard calibration techniques and observing strategies.

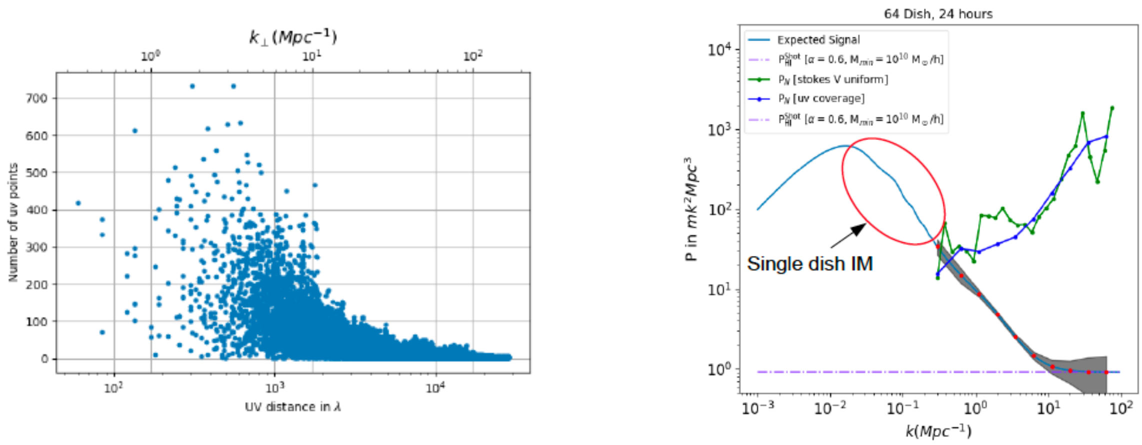

Figure 1. *Left*: the number of MeerKAT baselines as a function of uv distance or k perpendicular for a 24h track. *Right*: expected constraints on the HI power spectrum for a single 24h pointing at z~0.3 (grey area).

MIGHTEE is one of MeerKAT's Large Survey Projects. The observations will provide radio continuum, spectral line and polarisation information, thus giving us an unique view of galaxy evolution and structure formation (https://arxiv.org/abs/1709.01901). With about 1000 hours in L-band, it aims to reach a sensitivity of 1uJy in continuum over 20 deg2 on well known fields. Such data is perfect to study HI on small cosmological scales using the HI intensity mapping technique in interferometer mode. Figure 1, shows what constraints can be achieved on the HI power spectrum with a single pointing of 24h (e.g., the COSMOS field). This will be a worlds first. Larger sky areas will allow to further beat cosmic variance. Note that for this technique to be used, the 4k mode should be enough. The sensitivity allows to even probe the power spectrum at z~0.5 in L-band.

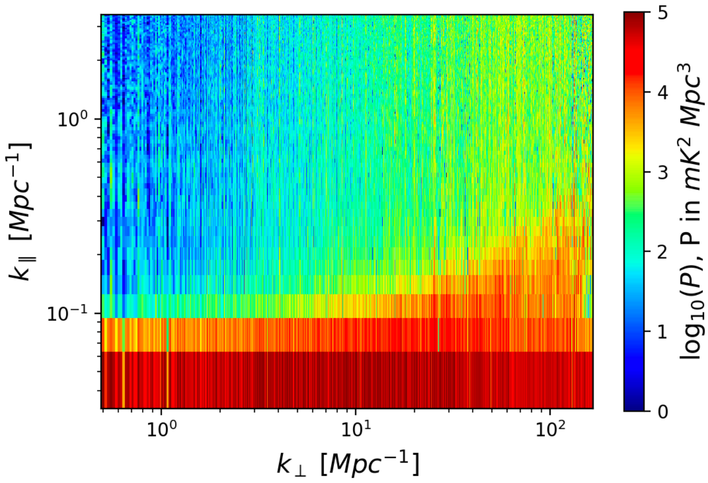

Figure 2: the 2d power spectrum from a simulation of a MeerKAT observation (L-band, 6 hours, one pointing). It includes the HI signal, thermal noise and point sources in the COSMOS field.

This PhD project aims to study the HI content of the Universe using the HI intensity mapping method with MIGHTEE. The student will work on both simulations and data analysis. One of the core objectives will be to help with the measurements of the cross-correlation signal between HI intensity mapping and optical galaxy surveys on the same fields. Such cross-correlation measurement should be much cleaner from systematics and has already be attempted using single dishes with the GBT (https://arxiv.org/abs/1007.3709) and Parkes (https://arxiv.org/abs/1710.00424) telescopes, which measured an unexpected lack of clustering in low-redshift 21-cm intensity maps cross-correlated with 2dF galaxies. The high signal to noise measurements with MIGHTEE on small scales will provide useful insights into the HI content of such galaxies and allow to better understand the differences between red and blue galaxies (or the correlation between HI content and star formation rates, e.g. https://arxiv.org/abs/1512.04189). On these scales, we can also expect to probe cosmological quantities such as the spectral index of fluctuations, constraints on warm dark matter or the measurement of the quasi-linear redshift space distortions.

In summary, the PhD project will entail the following: 1) development of simulations of the observation pipeline to compare to the MIGHTEE data. This will be improved with realistic HI galaxy simulations that can be used to test the cross-correlation with the external galaxy catalogues. 2) analysis of the MIGHTEE data and measurement of the cross-correlation signal with galaxy surveys. 3) provide constraints on cosmology and the HI simulations using such data.

2. Feasibility

The project will start with the development of simulations that can emulate the MIGHTEE observations. Such simulation pipeline is already being developed so the student is not expected to do this from scratch. See figure 2 for the output 2d power spectrum of the current simulation of the COSMOS field including HI, noise and point sources. Instead, the plan is for the student to add specific improvements to the simulation. One example is testing the effects of the MeerKAT primary beam on the contamination of the signal. This can be compared with the calibrated data and to check if further improvements are needed. Such work is expected to lead to a paper. A further improvement on the simulation will be to add more accurate HI simulations which also include information on the optical galaxies. Ultimately the HI simulations should allow for different cosmological parameters so we can also test our ability to measure them. Note that the scales probed by the MeerKAT interferometer (30 Kpc to 30 Mpc) are a really good fit to state of the art simulations. This part of the project will be an ongoing process involving several people specialised in HI simulations. At least one publication will also be expected from this component, although whether the student will be first author on this one will depend on how the second phase of data analysis will evolve. The student is expected to work on simulations for the first half of the PhD, while at the same time starting to learn how to analyse the MeerKAT data.

The second part of the project relates to data analysis of the MIGHTEE data and extraction of the HI intensity mapping signal. Again, this is already an ongoing process and the student is not expected to do this alone. In particular we can use the data already calibrated by the MIGHTEE team and we are already developing techniques to extract the HI power spectrum. The student is expected to focus on the cross-correlations with galaxy catalogues and comparison with the previous simulations. Depending on how the student progresses, they can either lead a "detection" paper or one of the papers focusing on constraining cosmological and astrophysical quantities. We are confident that the data will be good enough for such measurements, in particular when looking at the cross-correlations. However, if we find issues in the data, the student will still be able to publish a paper setting upper limits on the power spectra comparing with the simulations pipeline and test what effects could be contaminating the signal. If time allows, the student will also have the chance to work on HI stacking. This is intimately related to the HI intensity mapping technique and it will be a natural extension to the PhD project. In particular we can combine the HI mass function measurements typical of stacking with the HI clustering used in intensity mapping (e.g. combine the one-point and two-point correlation function in stacking). This will be done in a Bayesian framework as already explored in https://arxiv.org/abs/1908.05316 and https://arxiv.org/abs/1907.10404.

This project has been approved by MIGHTEE and the student will have access to the data. The supervisor is co-chair of the MIGHTEE "Noise and IM" working group and the co-supervisor is co-PI of MIGHTEE. Both the simulation and data analysis pipeline will be done in the IDIA/ILIFU system (the current pipelines have been already deployed in the system).

3. Link the proposed project to one or more of the SARAO research priority areas

This project is based on the MeerKAT MIGHTEE project. The research aims to probe the HI content of the Universe on cosmological scales using MeerKAT. The student will develop knowledge on how to use MeerKAT data and even how to run the MIGHTEE calibration pipeline itself. By the end of the PhD the student will be comfortable on working with interferometer data. The student will also acquire knowledge on HI galaxy simulations and cosmology with HI intensity mapping and will be well prepared to face the challenges with the future SKA1-MID.

4. Required qualifications, academic abilities, skills and/or experience

The student is expected to have prior knowledge of Python. Classes in Cosmology and statistical analysis would be an advantage.