Section A: Overview of the Research Project Proposal

1. Academic level of research project
Masters

2. Broad field of research
Astronomy

3. Title of research project
Investigating the effects of RFI flagging on HI source-recovery in LADUMA

4. Research project abstract/summary (max 250 words)
The LADUMA (Looking At the Distant Universe with the MeerKAT Array) survey will observe the neutral hydrogen (HI) gas in galaxies back to when the universe was less than a third of its current age ($z \sim 1.4$) using the MeerKAT radio telescope. However, due to satellite transmissions, parts of the radio frequency range observed with MeerKAT will be contaminated by radio frequency interference (RFI). As a result, key epochs of cosmic time could be unavailable for analysis in terms of the HI content of galaxies. In this project, the student will investigate the effect on recovered HI source properties of removing data from short array baselines, which are particularly prone to RFI, compared to results when the data from all baselines are used. This comparative analysis will be done on an RFI-clean region of the first LADUMA L-band data, with the aim of assessing the future potential for rescuing useful data from other, highly RFI-contaminated regions of the MeerKAT frequency band.

Section B: Supervisor Details

1. Primary supervisor's details
a. Title and full name
Dr Sarah Blyth

b. Name of South African or SKA Partner Country university at which the primary supervisor is a permanent academic staff member
University of Cape Town

c. Email address and/or contact telephone number
sarblyth@ast.uct.ac.za

d. Supervision of postgraduate students – please provide the details of all the previous and current postgraduate students supervised.

i. Doctoral Students

Name of student	Nationality	Date started Doctoral Degree (Month and Year)	Date completed / will complete Doctoral Degree (Month and Year)	Title of Research Project / Thesis	Co-Supervisor (if relevant)
Julia Healy	SA	Jan 2017	Dec 2020	Using HI Stacking to determine the cosmic neutral gas density at intermediate redshift	Marc Verheijen (University of Groningen)
Tshiamiso Makwela	SA	Jan 2018	Dec 2020	The impact of visualisation of 3D astronomy objects in school learning	Saalih Allie (UCT), Dale Skippers (UCT)
Gerald Balekaki	Ugandan	Jan 2016	Jun 2019	A robust online RFI database for SKA	Michelle Kuttel (UCT), Anja Shroeder (SAAO)
Narges Hatamkhani	Iranian	Aug 2016	Dec 2019	Unveiling of the newly discovered Vela Supercluster	Renee Kraan-Korteweg (UCT)
Jamie Bok	SA	Jun 2015 (upgraded from MSc)	Dec 2019	Exploring galaxy evolution with HI profile asymmetries of close pair galaxies	Tom Jarrett (UCT), Michelle Cluver (Swinburne), Rosalind Skelton (SAAO)
Tom Mutabazi	Ugandan	Jul 2012	Jul 2015	The distance to the Norma cluster and its relation to the Great Attractor	Patrick Woudt (UCT)

ii. Masters Students

Name of student	Nationality	Date started Masters Degree (Month and Year)	Date completed / will complete Masters Degree (Month and Year)	Title of Research Project / Thesis	Co-Supervisor (if relevant)
Nazir Makda	SA	Aug 2017	Dec 2018	Ultra-diffuse galaxies in Stripe 82 Clusters	Rosalind Skelton (SAAO)
Nadine Hank	SA	Feb 2018	Dec 2019	Exploring HI asymmetries in galaxy mergers with simulations	Nathan Deg (UCT)
Julia Healy	SA	Jan 2015	Jun 2017	Using HI Stacking to study galaxy properties in the nearby Universe	Ed Elson (UCT)
Christopher	SA	Jan 2012	Jun 2015	RFI Monitoring for	Michelle Kuttel

Schollar				the MeerKAT Radio Telescope	(UCT), Anja Schroeder (SAAO)
Scott Badenhorst	SA	Jan 2011	Jun 2015	Acceleration of the noise suppression component of the DUCHAMP source-finder	Michelle Kuttel (UCT)
Riona Ramraj	SA	Jun 2012	Jun 2014	Exploring star-formation in dwarf galaxies at z~1	David Gilbank (SAAO)
Zara Randria-manakoto	Madagascan	Jun 2009	Dec 2010	Super star clusters and star formation in interacting galaxies	Petri Vaisanen (SAAO)

1. Co-supervisor's details

a. Title and full name
Professor Andrew Baker

b. Name of the university/institute at which the co- supervisor is a permanent academic/research staff member
Rutgers, the State University of New Jersey

c. Email address and/or contact telephone number
ajbaker@physics.rutgers.edu

d. Supervision of postgraduate students – please provide the details of all the previous and current postgraduate students supervised.

i. Doctoral Students

Name of student	Nationality	Date started Doctoral Degree (Month and Year)	Date completed / will complete Doctoral Degree (Month and Year)	Title of Research Project / Thesis	Co-Supervisor (if relevant)
Giovanni Cresci (Firenze)	Italy	Jun 2004	2/2006	Galaxy Morphology and Star Formation with Adaptive Optics	Filippo Mannucci, Richard Davies, & Matthew Lehnert
Ross Fadely (Rutgers)	USA	Jul 2009	Jul 2010	Multi-Wavelength Applications of Gravitational Lensing	Charles Keeton
Lisa Wei (Maryland)	USA	Sep 2004	Sep 2010	A Study of Cold Gas and Star	Stuart Vogel & Sheila

				Formation in Low-Mass Blue-Sequence E/S0s	Kannappan
Paula Aguirre (Catolica)	Chile	Jun 2008	Oct 2012	Submillimeter Galaxies: Insights into Their Formation Mechanisms and the Link with Local Massive Ellipticals	Leopoldo Infante
Chelsea Sharon (Rutgers)	USA	Mar 2008	May 2013	Molecular Gas in Dusty High-Redshift Galaxies	
Robert Lindner (Rutgers)	USA	Jan 2009	Jul 2013	The Growth of Massive Galaxies and Clusters at High Redshift	
Amitpal Tagore (Rutgers)	USA	Feb 2011	Aug 2014	Uncertainties in Pixel-Based Source Reconstruction for Gravitationally Lensed Objects and Applications to Lensed Galaxies	Charles Keeton
Jesus Rivera (Rutgers)	USA	Jun 2012	May 2019	(tentative) "Detailed Studies of Lensed Dusty Star-Forming Galaxies"	
John Wu (Rutgers)	USA	Jun 2013	May 2019	(tentative) "Galaxy Evolution as Traced through Multiple Phases of the Interstellar Medium"	
Anthony Young (Rutgers)	USA	Jun 2017	May 2021	(tentative) "Applications of Gravitational Lensing at High Spectral Resolution"	

ii. Masters Students

Name of student	**Nationality**	**Date started Masters Degree (Month and Year)**	**Date completed / will complete Masters Degree (Month and Year)**	**Title of Research Project / Thesis**	**Co-Supervisor (if relevant)**

Section C: Full Research Project Proposal

1. Scientific merit: describe the objectives of the research project, placing them in the context of the current key questions and understanding of the field.

The LADUMA (Looking At the Distant Universe with the MeerKAT Array) survey will be able to observe neutral atomic gas (HI) in emission to a redshift z = 1.4, when the universe was only a third of its present age. Because of the one-to-one correspondence between frequency and redshift for the HI line, however, frequency ranges afflicted by extensive radio frequency interference (RFI) from satellites will undermine LADUMA's ability to probe the evolution of neutral atomic gas in galaxies over key redshift ranges. A possible mitigating factor is that RFI affects data on long baselines (i.e., involving widely separated antennas) much less than data on short baselines -- meaning that affected frequency ranges may be "rescued" if images are only produced from long-baseline data. The cost of such an approach is that HI sources in long-baseline-only maps are less likely to be detected with accurate parameters (size, flux, velocity width) if they are detected at all.

In this MSc project, the student will use MeerKAT observations in a frequency range with *minimal* RFI to compare the properties of HI sources detected in data cubes produced from (a) all baselines, and (b) only long baselines. The student will first determine a useful definition of "only long baselines" by considering frequency ranges where RFI is bad. The student will then produce a data cube from long-baseline MeerKAT data in the low-RFI frequency range, suitable for comparison with a standard pipeline-generated data cube produced using all baselines. The student will then apply source-finding software to both data cubes and compare the properties of the HI-emitting galaxies detected in each. The outcome of the project will inform the LADUMA team's overall strategy for imaging and analyzing data in RFI-affected frequency ranges; depending on results and student progress, it may be possible to move forward with "rescuing" such frequency ranges in the LADUMA data, and analyzing the HI properties of galaxies detected in the corresponding redshift ranges.

2. Feasibility: outline the methods that will be used to achieve the objectives. Provide details on the availability of required data / access to required equipment / availability of research facilities and other resources required. Include any relevant expected intermediate milestones and associated timeframes towards attaining the overall objectives of the project.

Data availability and analysis techniques
The project will make use of the first tranche of LADUMA L-band data from MeerKAT. Since LADUMA is due to start observing at the end of 2018, or at the very beginning of 2019 at the latest, appropriate MeerKAT data should be well in hand for this project. For training purposes, it will also be possible for the student to start applying relevant analysis techniques to existing MeerKAT commissioning data.

The student will use the CASA-based pipeline that is being developed at the Inter-University Institute for Data-Intensive Astronomy (IDIA) to process MeerKAT data, after flagging short baselines. This approach will ensure that systematic effects not related to removal of the short

baselines are kept constant, and will enable a fair comparison of the resulting data cubes produced from all baselines vs. only the long baselines.

For source-finding, the student will use the current state-of-the-art HI source-finder software, SoFiA (Serra et al., 2015). As part of the project, some investigation and optimisation of ideal parameters for the standard and long-baseline-only data cubes will be required.

Resources and equipment
As a LADUMA team member working at UCT with two of the survey PIs, the student will have access to the IDIA cloud compute resources where the LADUMA data will be stored, processed, and analysed. The student will also be able to interact with members of the LADUMA pipeline/calibration and source-finding working groups, who are experts in radio data processing and HI source detection and characterisation, respectively.

As a UCT student in the Department of Astronomy, the student will also have access to the usual desk and office space, internet access and library access afforded to all postgraduate students.

High-level breakdown of activities:
- Learn about the existing IDIA pipeline for HI data and how to produce data cubes, flag specific baselines, etc. (2019)
- Investigate and identify specific 'bad' spectral areas contaminated by RFI (2019)
- Produce LADUMA data cubes based on first L-band data with short baselines excised (2019)
- Optimise HI source-finder parameters based on 'clean' spectral ranges with full baseline complement. Some testing on LADUMA simulated data cubes may be required here as well (2019 - 2020)
- Compare source-finding results based on different underlying baseline selections to determine feasibility for RFI-heavy spectral regions (2020)
- Write up thesis (2020)

3. Link the proposed project to at least one SARAO research priority areas (refer to Annexure 1 of the Application Guide), and explain in some detail how the proposed research will contribute to the priority area(s).

The proposed project is linked to three of the SARAO research priority areas namely:
8. Interferometric Data Processing and Analysis, including calibration and imaging.
and
6. Science topics that involve the exploitation of MeerKAT data projected to be available by 2019 - 2020.
and
5. Instrumentation and systems for Radio Frequency Interference (RFI) detection and data analysis/archival/interrogation/visualization.

This project is directly relevant to exploitation of the LADUMA survey dataset and therefore applies directly to the priority area "6. Science topics that involve exploitation of MeerKAT data". It will involve testing the imaging process after removal of different baselines to investigate the effect on recoverability of HI sources, which fits into the priority area "8. Interferometric data

processing and analysis". It is also strongly linked to priority area "5. Instrumentation and systems for RFI detection and data analysis/archival/interrogation/visualisation" since the aim is to test the feasibility of extracting useful science data from RFI-afflicted parts of the spectrum.

4. If relevant, describe any particular qualifications, academic abilities, skills and/or experience that a student should have in order to successfully deliver on the objectives of the research proposed.
Good Python programming and analytical skills will be needed for this project. The student will gain hands-on experience working with interferometric imaging and HI source-finding during the project.

Section D: Signatures

1. Signature of the primary supervisor, with date

_____ 31 August 2018

Dr. Sarah Blyth Date

2. If relevant, signature of the co-supervisor/research supervisor, with date

_____ 31 August 2018

Prof. Andrew Baker Date