Section A: Overview of the Research Project

1. **Title**: Adding VLBI tools to Stimela and the CARACal Pipeline
2. **Research area**: Science
3. **Academic leve**l: Masters
4. **Abstract**:
    a. Container technology has had a big impact on radio interferometry (RI) software packages and pipelines. For example, the stimela tool (Makhathini 2017), which leverages containers to provide a system-agnostic framework for writing pipelines that can use a diverse set of RI software packers was used to produce an image with a dynamic range (DR) of 8.4 million to 1 (the highest DR in a radio synthesis image to date; Makhathini 2017). More recently, the stimela-based CARACal pipeline has been used to produce reductions that have led to at least 12 publications, including 8 from that are based on MeerKAT data. However, besides the VLBI data simulation package MeqSilhouette (Blecher *et al* 2018, Natarajan et al. 2021), the impact of container technology in VLBI data processing has been minimal. We, therefore, propose an MSc project to incorporate VLBI software tools into the stimela framework as well as the CARACal pipeline. This is particularly important in the context of reproducibility in several high-impact VLBI experiments (e.g. EHT black hole imaging, gravitational wave afterglows, binary supermassive black holes, strong gravitational lenses). Moreover, this endeavour is timely given recent developments in CASA-VLBI, enabling more modern and flexible VLBI calibration and imaging pipeline capabilities than was possible in the past.
5. **Primary supervisor**:
    a. Sphesihle Makhathini
    b. sphesihle.makhathini@wits.ac.za
    c. University of the Witwatersrand
    d. Roger Deane
    e. University of the Witwatersrand

Section B: Details of Research Project

1. **Scientific merit**:

Reproducibility is becoming an ever-challenging topic in modern science, particularly in a field like radio astronomy with complex instruments, complex software, and ever-changing environmental conditions, such as intrinsic sky variability and ionospheric/tropospheric corruptions. Nowhere is this more important to achieve than far-reaching VLBI experiments like the Event Horizon Telescope, or mapping the changing source evolution of gravitational wave electromagnetic counterparts, isolating binary supermassive black holes, or constraining dark matter substructure with VLBI lensing.

Containerized VLBI calibration and imaging pipelines have received little attention, despite being so important in the context of high-value targets that in many cases require repeatability from the simple and essential perspective of performing multi-epoch observations and the required (sub) milliarcsecond resolution. Not only that, but the CARACal pipeline enables the inclusion of several powerful calibration algorithms (cubical) and plotting libraries (ragavi, shadems) that allow insightful inspection of a complex dataset. These should prove even more useful in a VLBI application, with fewer stations and more opportunity for meaningful and impactful user inspections and editing.

In the past, easy-to-use tools for the calibration and imaging of VLBI datasets has been a limiting factor in enabling progress in the scientific applications thereof. This project aims to significantly lower the barrier to entry, as CARACal has done for MeerKAT observations. To demonstrate this, the project will include a suite of well-known, high-profile datasets to demonstrate the capabilities developed, as well as exposure the student to several cutting-edge scientific applications using different VLBI facilities.

The student will likely travel significantly in the second year of this MSc in order to interact with VLBI collaborators around the world and to further extend the capabilities of the software. This travel will be supported by Wits.

2. **Feasibility:**
    a. The main supervisor is a leading developer of both stimela and CARACal, so The student will get expert technical support. The co-supervisor is also a VLBI expert who has extensive experience with tools and techniques used in VLBI data processing.
    b. We can use the MeqSilhoette package to generate realistic simulated VLBI data that can be used to test the eventual calibration pipeline as well as the individual components. This pipeline can also be benchmarked against the fully vetted and tested CASA-VLBI pipeline (rPICARD, Janssen et al. 2019)
    c. There is also a suite of max-entropy imagers developed by the EHT Collaboration (of which the co-supervisor is an official member) that can be added in parallel in a modular fashion, and be used for side-by-side benchmarking tests for a suite of VLBI datasets, across ~3 orders of magnitude in observing frequency, with applications in astronomy and geodesy.
    d. Simulation, calibration and imaging of VLBI datasets is a computationally expensive process, however, Wits University is extremely well equipped to perform this with purpose-installed compute servers running all necessary software.

3. **Link to SARAO research priorities**:
    a. VLBI and connected-element interferometry are slowly merging en route to the SKA, therefore, we need to start the process of merging (or at least migrating to a common framework) the development of software packages and techniques. This project is a push toward that, leveraging off the

investments made and expertise developed by SARAO in interferometric calibration and imaging.

4. **Required student skills and qualifications**:

Some programming and/or data calibration experience at Honours level will be helpful but not essential.