Mock HI spectral line stacking experiments for large galaxy surveys

Submitted in November 2019, in partial fulfillment of
the conditions for the award of the degree of **MSc Physical Science**

Modisha Tladi

Supervisor: Dr Ed Elson

Co-supervisor: Prof. Roy Maartens

Department of Physics and Astronomy
University of the Western Cape

UNIVERSITY *of the*
WESTERN CAPE

https://etd.uwc.ac.za

Plagiarism declaration

I, Modisha Tladi, know the meaning of plagiarism and declare that all the work in the document, save for that which is properly acknowledged, is entirely mine.

https://etd.uwc.ac.za

Abstract

The study of neutral atomic hydrogen (HI) plays a vital role in improving the knowledge about galaxy evolution since HI serves as the raw fuel for star formation. However, the current knowledge about it is limited to the nearby Universe due to its intrinsic faintness and the inadequate sensitivities of the current radio telescopes. This work focuses on applying the HI spectral line stacking technique to a suite of simulated data sets in order to measure the average HI content of galaxies at high redshifts. The method aims to obtain a high signal-to-noise (S/N) ratio detection of the total HI content of a galaxy sample by co-adding individual galaxy spectra.

The main aim in this work is to quantify the uncertainties in the results from HI stacking experiments due to the inclusion of additional flux from source confusion and point spread function (PSF) contamination, respectively. HI data cubes that contain an extensive amount of realism are produced, based on the various VLA configurations, using simulation methods in CASA (Common Astronomy Software Applications). Furthermore, a set of cubes that are smoothed with the Gaussian beams equal in size to the central lobes of the PSFs from the CASA-generated cubes are produced. The Gaussian-smoothed cubes are representative of the perfect case such that they only have source confusion effects and no component due to PSF contamination.

Stacking methods are then written in Python and are applied on all the newly-generated data cubes to study the aforementioned sources of additional flux. Comparing the stacked results for the smoothed and the CASA-generated cubes allows for the proper quantification of the additional flux contributions that are solely due to PSF contamination.
All of the co-added spectra significantly mis-estimate the average HI galaxy mass. Typically, the measured average galaxy mass is $8.99 \times 10^9\ M_\odot$ whereas the evaluated (true) average mass is $2.88 \times 10^9\ M_\odot$. From this, the total amount of mass due to source confusion is $4.19 \times 10^9\ M_\odot$ while the amount of mass due to PSF contamination is $1.92 \times 10^9\ M_\odot$. Results such as these indicate that upcoming surveys will suffer from these additional flux effects and this further shows that detailed modelling will be required to properly quantify the uncertainties. This will allow for the extraction of the most accurate cosmological parameters from these experiments.

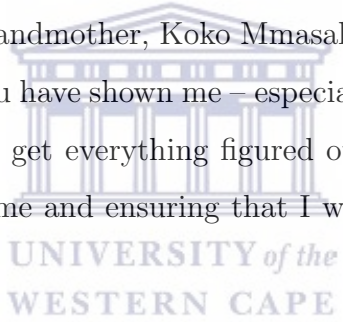
UNIVERSITY of the
WESTERN CAPE

https://etd.uwc.ac.za

Acknowledgements

I would first like to extend my sincere gratitude to my supervisor, Dr Ed Elson, for guiding me throughout this thesis. This work would have not been possible without the constant support and guidance that I received from him. Ed, I am grateful for having to work under your supervision and I appreciate your patience throughout this work – especially during the times when I was still working on gaining a clear understanding on this project. I would also like to thank my co-supervisor, Prof. Roy Maartens, for his support – more especially with the help that he offered when I was still doing my master's coursework.

To my family, thanks for the support that you have shown me throughout. Mom and Dad, I really appreciate your moral support and prayers throughout my years of study. Thanks for always ensuring that I had everything that I needed during my tough undergraduate years. Rose and Gladwin, visiting home has always been worthwhile because of you. Listening to music and playing video games with you has always been therapeutic and it kept me going. To my grandmother, Koko Mmasalemone, I will always be grateful for the love and support that you have shown me – especially during my first year of study in varsity when I was trying to get everything figured out. To my uncle, Moses, thank you for always checking up on me and ensuring that I was well-taken care of during my undergraduate years.



Mbalenhle, I get lost for words to express my gratitude for your constant support throughout my master's research. Your words of encouragement kept me going and helped me stay motivated through the difficult times that I encountered while doing this research. Thank you so much for always being there. Thabiso "Warriks", thanks for being a good friend whom I knew I could always count on. Your sense of humour and the good music that we shared really kept me going. Thato "Nams", thanks for being a good friend and for always being supportive when we had to get through our MSc coursework.

I would also like to thank the Centre for Radio Cosmology (CRC) for funding my master's research. I would have not been able to undertake this work without the financial support that I received from the CRC.

https://etd.uwc.ac.za

Contents

Abstract i

Acknowledgements ii

1 Introduction 1

1.1 Research background 1

1.2 Redshift evolution of the cosmic HI density and star formation rate 2

1.3 HI stacking methods 6

1.4 Uncertainties in the HI stacking experiments 7

1.5 Blind HI surveys 8

1.6 Upcoming deep HI surveys 9

1.7 Aims 10

2 Synthetic data products 13

2.1 Existing data products 13

2.1.1 Semi-analytic models 14

2.1.2 Galaxy models 21

2.2 New data products produced in this work 25

2.2.1 Simobserve task 26

2.2.2 CLEAN task 29

3 HI stacking methods 31

3.1 The stacking method 31

3.2 Verification of the stacking method 36

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

4 Mock stacking experiments **40**

4.1 Data cube representations 40

4.2 Descriptions of the stacks obtained in this project 42

5 Discussions **52**

5.1 Additional flux in HI co-added spectra 52

5.2 Total co-added mass 54

5.3 Additional flux due to source confusion 57

5.4 Additional flux due to PSF contamination 58

6 Summary and future outlooks **63**

6.1 Summary 63

6.2 Future outlooks 65

6.2.1 MeerKAT surveys 66

6.2.2 ASKAP surveys 67

6.2.3 Overview 68

Bibliography **68**

UNIVERSITY of the
WESTERN CAPE

https://etd.uwc.ac.za

List of Tables

2.1 Description of the columns of the mock catalogue in Obreschkow and Meyer (2014) 20
2.2 Baseline properties and the major beam axis (BMAJ) corresponding to the various VLA configurations 26
2.3 Input parameters used in the 'Simobserve' task 27

3.1 Comparisons of the co-added HI masses obtained in this work and those from Elson et al. (2019) 38

4.1 Beam parameters of the CASA-generated data cubes 42

5.1 VLA-B CASA cube results 61
5.2 VLA-C CASA cube results 61
5.3 VLA-D CASA cube results 61

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

List of Figures

1.1 Evolution of the cosmic HI gas density $\Omega_{\rm HI}$ with redshift 3

1.2 Cosmic star formation history measurements 5

1.3 Illustration of an HI data cube 7

2.1 Column density profiles for two spiral galaxies (NGC 5055 and NGC 3184) 16

2.2 Galaxy mass functions for HI and H_2 17

2.3 Tully Fischer relations (TFRs) from Obreschkow et al. (2009a) 18

2.4 The HI mass-size relation 19

2.5 An example 3-D model for each galaxy in the Elson et al. (2016) simulation 23

2.6 Evaluated galaxy properties for the Elson et al. (2016) cube 24

2.7 Point spread functions (PSFs) for the various VLA configurations 25

2.8 Outputs generated from the 'Simobserve' task 29

3.1 Individual HI spectra for an isolated galaxy from the Elson et al. (2019) data cube 34

3.2 Individual HI spectra for a crowded galaxy from the Elson et al. (2019) data cube 35

3.3 Contamination as a function of spatial aperture size for an isolated and a crowded galaxy from the Elson et al. (2019) data cube 36

3.4 Co-added HI spectra obtained for the full-sized Elson et al. (2019) data cube 37

4.1 Channel maps for the Elson et al. (2016) data cube 41

4.2 Channel maps for the VLA-B CASA cube 43

4.3 Channel maps for the VLA-C CASA cube 44

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

4.4 Channel maps for the VLA-D CASA cube 45
4.5 P-V diagrams for the Elson et al. (2016) data cube 46
4.6 P-V diagrams for the VLA-B CASA data cube 47
4.7 P-V diagrams for the VLA-C CASA data cube 48
4.8 P-V diagrams for the VLA-D CASA data cube 49
4.9 Total intensity maps for some of the massive galaxies in the Elson et al. (2016) cube 50
4.10 Combined co-added HI spectra for the CASA-generated cubes based on the various VLA configurations 51

5.1 Individual co-added HI spectra for the CASA-generated cubes based on the various VLA configurations 55
5.2 Fractional mass as a function of spatial aperture size for the Gaussian-smoothed and the CASA cubes 56
5.3 Co-added HI spectra for the Gaussian-smoothed cubes based on the beam parameters of the various VLA CASA cubes 58
5.4 Illustration of a 2-D array convolved with a Gaussian beam and a realistic PSF 60

UNIVERSITY of the
WESTERN CAPE

https://etd.uwc.ac.za

Chapter 1

Introduction

1.1 Research background

Quantification of the neutral atomic hydrogen (HI) gas in galaxies – as a function of redshift – is crucial as it enhances the knowledge about galaxy formation and evolution. This is mainly because HI is supplied through various merger, feedback and accretion processes where it then condenses into molecular clouds and stars – thus making it the raw fuel for star formation (Delhaize et al., 2013; Maddox et al., 2013; Geréb et al., 2015; Rhee et al., 2016). The current knowledge of the gas content around galaxies in the nearby Universe has increased amply, thanks to the HI surveys such as HIPASS (HI Parkes All-Sky Survey, Barnes et al., 2001), ALFALFA (Arecibo Legacy Fast ALFA, Giovanelli et al., 2005) and THINGS (The HI Nearby Galaxy Survey, Walter et al., 2008).

UNIVERSITY of the
WESTERN CAPE

To explore the star formation properties of different galaxy populations in relation to their total HI gas content, large samples are needed. This can be achieved by using large sky surveys such as ALFALFA (Giovanelli et al., 2005) or by extending observations out to higher redshifts (Geréb et al., 2015). However, when conducting experiments beyond the local Universe, direct HI detection methods are not reliable since the intrinsic faintness of the HI signal falls off with distance. The inadequate sensitivities of the current radio telescope also make current direct detection methods unreliable. Another major drawback with direct detection methods on high-redshift experiments lies in the fact that

https://etd.uwc.ac.za

they require prohibitively long observation times even with the most sensitive telescopes currently available. Such limitations are an indication that alternative methods are needed for studying and understanding the HI properties in the distant Universe. For this reason, a method called HI spectral line stacking is used to acquire a better understanding of the HI gas around galaxies at high redshifts. Further details on this method will be explored in Section 1.3.

1.2 Redshift evolution of the cosmic HI density and star formation rate

A complete picture of the evolution of galaxies can be obtained if sufficient information about the variation of cold gas with cosmic time is provided. Particularly, improving an understanding of things such as the star formation rate density first requires a clear understanding of the changes in the cosmic HI gas density ($\Omega_{\rm HI}$). The evolution of $\Omega_{\rm HI}$ as a function of redshift (z) has been explored in various studies and the measurements obtained are shown in Figure 1.1. Clearly visible in Figure 1.1 is the good agreement in the measurements made at low redshifts ($z < 0.5$). However, large uncertainties start arising at intermediate redshifts and beyond. The results obtained by studies that apply the HI spectral line stacking technique (in Figure 1.1) indicate the usefulness of the stacking method, which assists in bridging the redshift gap between estimates done using direct detections and the measurements from damped Lyman-alpha (Lyα) systems (concentrations of neutral gas found on the spectra of quasars).

UNIVERSITY of the
WESTERN CAPE

As indicated in Zwaan et al. (2005), accurate measurements of the HI mass function (HIMF) can be used to measure the total mass density of HI atoms at $z = 0$. Zwaan et al. (2005) made use of the HIPASS to measure the HIMF of galaxies and ultimately used the measurements to determine the mass density of neutral hydrogen at $z = 0$. Given the large sky coverage of the HIPASS, the number of detections made is very large, consisting of 5317 galaxies identified purely by their HI content (e.g., Wong et al., 2006), and thus

https://etd.uwc.ac.za

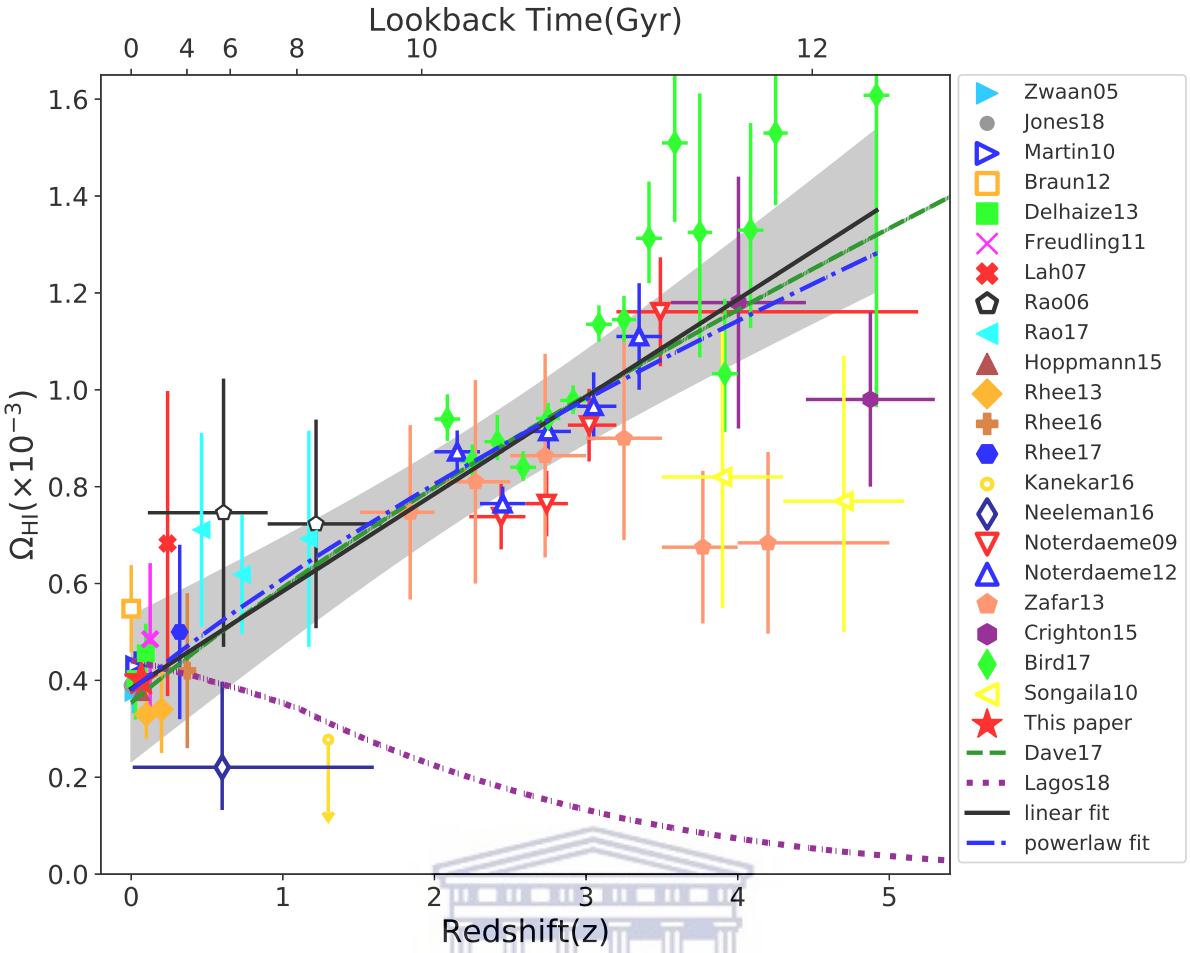


Figure 1.1: A plot of the evolution of the cosmic HI gas density with redshift obtained from different measurement methods. The measurements shown on the right are from different sources and they are as follows (from top to bottom): 21-cm emission measurements from HIPASS (Zwaan et al., 2005); 21-cm emission measurements from α40 ALFALFA (Jones et al., 2018); Parkes HI stacking measurements (Delhaize et al., 2013); Arecibo Ultra Deep Survey (AUDS) (Hoppmann et al., 2015); WSRT HI stacking (Rhee et al., 2013); 21-cm emission stacking from the Giant Metrewave Radio Telescope (GMRT) (Lah et al., 2007); Hubble Space Telescope (HST) and SDSS damped Lyα measurements (Rao et al., 2006, 2017); self-opaque effect corrected detections of damped Lyαs with the Green Bank Telescope (GBT) and the Westerbork Synthesis Radio Telescope (WSRT) (Braun, 2012); European Southern Observatory (ESO) Ultraviolet and Visual Echelle Spectrograph (UVES) measurements of damped Lyαs (Zafar et al., 2013); Gemini Multi-Object Spectrograph (GMOS) measurements of damped Lyα (Crighton et al., 2015); Galaxy Evolution Explorer (GALEX) and Keck measurements of damped Lyαs (Songaila and Cowie, 2010); MUFASA cosmological hydrodynamical simulations (Davé et al., 2017); SHARK semi-analytic galaxy formation model (Lagos et al., 2018). Results from Hu et al., 2019 are shown by the red star. (Hu et al., 2019, Figure 14).

allows for the measurement of the HIMF with high accuracy. Given the good statistics provided by the HIPASS at the high-mass end of the HIMF, the measurements of $\Omega_{\rm HI}$ by Zwaan et al. (2005) were made with high precision since the high mass-end dominates the

https://etd.uwc.ac.za

mass density of neutral hydrogen.

Several observational techniques exist that focus on tracing the HI gas density at high redshifts and one particular example is the Gunn-Peterson effect applied to damped Lyα systems. The Gunn-Peterson effect is a feature of the spectra of quasars which is caused by the presence of neutral hydrogen in the Intergalactic Medium. Lah et al. (2007) indicate that measurements of the damped Lyα absorption in quasi-stellar objects (QSO) spectra at high redshifts out to $z > 1.5$ have measured the density of HI very effectively (e.g., Storrie-Lombardi and Wolfe, 2000; Prochaska et al., 2005) whereas it is difficult to constrain the $\Omega_{\rm HI}$ at intermediate redshifts ($0.5 < z < 1.0$). The latter effect occurs since the average number of damped Lyα systems are low per unit redshift, which then leads to shortcomings such as the need for impractical observation times when attempting to make estimates of $\Omega_{\rm HI}$ at intermediate redshifts. Due to the difficulty in detecting the Lyα line with ground-based telescopes, this ultimately leads to the requirement of using HI stacking techniques to study the $\Omega_{\rm HI}$ at intermediate redshifts.

Furthermore, constraining the $\Omega_{\rm HI}$ at $z > 0$ using radio observations of HI-21 cm emission is difficult since most of the flux from galaxies at these redshifts falls below the detection limits of the current radio telescopes. This result further indicates the necessity to resort to alternative methods such as stacking to further gain an understanding of the $\Omega_{\rm HI}$ beyond the local Universe.

UNIVERSITY of the
WESTERN CAPE

Madau and Dickinson (2014) studied the cosmic star-formation history (SFH) from ultraviolet (UV) and infrared (IR) data, shown in Figure 1.2. They presented their measurements in both redshift and lookback time (Gyr). Their results show a steady rising phase at redshifts $3 \leq z \leq 8$ and peaks between $z = 2$ and $z = 1.5$, which is then followed by a steady decline up to $z = 0$ (present day).

Understanding the link between the redshift evolution of $\Omega_{\rm HI}$ and SFR is crucial as it helps

https://etd.uwc.ac.za

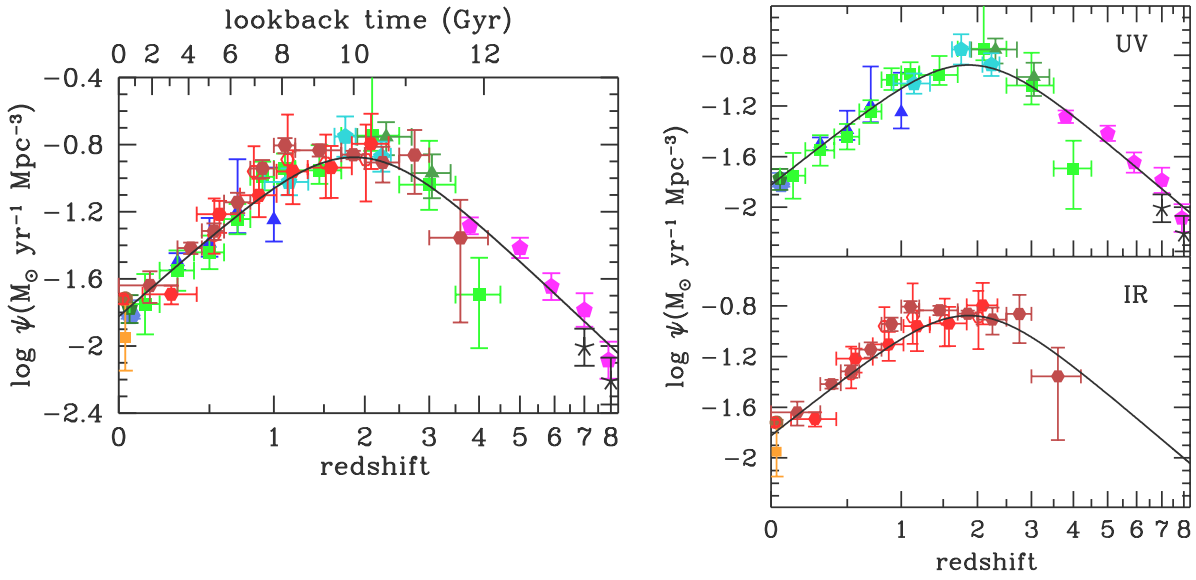


Figure 1.2: Cosmic star formation history from the rest-frame measurements of FUV (top right), IR (bottom right) and a combination of FUV and IR (left). Measurements were made from various studies as follows: blue-gray hexagon from Wyder et al. (2005); blue triangles from Schiminovich et al. (2005); dark green pentagon from Robotham and Driver (2011); green squares from Cucciati et al. (2012); turquoise pentagons from Dahlen et al. (2007); dark green triangles from Reddy and Steidel (2009); magenta pentagons from Bouwens et al. (2012a,b); black crosses from Schenker et al. (2013); brown circle from Sanders et al. (2003); dark orange square from Takeuchi et al. (2003); red open hexagons from Magnelli et al. (2011); red filled hexagons from Magnelli et al. (2013); dark red filled hexagons from Gruppioni et al. (2013). (Madau and Dickinson, 2014, Figure 9).

UNIVERSITY of the
WESTERN CAPE

in understanding galaxy evolution. To explore this, this work reports on the investigations of Hopkins et al. (2008), who studied the link between the redshift evolution of $\Omega_{\rm HI}$ and SFR in detail. To maintain the star formation rates in galaxies, they emphasize the need for HI to be continually replenished from a reservoir of ionized gas. To explain the reason for the little evolution (from $z \approx 5$ to $z \approx 0.2$ in Figure 1.1) of the HI mass density in the Universe for so much of cosmic history, they proposed that the replenishment is steered by the supershells linked with star-forming complexes within galaxies. In the supershell walls, precursory ionized gas cools and recombines. Molecular gas forms in these walls and compression of the molecular clouds by these walls occurs – which then triggers collapse and star formation (Hopkins et al., 2008).

https://etd.uwc.ac.za

1.3 HI stacking methods

Due to the limitations of the direct detection methods discussed in section 1.1, this work focuses on applying the HI spectral line stacking method to study the HI properties of galaxies at high redshifts. Stacking is defined as the method of co-adding individual spectra with no HI detections with the aim of boosting the signal-to-noise ratio of the data – thus recovering a detection that is more significant (Delhaize et al., 2013). This method has been proven to be reliable and efficient in a variety of studies (e.g., Lah et al., 2007; Delhaize et al., 2013; Rhee et al., 2013, 2016, 2017; Geréb et al., 2015; Elson et al., 2016, 2019). The method uses known redshifts and sky positions of a sample of galaxies – which are usually obtained from optical imaging and spectroscopy – to extract individual galaxy spectra from an HI data cube (Maddox et al., 2013). Spectra are extracted from these known locations, shifted to the rest frame of each galaxy and then finally get co-added to yield an HI spectrum representative of the whole sample of galaxies in the data cube. Shown in Figure 1.3 is a demonstration of the extraction of sub-cubes within a data cube. These sub-cubes are then stacked to yield the total HI content representative of the entire sample of galaxies.

The main goal in HI stacking experiments is to co-add a large number of non-detections (an example of which is shown as the spectrum corresponding to the red sub-volume in Figure 1.3) to ultimately build up a single total detection and recover the statistical properties of the galaxies in the data cube.

Gaining an understanding of the HI properties of galaxies in the distant Universe will help answer a lot of key science questions such as the processes involved in galaxy formation and evolution, which will be better understood using the upcoming deep HI surveys. Since these surveys are not yet operational, theoretical models from the HI stacking experiments are some of the most reliable methods used to study the HI content of galaxies at high redshifts. Despite the reliability of these methods, a major drawback in them is that they always overestimate the true HI mass expected for a sample of galaxies. The uncertainties that come with this shortcoming will be explored further in Section 1.4.

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

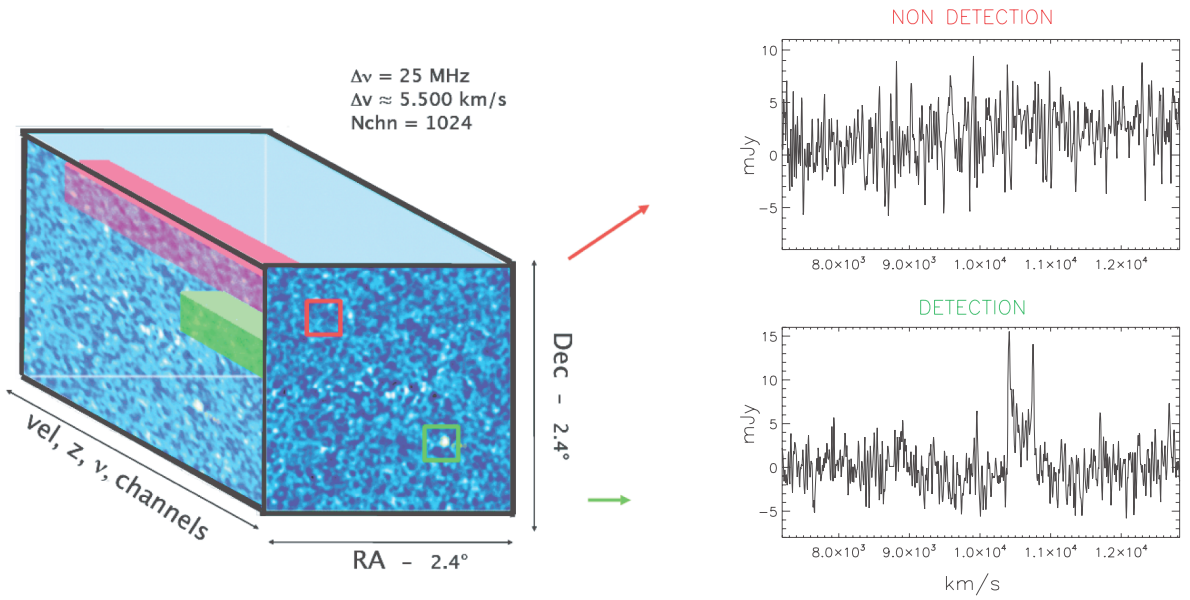


Figure 1.3: Illustration of the extraction of sub-cubes (indicated by the green and red sub-volumes) from a large data cube. On the right are the HI spectra for these sub-volumes, where one shows a detection (having a prominent double horn global profile) and the other shows a non-detection (noisy spectrum). The size of each sub-cube extracted is controlled by the choice of spatial aperture size (in right ascension and declination) and the spectral range (number of channels) specified. (Fabello et al., 2011, Figure 4).

1.4 Uncertainties in the HI stacking experiments

HI stacking methods are an effective tool for studying the HI content of galaxies at high redshifts and they have proven to be efficient as they have produced consistent results in various studies. Until the next generation of telescopes that will push out to higher redshifts are available, HI stacking methods will remain highly efficient in recovering the total HI content of galaxies beyond the local Universe. Despite the effectiveness of these methods, a major setback about them is that they always yield an overestimate of the true HI mass contained within a sample of galaxies in a data cube. A large fraction of this overestimate is due to source confusion, which is a result that occurs when the extracted spectrum of a target galaxy is contaminated by flux from nearby neighbouring galaxies due to overcrowding in a data cube. Delhaize et al. (2013) investigated the contributions of source confusion in HI stacking experiments for Parkes observations at a redshift range $0.04 < z < 0.13$ and they found that their stacked sample had a total amount of confusion of $\sim 1.1\times\ 10^{10}\ h_{70}^{-2}\ M_{\odot}$ per stacked galaxy. Jones et al. (2015) carried out the same

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

experiment as that of Delhaize et al. (2013) but took a different approach. Their method made use of the currently available HI correlation function (CF) of the ALFALFA and the measurements of the mean HI density to measure the amount of additional mass in a co-added spectrum, due to source confusion. Their results predicted the amount of confused mass to be $\sim 1.3\times 10^{10}\ h_{70}^{-2}\ M_{\odot}$ per stacked galaxy. Furthermore, Elson et al. (2016) created a set of synthetic HI data cubes and carried out HI stacking experiments on them to quantify the contributions of additional mass from source confusion. When considering the redshift range $0.04 < z < 0.13$, they found the total amount of confused mass in a stacked spectrum to be $\sim 1.5\times 10^{10}\ h_{70}^{-2}\ M_{\odot}$ per stacked galaxy, which is similar to that of Jones et al. (2015). The efforts carried out in the aforementioned studies have contributed significantly towards improving the current understanding of the effects of additional flux in HI stacking experiments due to source confusion. However, these studies did not consider the contributions of additional flux due to point spread function (PSF) effects. This work considers the contributions of additional flux in HI stacking experiments due to source confusion and, for the first time, the contributions of additional flux due to PSF effects. For HI interferometric observations, the characteristics of the PSF are determined by the Fourier transform of the baseline distribution. In practice, this implies that the PSF can be very large PSFs with spatially-extended side lobes. The latter occurs since an incomplete UV coverage leads to PSFs with spatially-extended features such as lobes and spikes. When the flux distribution on the sky is convolved with such a PSF, the flux from a point source gets distributed to other sources much further away from it. The main focus of this study is on exploring the effects of PSFs in the context of the various configurations of the Very Large Array (VLA). Each of the configurations has a unique baseline distribution, which determines the data products generated and thus allows for the quantification of PSF effects for various cases.

UNIVERSITY of the WESTERN CAPE

1.5 Blind HI surveys

Radio telescopes with fast survey speeds allow for conducting blind HI surveys. One significant survey was the HIPASS, which was conducted using the Parkes 64 m radio

https://etd.uwc.ac.za

telescope located in New South Wales, Australia. A survey such as this was sensitive to low-surface-brightness galaxies which are missing from existing catalogues because of the limitations of optical data due to high optical extinction. One of the key objectives of the HIPASS was to conduct an extragalactic HI survey of the whole southern sky which contains crucial information such as the distribution of galaxies, the HI mass function (HIMF) and the dynamics of groups and superclusters. The stacking experiment of Delhaize et al. (2013) was based on the HIPASS data.

One other significant survey was the ALFALFA, which mapped nearly 7000 $\deg^2$ of the entire sky visible from the Arecibo observatory. One of the core aims of ALFALFA was to study the faint-end of the HIMF in the local Universe. When compared to HIPASS, ALFALFA is about 8 times more sensitive with about 4 times better angular resolution (Giovanelli et al., 2005). As blind surveys, both HIPASS and ALFALFA were not biased to high-surface-brightness galaxies as it is usually the case in optical galaxy catalogues. Fabello et al. (2011) made use of early ($\sim$ 40 %) ALFALFA data for their HI stacking experiments.

UNIVERSITY of the WESTERN CAPE

1.6 Upcoming deep HI surveys

Recent single-dish blind HI surveys have been of great significance as they have shed light in providing information on the global HI properties and the relationship between star formation and the interstellar medium. The main drawback with these surveys is that they have been limited to the nearby Universe. Upcoming deep HI surveys are underway and they will tackle this problem by extending current measurements to higher redshifts, beyond the local Universe. These surveys will allow for probing the HI content of galaxies beyond $z = 0$, which will be done indirectly using stacking. This work focuses on exploring some of the details of the three deepest upcoming surveys namely the LADUMA (Looking At the Distant Universe with MeerKAT Array, Baker et al., 2018), DINGO UDEEP (Deep Investigation of Neutral Gas Origins - Ultra Deep, Jones et al., 2015) and CHILES (COSMOS HI Large Extragalactic Survey, Fernández et al., 2013). For the LADUMA

https://etd.uwc.ac.za

survey, the HI spectral line stacking technique will be one of the primary analysis methods.

LADUMA is a single pointing survey with an integration time of 3424 hours and the synthesized beam is $\sim$ 10 arcsec across – with a total field of view of 0.9 deg^2 at $z = 0$. Its main aim is to detect HI out to redshift $z \sim 1$ and beyond, which will ultimately help answer some of the key questions in science such as the dynamics behind the formation and evolution of galaxies. Within $z \sim 0.5$, galaxies will be directly detected (but not all of them). Beyond $z \sim 0.5$, most, if not all, galaxies will not be directly detected.

CHILES is also a single pointing survey with an integration time of 1000 hours. The minimum synthesized beam from the survey is $\sim$ 5 arcsec across and its field of view is 0.25 deg^2 at $z = 0$. CHILES adopts a narrower bandwidth compared to LADUMA, which basically sets its maximum redshift detection for HI at redshift $z \sim 0.45$. The highest HI 21 cm direct detection of a galaxy to date was made by Fernández et al. (2016) at redshift $z = 0.376$, using the first 178 hr of the CHILES data. This detection was of the luminous galaxy COSMOS J100054.83+023126.2 and the total HI mass that they obtained for this galaxy is $(2.9 \pm 1.0) \times 10^{10}\ M_{\odot}$.

UNIVERSITY of the
WESTERN CAPE

Unlike LADUMA and CHILES, DINGO UDEEP is not a single pointing survey. The survey will make use of the Australian SKA Pathfinder (ASKAP), which covers a field of view of 60 deg^2 between the redshift range $z \sim 0.1$ to $z \sim 0.43$. The integration time that it will have is 5000 hours and the survey may reach a resolution of 10 or 30 arcsec, which has not yet been ascertained since it is computationally expensive to form multiple beams from the phased array feeds of the ASKAP.

1.7 Aims

The main aims of this project are as follows:

- To gain access to various simulated data products that can be used to extract

https://etd.uwc.ac.za

galaxy properties and allow for the study of neutral atomic hydrogen (HI) stacking experiments.

- To explore the uncertainties in the results from HI stacking experiments due to the inclusion of additional flux from source confusion and PSF contamination components, respectively.

- To produce realistic data cubes containing PSF effects using simulation methods on the Common Astronomy Software Application (CASA) software. This will be done by passing the Elson et al. (2016) cube through CASA to incorporate the effects of realistic PSFs and interferometric noise. This will yield variations of the Elson et al. (2016) cube which will be based on the various VLA configurations. Each of the configurations that are considered consists of a unique baseline distribution that determines the angular resolution of the data products generated, as well as the relative prominent PSF side lobes.

- To write stacking codes using Python and to verify their correctness by applying them on the Elson et al. (2019) cubes and making comparisons of the results from this work with their results. Upon confirming the correctness of the stacking methods, the next step will be to apply them on the realistic CASA-generated cubes to quantify the amount of additional flux due to source confusion and PSF contamination.

- To produce cubes that are smoothed with the Gaussian beams equal in size to the central lobes of the PSFs from the CASA-generated cubes. These cubes will be free from the effects of PSFs. Comparing the stacked results for the smoothed and the CASA-generated cubes will allow for the proper quantification of the additional flux contributions that are solely due to PSFs.

- To study the variation of additional flux in a stacked spectrum as a function of changing spatial aperture size for both the perfect and the realistic cubes.

UNIVERSITY of the
WESTERN CAPE

https://etd.uwc.ac.za

The structure of this thesis is as follows: Chapter 2 outlines the existing and new synthetic data products used in this work. Chapter 3 outlines the HI stacking methods implemented together with the verification of those methods. Chapter 4 outlines the various representations of the data cubes used together with the stacking results obtained from those cubes. Chapter 5 provides a discussion of the results obtained in this work – based on the methods of additional flux quantification due to source confusion and PSF contamination. Finally, Chapter 6 provides a summary of this work together with the future outlooks on HI stacking experiments.

https://etd.uwc.ac.za

Chapter 2

Synthetic data products

This chapter describes the details of the simulated data products used in this thesis. The final data cubes produced in this work are variations of the Elson et al. (2016, hereafter E16) cube, which were passed through CASA (Common Astronomy Software Applications) to introduce more realism. The E16 cube is based on the semi-analytic models of Obreschkow and Meyer (2014, hereafter O&M14).

The various simulation methods that led to the final products used in this work are described in Section 2.1. First, an explanation of the semi-analytic models from Obreschkow et al. (2009a,b,c) – which led to the mock galaxy catalogue produced in the O&M14 study – is given. This is then followed by the presentation of the methods applied by E16 to the O&M14 catalogue to produce synthetic HI data cubes. Finally, explanations of the methods that were applied in this study – using CASA – to produce several versions of the E16 cube with more realism are given.

UNIVERSITY of the
WESTERN CAPE

2.1 Existing data products

This section is devoted to describing the existing products that were converted into the final data products used in this work.

https://etd.uwc.ac.za

2.1.1 Semi-analytic models

Obreschkow et al. (2009a,b,c) present the evolution of the molecular (H_2) and atomic (HI) phases of cold hydrogen gas in 3×10^7 galaxies, based on the Millennium simulation. Their methods mainly post-process the virtual galaxy catalogue of De Lucia and Blaizot (2007), where they set physical prescriptions that assign HI and H_2 to galaxies. They approximate the surface density of the total hydrogen component (HI + H_2) by an axially symmetric exponential profile presented by Leroy et al. (2008):

$$\sum\nolimits_{\rm H}(r) = \sum\nolimits_{\rm H} \exp\left(\frac{-r}{r_{\rm disc}}\right), \tag{2.1}$$

where $r_{\rm disc}$ is the exponential scale length of the disc. The normalization factor, $\sum_{\rm H}$, can be calculated using $\sum_{\rm H} \equiv M_{\rm H}/(2\pi r_{\rm disc}^2)$, where $M_{\rm H}$ represents the total mass of cold hydrogen in the disc.

UNIVERSITY of the
WESTERN CAPE

They then derived analytic expressions for the surface density profiles of HI and H_2, which were expressed as:

$$\sum\nolimits_{\rm HI}(r) = \frac{\sum_{\rm H} \exp\left(\frac{-r}{r_{\rm disc}}\right)}{1 + R^{\rm c}_{\rm mol} \exp(-1.6r/r_{\rm disc})}, \tag{2.2}$$

$$\sum\nolimits_{\rm H_2}(r) = \frac{\sum_{\rm H} R^{\rm c}_{\rm mol} \exp(-2.6r/r_{\rm disc})}{1 + R^{\rm c}_{\rm mol} \exp(-1.6r/r_{\rm disc})}, \tag{2.3}$$

where $R^{\rm c}_{\rm mol}$ is the H_2/HI mass ratio at the galaxy centre. Their physical model is based on the midplane pressure, which plays a vital role in these simulations as it determines the HI and H_2 radial profiles of galaxies. The midplane pressure represents the pressure found in the diffuse part of the interstellar medium (ISM) which is closest to star-forming regions (e.g., Koyama and Ostriker, 2009). It has been included in semi-analytic simulations as it predicts the global evolution of the H_2 content in the Universe and also estimates the star formation rates of galaxies (e.g., Obreschkow and Rawlings, 2009). Using Equations 2.2 and 2.3, (Obreschkow et al., 2009a, hereafter O2009a) characterized the HI- and H_2-

https://etd.uwc.ac.za

content of the simulated galaxies found in the catalogue of De Lucia and Blaizot (2007).

To confirm that their derived analytical expressions (Equations 2.2 and 2.3) for assigning HI and H_2 to galaxies work well, O2009a selected two observed spiral galaxies (NGC 3184 and NGC 5505) to produce HI and H_2-density profiles that could be compared to the results obtained from other studies. They tested the limitations set by their assumptions that $r_{\rm disc} = 2\tilde{r}_{\rm disc}$, where $\tilde{r}_{\rm disc}$ is the scale length of the stellar disc that is $30 - 50\%$ smaller than $r_{\rm disc}$. The galaxies that they consider have $r_{\rm disc} \approx \tilde{r}_{\rm disc}$ (NGC 5505) and $r_{\rm disc} \approx 2\tilde{r}_{\rm disc}$ (NGC 3184). To evaluate their equations, they require the quantities $M_{\rm stars}^{\rm disc}$ (disc stellar mass), $M_{\rm gas}$ (cold gas mass), $r_{\rm disc}$ (the scale radius of the cold gas disc) and $\langle f_\sigma \rangle$ (the dispersion parameter). They determined $M_{\rm stars}^{\rm disc}$, $M_{\rm gas}$ and $r_{\rm disc}$ by fitting a single exponential profile to the stellar component. For $\langle f_\sigma \rangle$, they used $\langle f_\sigma \rangle = 0.4$, which is the value provided in the work of Elmegreen (1993), chosen for nearby galaxies. They produced model profiles and made comparisons with results from the work of Leroy et al. (2008), who also considered NGC 5505 and NGC 3184 when producing column density profiles. Shown in Figure 2.1 are the results that they obtained.
The results of O2009a in Figure 2.1 show the accuracy of their predictions since their model profiles, given by Equations 2.2 and 2.3, match empirical data.

UNIVERSITY of the
WESTERN CAPE

A huge success in these resulting hydrogen simulations is that they have reproduced, with high accuracy, several empirical relations known for the local Universe. One other significant result that was obtained by O2009a was on the Mass Functions (MFs) of HI and H_2 that they produced. MFs determine the number density of galaxies in the Universe as a function of their HI and H_2 masses. The shapes of MFs give information on how HI and H_2 are distributed over galaxies of different masses in the Universe and, as a result, they are important probes of galaxy evolution. Shown in Figure 2.2 are the Mass Functions (MFs) for HI and H_2 from O2009a compared with the results from other work.
Simulations that can produce mass functions for the local Universe can be regarded as being reliable and able to predict real galaxy properties. Given the fact that the O2009a

https://etd.uwc.ac.za

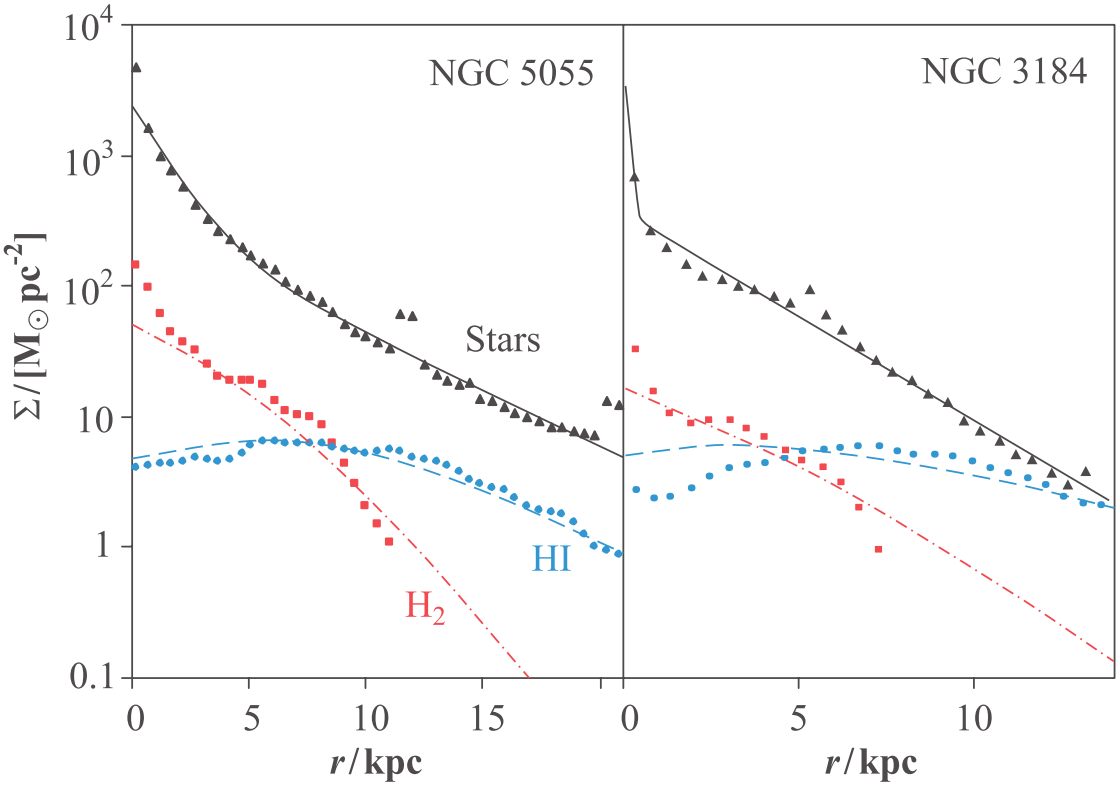


Figure 2.1: Plots showing the column density profiles for two spiral galaxies (NGC 5055 and NGC 3184). The dashed and dashed-dotted lines represent the predictions of the model of O2009a, which is based on Equations 2.2 and 2.3. Circles, squares and triangles represent the column density profiles for HI, H_2 and stars, respectively, from Leroy et al. (2008). Solid black lines show the best-fitting functions for stellar densities. (O2009a, Figure 1).

UNIVERSITY of the
WESTERN CAPE

simulations can produce MFs that are comparable to the empirical MFs from Zwaan et al. (2005) and Obreschkow and Rawlings (2009), they can safely be regarded as being reliable.

Part of the significant work done by O2009a was on the Tully-Fisher relations (TFRs) that they obtained. The TFR is an empirical scaling relation between the mass or intrinsic luminosity and the rotational velocity of a spiral galaxy. One of the main important uses of the TFR is to estimate the distance to spiral galaxies, which can be found by comparing the luminosity to the apparent brightness. O2009a computed TFRs at a range of redshifts for a sample of 1000 galaxies that were randomly drawn from their simulation. They showed this at redshift z = 0. They produced the TFRs based on the masses of galaxies versus the simulated line widths. Shown in Figure 2.3 are the results that they

https://etd.uwc.ac.za

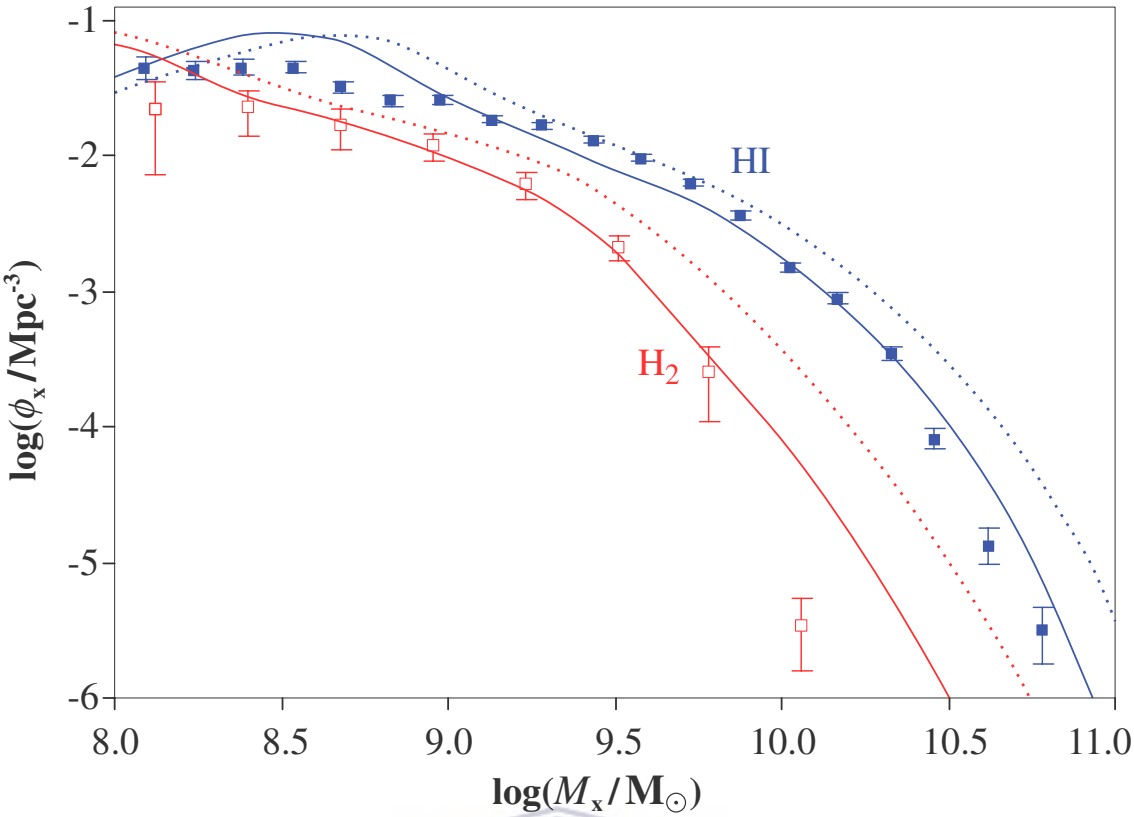


Figure 2.2: Galaxy mass functions (MFs) for HI (blue) and H_2(red) for the results obtained by O2009a. Solid lines show results of the MFs with the constant correction for all cold gas masses while the dotted lines show MFs without the correction. The filled squares with error bars represent empirical MFs from Zwaan et al. (2005) while the open squares with error bars represent empirical mass functions from Obreschkow and Rawlings (2009). (O2009a, Figure 3).

UNIVERSITY of the
WESTERN CAPE

obtained.

The comparisons made for the TFR results in Figure 2.3 indicates good consistency between the results obtained from the O2009a simulations and observations from McGaugh et al. (2000). Given the fact that the TFR is one of the most important scaling relations, the ability of the O2009a simulations in producing such a relation further supports their correctness and reliability.

A final crucial test that this work considered for the simulations of O2009a was on the relation between the HI mass ($M_{\rm HI}$) and HI radius ($r_{\rm HI}$) for galaxies in the local Universe ($z = 0$) (e.g., Verheijen, 2001). This relation is very crucial as it can be used to predict

https://etd.uwc.ac.za

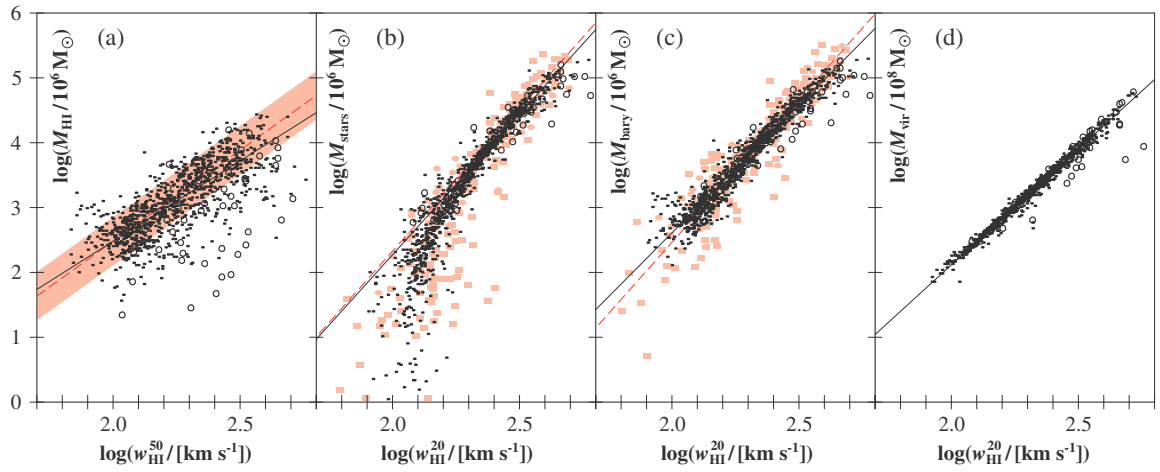


Figure 2.3: TFRs from O2009a. The black dots and circles represent simulated spiral and elliptical galaxies. Solid lines represent power-law fits to the simulated spiral galaxies. Plot (a) shows the TFR between $M_{\rm HI}$ and $w^{50}_{\rm HI}$. The red-dashed line and the red-shaded zone represent O2009as observational determination and the 1σ scatter of the HI-TFR from the HIPASS data. Plot (b) shows the simulated stellar-mass TFR together with observational data from McGaugh et al. (2000). The red dots and dashed line represent observational data and the power-law regression from the study of McGaugh et al. (2000). Plot (c) shows the simulated baryonic TFR and observational data from McGaugh et al. (2000). The red dots and dashed line are defined in the same manner as in Plot (b). Plot (d) shows the simulated total dynamical mass, considered as the virial mass $M_{\rm vir}$ and the circular velocity, given by the line width $w^{20}_{\rm HI}$. (O2009a, Figure 13).

the distribution of the bright, well-resolved galaxies for some existing catalogues such as HIPASS. The relation can also be used to estimate the HI diameters of unresolved galaxies based on the HI masses that will be detected by forthcoming experiments such as WALLABY (Widefield ASKAP L-band Legacy All-sky Blind surveY, Duffy et al., 2012). The relation is also able to predict the HI disc size distribution based on the HI mass obtained from single-dish surveys. Shown in Figure 2.4 are the results for the relation between HI mass and HI radius obtained by O2009a at $z = 0$.

Figure 2.4 further supports the correctness of the O2009a simulations since their results – for the relation between $M_{\rm HI}$ and $r_{\rm HI}$ – match observational data of Verheijen (2001).

The good match obtained between simulations and observations by O2009a supports their semi-analytic models and renders their simulations as being highly reliable.

Based on the fact that the physical models described in Obreschkow et al. (2009a,b,c) were

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

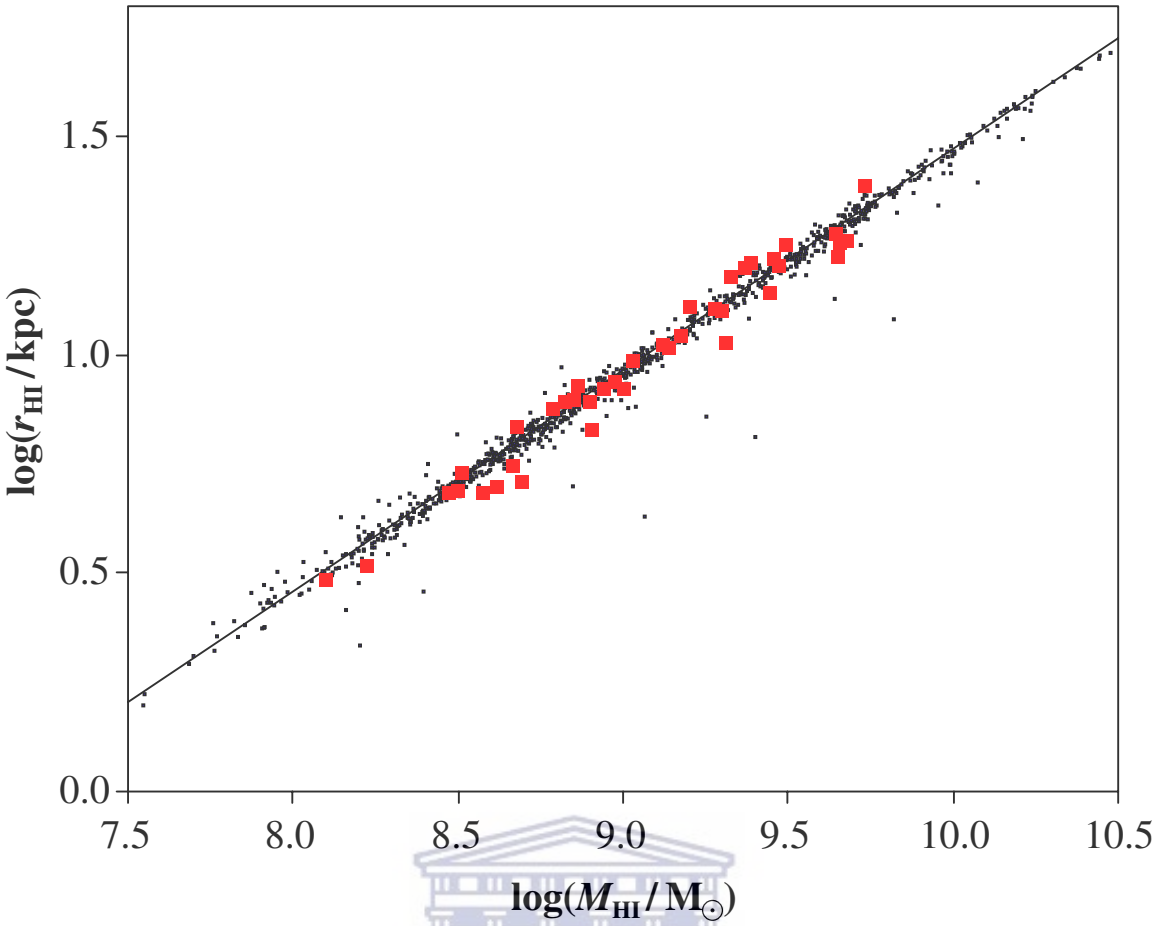


Figure 2.4: The relation between HI radius and HI mass for galaxies in the local Universe ($z = 0$). Black dots represent the simulated galaxies from O2009a whereas the red squares show the measurements done by Verheijen (2001) in the Ursa Major group, using the same definition of r_{HI} as in O2009a. (O2009a, Figure 8).

UNIVERSITY of the
WESTERN CAPE

successful, O&M14 further went on to use them to produce a mock galaxy catalogue. The catalogue covers a 10-by-10 degrees field and a redshift range of $z = 0 - 1.2$. It deals with smaller cosmological volumes and is a lite version of the semi-analytic suite of the SKA (Square Kilometre Array) Simulated Skies (S^3-SAX), which was delivered in 2009 as part of the European SKA design studies (SKADS); based on the simulations of Obreschkow et al. (2009a,b,c). S^3-SAX covers hundreds of squared degrees and is dependent on the user-defined maximal redshift. The O&M14 catalogue is complete down to HI masses above 10^8 $M_\odot$ for 21 cm peak flux densities above 1 Jy. The main distinguishing feature in the O&M14 catalogue – when compared to the original S^3-SAX – is that it includes optical properties such as the stellar masses, optical radii and apparent Vega R-band magnitudes. These properties were derived from the extinction-corrected

https://etd.uwc.ac.za

absolute magnitudes that were simulated using a variable K-correction. The optical properties were fitted to empirically-determined K-corrections as a function of redshift and morphology.

The O&M14 catalogue contains 20 properties per galaxy, which are shown in Table 2.1.

Col	Symbol	Unit	Description
1	ID	–	Unique galaxy identifier in the Munich Semi-Analytic Model of "DeLucia2006a" Unique galaxy identifier in the Munich Semi-Analytic Model of "DeLucia2006a"
2	RA	deg	Right ascension of galaxy centre
3	Dec	deg	Declination of galaxy centre
4	z	–	Apparent redshift of galaxy centre, including the Doppler component due to peculiar motion relative to the Hubble expansion
5	i	deg	Galaxy inclination defined as the smaller angle ($0° - 90°$) the line-of-sight and the rotational axis of the galaxy
6	T	–	Numerical Hubble type ($-6...0$ for ellipticals, $0...10$ for spirals, 99 for morphologically-unresolved objects, mostly dwarfs)
7	M_*	$M_\odot$	Stellar mass
8	$M_{\rm HI}$	$M_\odot$	Mass of neutral atomic hydrogen HI, without helium
9	$M_{\rm H_2}$	$M_\odot$	Mass of molecular hydrogen H_2, without helium
10	$S_{\rm HI}^{\rm int}$	Jy km s^{-1}	Velocity-integrated flux of the redshifted 21 cm HI emission line, with velocity units defined in the galaxy rest-frame
11	$S_{\rm HI}^{\rm peak}$	Jy	Peak flux density of the HI emission line; typically the flux density of the 'horns'
12	$S_{\rm CO}^{\rm int}$	Jy km s^{-1}	Velocity-integrated flux of the redshifted 115.27 GHz ^{12}CO$(1-0)$ emission line, with velocity units defined in the galaxy rest-frame
13	$S_{\rm CO}^{\rm peak}$	Jy	Peak flux density of the ^{12}CO$(1-0)$ emission line; typically the flux density of the 'horns'
14	$W_{\rm HI}^{50}$	km s^{-1}	Width of the HI emission line, in galaxy rest-frame velocity units, measured at 50% of the peak flux density
15	$W_{\rm HI}^{20}$	km s^{-1}	Width of the HI emission line, in galaxy rest-frame velocity units, measured at 20% of the peak flux density
16	$r_{\rm HI}^{\rm edge}$	arcsec	Apparent HI radius along the major axis out to a HI disk surface density of 1 $M_\odot$ pc^{-2}, corresponding to a face-on column density of $1.25 \cdot 10^{20}$ cm^{-2}
17	$r_{\rm HI}^{\rm half}$	arcsec	Apparent HI half-mass radius along the major axis
18	$M_{\rm R}$	mag	Absolute Vega R-band magnitude, corrected for intrinsic dust extinction; 99 if stellar mass and star formation history are insufficiently resolved to compute $M_{\rm R}$
19	$m_{\rm R}$	mag	Apparent Vega R-band magnitude; value 99 if no absolute magnitudes available
20	$r_{\rm e}$	arcsec	Effective radius, here approximated as the radius containing half the stellar mass if the galaxy were viewed face-on

Table 2.1: Description of the columns of the mock catalogue in the O&M14 study. (O&M14, Table 1).

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

2.1.2 Galaxy models

Given the fact that the Obreschkow et al. (2009a,b,c) simulations were able to predict the properties of real galaxies in statistically accurate ways, O&M14 further went on to use the models to create a mock galaxy catalogue that contained additional optical properties. E16 then created a set of methods that convert the O&M14 catalogue into synthetic HI data cubes where they realistically model, in a fully three-dimensional way, the spatial and spectral distributions of HI for every galaxy in the catalogue. The methods followed by E16 in producing HI data cubes are explained in this subsection.

HI mass profiles

For the parameterisation of the radial distribution of HI mass for each galaxy, the following equation was adopted in E16:

$$\sum_{\rm HI}(R) = \frac{A\exp(-R^2/2h^2)}{1+\beta\exp(-1.6R^2/2h^2)}, \tag{2.4}$$

where A is a normalisation factor used to control the total HI mass. The numerator in Equation 2.4 is a one-dimensional Gaussian and the h parameter represents the standard deviation. β is a parameter that controls the central concentration of HI. For the construction of an HI mass profile for each galaxy in the O&M14 catalogue, E16 made a conversion of the half-mass radius along the major axis ($R_{\rm HI}^{\rm half}$) into an exponential disc scale length ($R_{\rm h}$). In Equation 2.4, E16 set h to $R_{\rm h}$ and β equal to the H_2/HI mass ratio at the galaxy centre ($R_{\rm mol}^{\rm c}$) obtained from the O&M14 catalogue. E16 modelled the cubes out to radius $R = 3.5h$.

Rotation curves (circular velocity profiles)

The modelling of the circular velocity profiles in E16 was done using the template rotation curves of disc galaxies obtained from Catinella et al. (2006). Catinella et al. (2006) constructed average rotation curves in separate I-band luminosity classes using a homogeneous sample of 2200 low-redshift disc galaxies. Their rotation curves are fitted with the Polyex model (Giovanelli and Haynes, 2002), which is an analytic function given by

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

the following equation:

$$V_{\rm PE}(R) = V_0(1 - e^{-R/R_{\rm PE}})\left(1 + \frac{\alpha R}{R_{\rm PE}}\right), \tag{2.5}$$

where V_0 determines the amplitude, $R_{\rm PE}$ determines the exponential scale of the inner rotation curve and α determines the slope of the outer rotation curve. To construct the rotation curves of galaxies based on their R-band magnitudes, E16 used the results from Table 2 of Catinella et al. (2006), which presents the Polyex model fits to the template rotation curves parameterised as functions of optical radius for various luminosity classes. E16 then converted from R- to I-band absolute magnitudes using the relation: $M_{\rm I} = M_{\rm R}$ 0.37, which is based on the evaluated optical magnitudes for $z < 0.1$ galaxies with stellar masses of $M_* > 10^{10}$ $M_\odot$ from the semi-analytic simulations of Duffy et al. (2012).

Three-dimensional modelling

To realistically model both the spatial and spectral distributions of the HI line emission in the simulated galaxies, E16 used 3D modelling. To do this, E16 used a routine that assumes axisymmetry and models a galaxy as a collection of many HI clouds to generate HI data cubes. These cubes are based on the radial profile parameterisation equations given by Equations 2.4 and 2.5. Shown in Figure 2.5 is an illustration showing a model generated for a single galaxy in E16.

UNIVERSITY of the
WESTERN CAPE

For each galaxy, the model requires the inclination and position of the HI disc together with a rotation curve. It also requires HI mass profiles and the HI velocity dispersions as functions of radius. It assumes an infinitely thin HI cloud and uses the HI mass profile and rotation curve as probability density functions to generate many random numbers that are representative of the spatial and spectral coordinates of the individual HI clouds. As a final step of the routine, the model places the cloud at the appropriate location in the data cube using its position and velocity.

To create a full synthetic HI data cube containing the HI line emission from many galaxies, E16 generated a unique 3-D model for each galaxy in a sub-volume from the O&M14

https://etd.uwc.ac.za

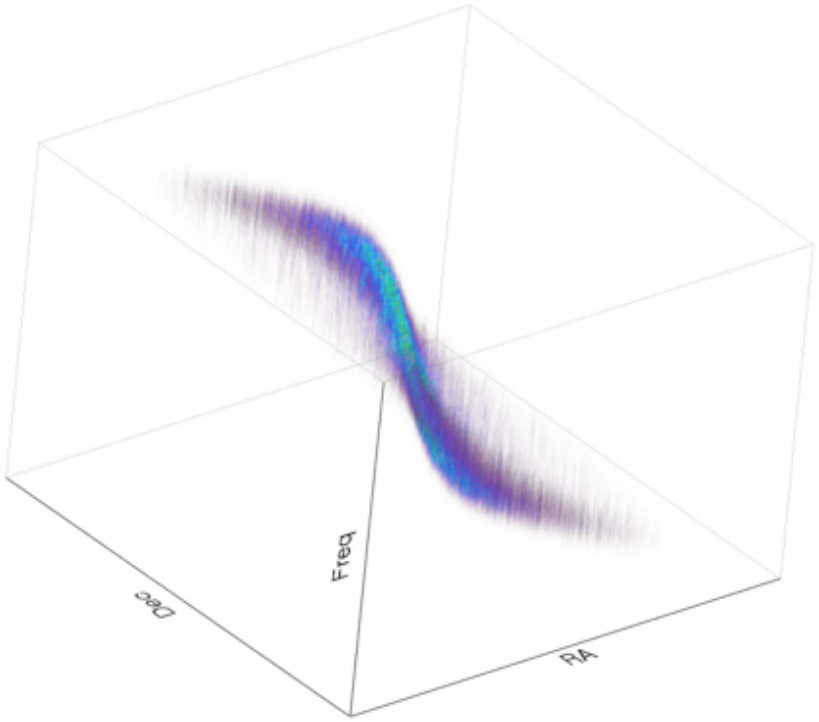


Figure 2.5: An example 3-D model generated for each galaxy in the E16 simulation.

UNIVERSITY of the
WESTERN CAPE

catalogue. The version of the cube that they created spans a sky area of $1.4° \times 1.4°$ at a redshift range of $0.7 < z < 0.758$. It contains 52 662 galaxies and a total HI mass of 3.53×10^{13} $M_{\odot}$. It has a spatial resolution of 15 arcsec and a spatial pixel size of 3 arcsec/pix. It has a spectral pixel size of 26 kHz, which corresponds to a velocity width $\Delta\nu = 9.497$ km·s^{-1} at the mean redshift of the cube. Elson et al. (2019, hereafter E19) used the same procedures above and created a cube similar to the one from E16, except that their cube includes noise. This version of the cube was used when testing the stacking codes written in this work, which will be described fully in Chapter 4. The E19 cube has Gaussian-distributed noise in each channel that has a standard deviation of 28.8 μJy·beam^{-1}. Figure 2.6 shows the distribution of various parameters for the galaxies in the E16 and E19 data cubes.

Of particular importance in Figure 2.6 is the fact that all galaxies have HI diameters less than 20 arcsec, shown by the HI diameter distributions in Figure 2.6 (e). This clearly shows how the typical galaxy properties compare with the properties expected from the simulated cubes, which will be discussed in Section 2.2. The spatial resolutions that

https://etd.uwc.ac.za

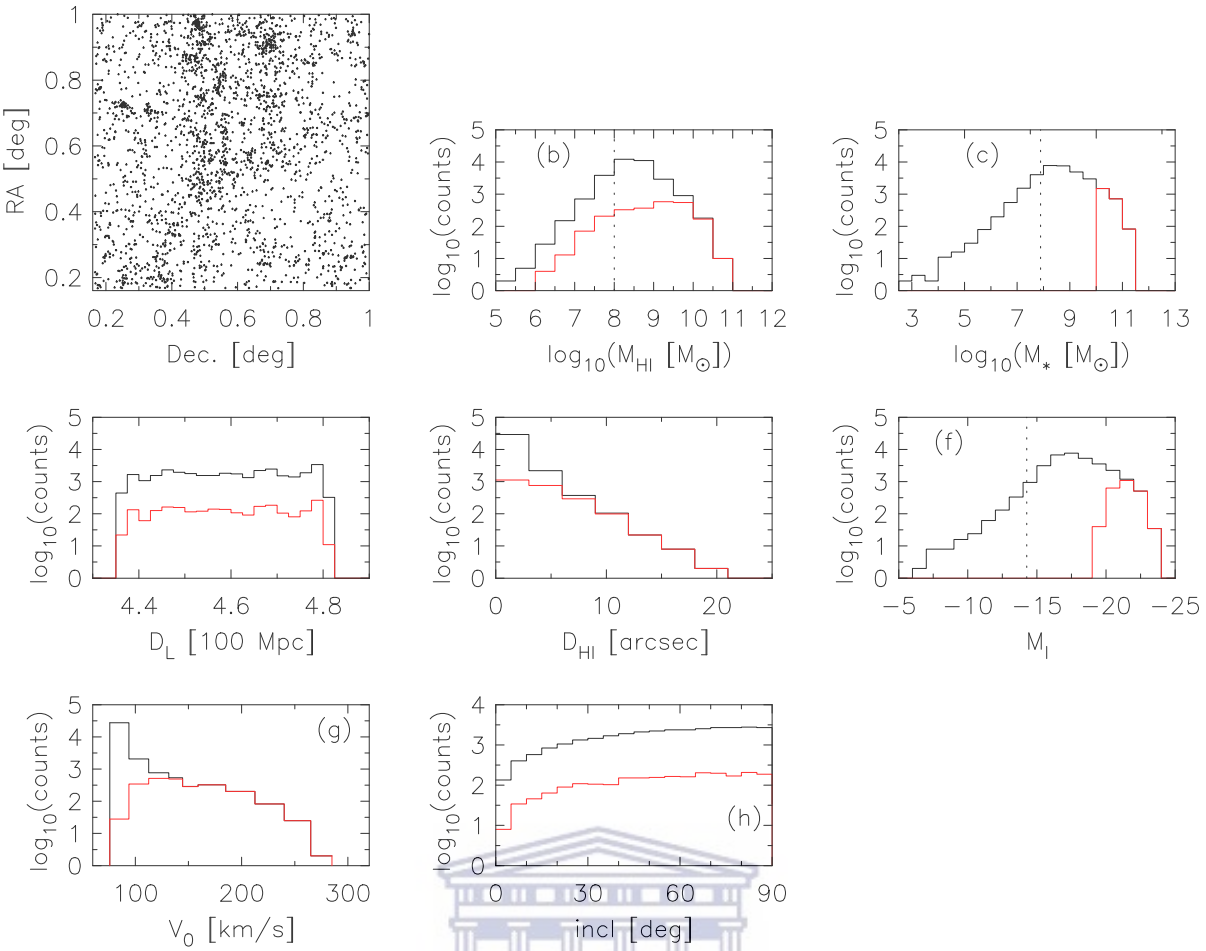


Figure 2.6: Evaluated galaxy properties for the E16 cube. The black histograms correspond to the full sample of galaxies in the data cube whereas the red histograms show results for a sample of galaxies having stellar masses $M_* > 10^{10}$ $M_\odot$, which were used to produce the co-added HI spectra. Plot (a) shows the RA-Dec positions of a random subset of galaxies in the E16 full-sized cube. Plot (b) shows the HI mass distributions. Plot (c) shows the stellar mass distributions. Plot (d) shows the luminosity distance distributions. Plot (e) shows the HI diameter distributions. Plot (f) shows the I-band absolute magnitudes distributions. Plot (g) shows the maximum rotation speed parameter (V_0) distributions. Plot (h) shows the distributions of the inclination angles of galaxies. (E16, Figure 4).

are considered in this work are from 7.5 to 110 arcsec. For almost all of the stacking experiments conducted in this work, the galaxies for which the spectra are extracted are completely spatially-unresolved.

https://etd.uwc.ac.za

2.2 New data products produced in this work

As already discussed in Section 2.1, galaxy properties have been predicted by Obreschkow et al. (2009a,b,c) and O&M14 created a mock catalogue of galaxy properties from their work. E16 used the O&M14 catalogue to produce synthetic HI data cubes out of it. The flux of each galaxy in the E16 cubes was realistically distributed, but the main drawback in the cubes was that when it comes to noise properties and PSF (Point Spread Function) effects, the HI data cubes lacked an extensive amount of realism, i.e., they did not take into account all the additional flux contributions that would occur in the most realistic case. This work further went on to produce several versions of the E16 cube that contain more realism by properly incorporating the characteristics of interferometric noise and PSF generated by the interferometers. This was done in CASA and was carried out for the VLA in its B, C and D configurations. The three considered configurations differ in terms of the layout of dishes. For convenience, this work refers to these three configurations as VLA-B, VLA-C and VLA-D. Shown in Figure 2.7 are the PSFs obtained from CASA simulation tasks using the aforementioned configurations.

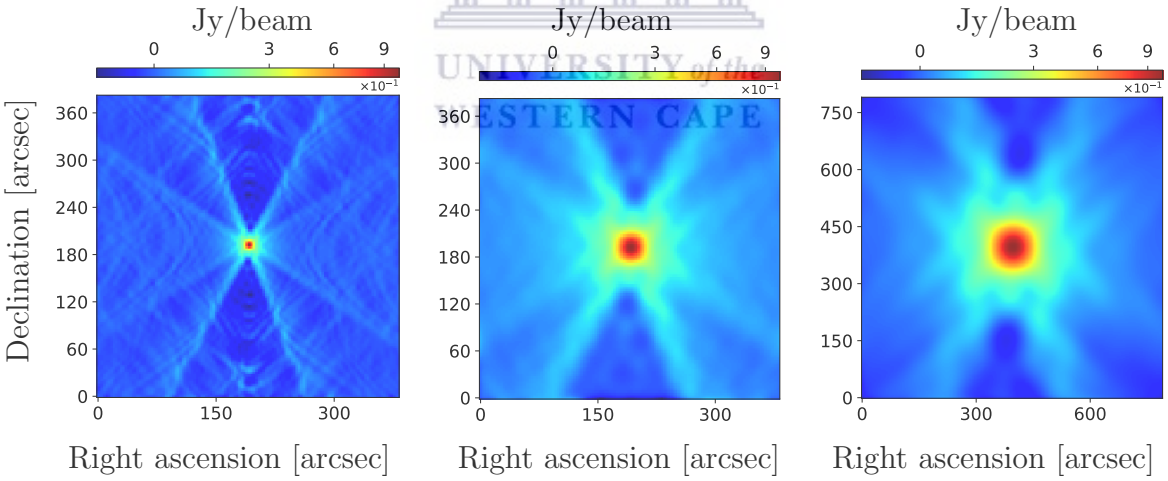


Figure 2.7: PSFs for the VLA-B (left), VLA-C (middle) and VLA-D (right) configurations.

The VLA-A configuration results were excluded since the size of its beam is smaller than the pixel size of the data cubes used in this work. Each of the configurations that are considered in this work has a unique baseline distribution which determines the typical angular resolution of the data products produced. Table 2.2 shows the maximum baseline

https://etd.uwc.ac.za

and the corresponding major beam axis for each of the three configurations of the VLA. Among the three configurations considered, the VLA-B configuration has the longest baseline distribution and thus has the highest angular resolution for a given frequency. The VLA-D configuration has the shortest baseline distribution and the lowest angular resolution for a given frequency.

Configuration	$B_{\rm max}$ [km]	BMAJ [arcsec]
VLA-B	11.10	9.89
VLA-C	3.40	31.78
VLA-D	1.03	99.00

Table 2.2: Table showing the maximum baseline ($B_{\rm max}$) and the major beam axis (BMAJ) corresponding to the various VLA configurations.

To generate new HI data cubes containing realistic PSF effects (which are henceforth referred to as the CASA cubes), two main steps were applied in CASA. The first step involved producing a set of simulated visibilities/Measurement Sets (MS) using the Simobserve task. Once the visibilities were produced, a second step involved imaging the visibilities using the CLEAN task. These steps were undertaken to ultimately produce HI data cubes that would later be used for the stacking analysis in this work. The cube used in the E16 study was smoothed down to a spatial resolution of the LADUMA survey (15 arcsec) and since CASA requires a pure skymodel, this work obtained and used a subset of the unsmoothed version of the E16 cube as the input cube for the CASA simulations. The dimensions of the new version of the E16 cube that was used are 600 $\times$ 600 pixels in right ascension and declination, with a spectral range of 1000 channels. The cube has spatial and spectral pixel sizes of 3″/pix and 26 kHz, respectively. The steps that were followed when generating the CASA cubes are described in subsections 2.2.1 and 2.2.2.

2.2.1 Simobserve task

The Simobserve task creates visibility data from a model image. It simulates total power measurements from an interferometer. The task takes in the central position of the field,

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

the total observing time, the integration time, the input model cube (which is entered as the skymodel parameter), the redshift range of the cube, the channel width of the cube, the angular size of each pixel, the frequency range of the cube and the VLA configuration file names. The task was run for three trials, where a significant change in each trial was made on the configuration file parameter for specifying each of the three different VLA configurations. This resulted in three folders of visibilities corresponding to each of the three VLA configurations.

Table 2.3 shows the parameters that were used when running the Simobserve task.

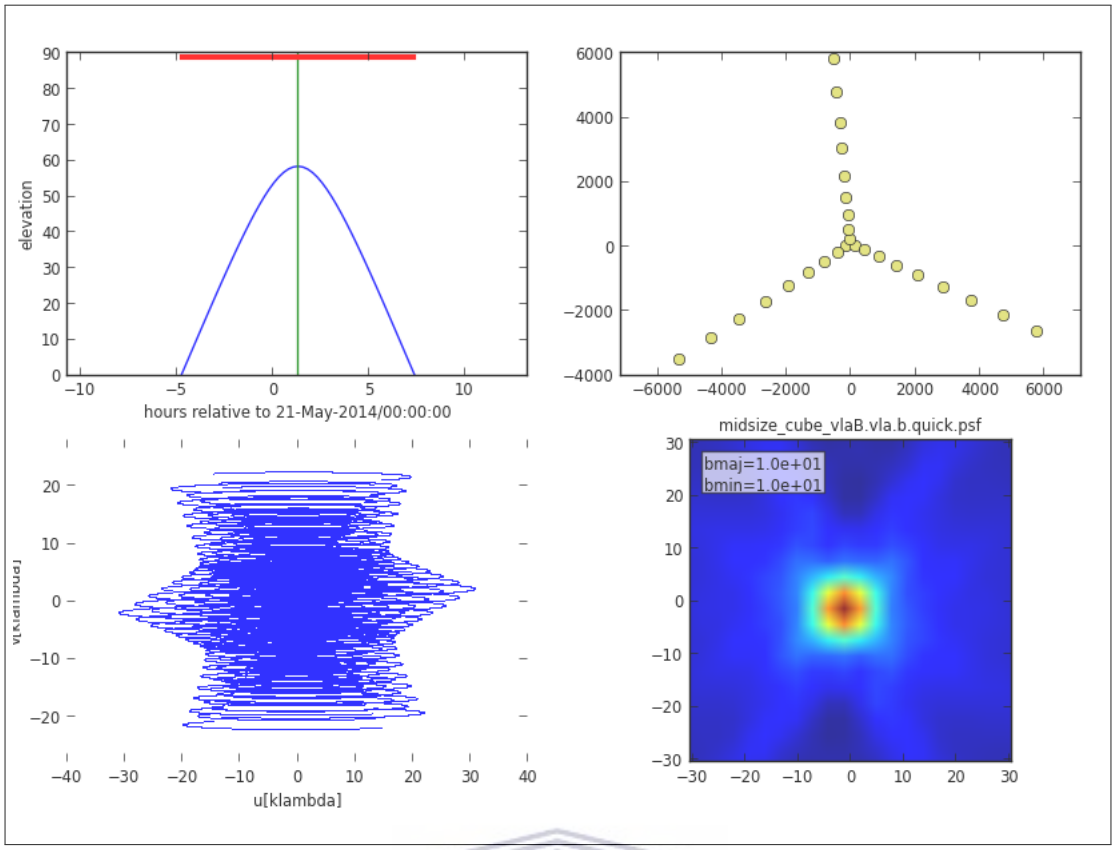
Total observing time	12 hours
Integration time	10 seconds
Field centre	J2000 10h00m00 02d30m00
Redshift range	$0.7 \leq z \leq 0.758$
Frequency range	$1420/(1+z_{\rm min}) \leq f \leq 1420/(1+z_{\rm max})$
Channel width (df)	26 kHz
VLA configuration files	VLA-B = vla.b.cfg, VLA-C = vla.c.cfg and VLA-D = vla.d.cfg

Table 2.3: Parameters used in the Simoberve task in CASA.

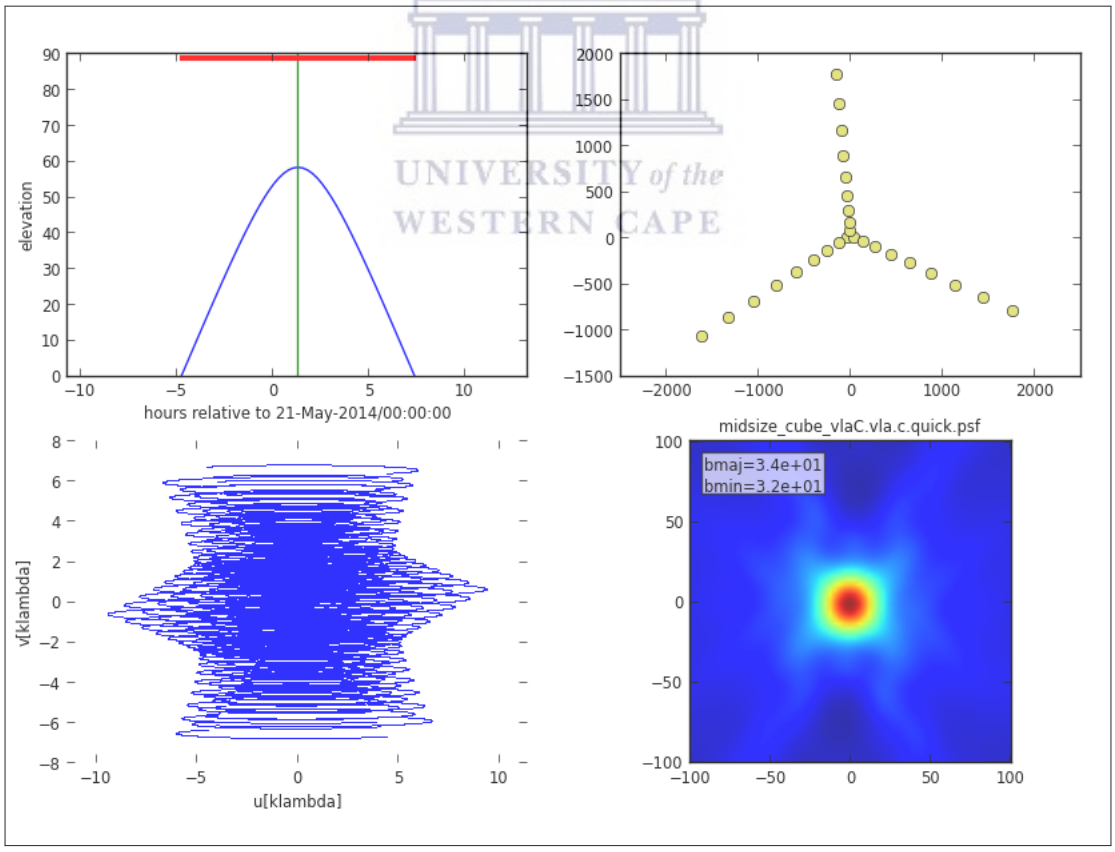
UNIVERSITY of the
WESTERN CAPE

The field centre in Table 2.3 was chosen as an intermediate case which gives a good baseline distribution and a good PSF due to the source attaining a maximum elevation of approximately 60°. Low-elevation cases that would yield a messy PSF were avoided. Furthermore, the cases where the sources would appear overhead were avoided as this is seldom the case. Upon running the Simboserve task, visibilities were created and the outputs shown in Figure 2.8 were obtained.

https://etd.uwc.ac.za



VLA-B results



VLA-C results

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

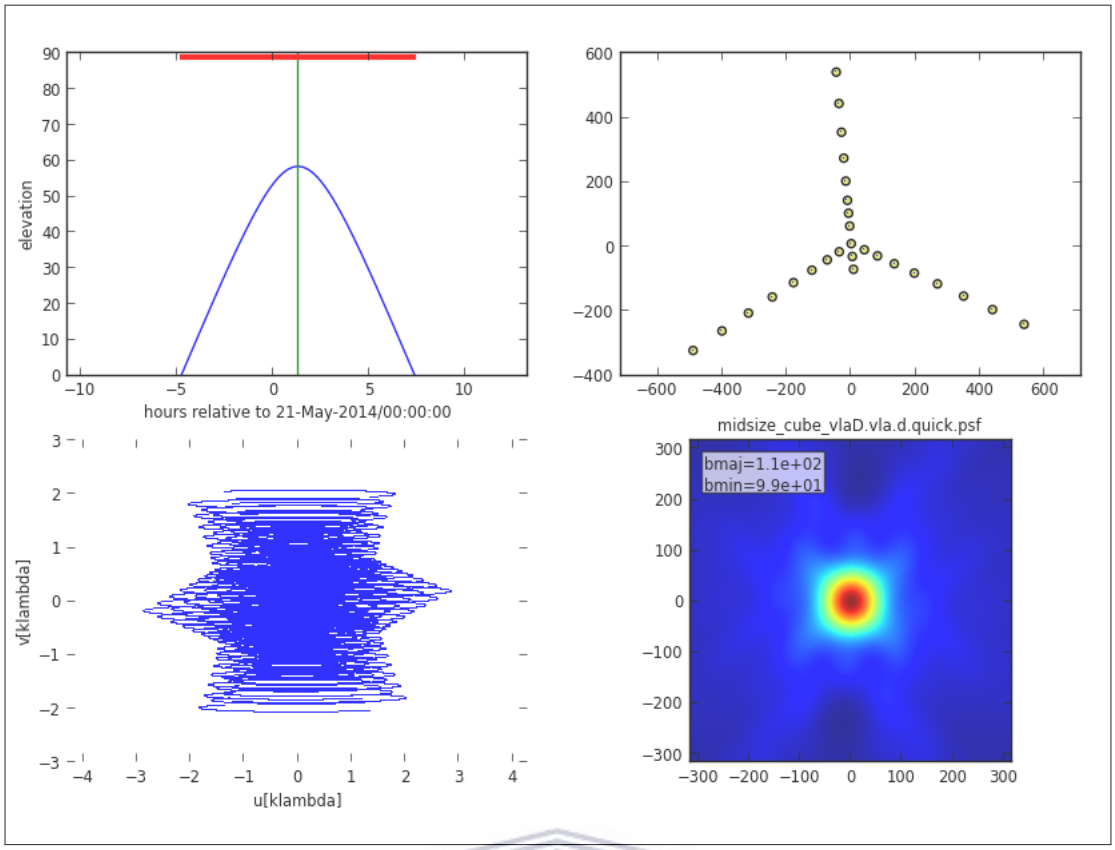


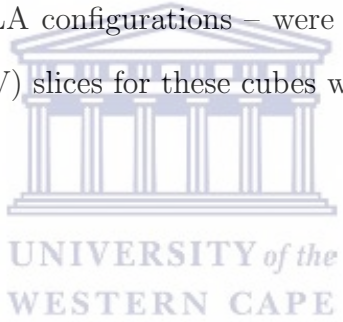
Figure 2.8: Outputs generated from the Simobserve task. Shown from top to bottom are sets of panels containing results for the VLA-B, VLA-C and VLA-D configurations (indicated at the bottom of each set of panels). In each of the three sets of panels shown, the top-left plot represents the elevation as a function of time for the modelled source, the top-right plot represents the VLA antenna positions in units of meters to the central VLA reference point, the bottom-left plot shows the uv coverage of the simulations and the bottom-right plot represents the PSF of the observation in units of arcsec for both the vertical (declination) and the horizontal (right ascension) axes. The PSF is fitted with a Gaussian of the specified major and minor beam sizes.

2.2.2 CLEAN task

Having used Simobserve to simulate the visibilities of the sky model for each of the VLA configurations, the CLEAN task was used to image the visibilities in order to produce HI spectral line cubes. This task inverts and deconvolves images with selected algorithms to form images from visibilities. The task takes in the name of the input visibilities folder, the x and y image size in pixels, the x and y cell size(s) and the name of the output image. The image size was set to match the size of the input cube, which resulted in spectral

https://etd.uwc.ac.za

line cubes of dimensions 600 × 600 pixels and a spectral range of 1000 channels. Since the upcoming deep HI surveys – such as LADUMA – will produce "dirty" cubes because the galaxy emission in them will be buried below the noise level, this work focused on producing "dirty" simulated cubes to represent the realistic case. To produce such cubes, the number of clean iterations (niter) was set to zero when running the CLEAN task. The task was run on each of the three visibilities folders produced from the Simobserve task. After running the task for three trials (on all the visibilities of the three VLA configurations), three folders containing information in the form of CASA tables were produced. The folders were converted to FITS (Flexible Image Transport System) format using the exportfits task in CASA. This task takes in the name of the folder containing the CASA tables – corresponding to a specific VLA configuration – and the desired name for the output file. After running the task on all three folders for each of the respective VLA configurations created from the CLEAN task, three cubes with PSF effects (CASA cubes) – corresponding to all three VLA configurations – were produced. The various channel maps and position-velocity (P-V) slices for these cubes will be presented in Chapter 4.



https://etd.uwc.ac.za

Chapter 3

HI stacking methods

This chapter presents the basic details of the HI spectral line stacking methods developed and implemented in this work. To extract spectra of individual galaxies, the methods make use of the positions of galaxies obtained from the Obreschkow and Meyer (2014, hereafter O&M14) catalogue together with spatial and spectral information specified in the code. The obtained spectra for individual galaxies are then stacked to yield an average HI spectrum. A package that extracts and stacks spectra from the Elson et al. (2019, hereafter E19) cube was generated and the reliability of the package was tested by making comparisons of the results obtained here with some of the results from E19. After confirming that the stacking techniques work well, they were then applied on the CASA cubes that were produced using the methods described in chapter 2.

UNIVERSITY of the
WESTERN CAPE

3.1 The stacking method

The stacking code used in this work was written in Python. The code reads in an input HI data cube specified and the galaxy properties file from O&M14. The cube used in this section is the one produced in E19. In order to extract the spectrum of an individual galaxy, the central position of the target galaxy in the cube needs to be determined. Thereafter, a sub-volume must be extracted at that point using the user-specified spatial and spectral apertures. A spectral range of 230 channels (corresponding to a velocity range of 2184.31 km·s^{-1}) was used together with spatial apertures of sizes 9, 15, 21, 27,

https://etd.uwc.ac.za

30, 33, 39 and 45 arcsec. These spatial aperture sizes correspond to 0.6, 1, 1.4, 1.8, 2, 2.2, 2.6 and 3 half-power widths of the Gaussian beam used by E19 to spatially smooth their cube. Given the pixel scale of 3 arcsec/pixel, the spatial apertures have sizes of 3, 5, 7, 9, 10, 11, 13 and 15 pixels, respectively. This work assumes ΛCDM cosmology with the following parameters: $H_0 = 67.3$ km·s^{-1} Mpc^{-1}, $\Omega_\Lambda = 0.685$ and $\Omega_\mathrm{M} = 0.315$, which were all specified in the code. As mentioned above, the first part carried out in creating the sub-cubes for each target galaxy involved determining the centres of those galaxies. To do this, information from the O&M14 catalogue about the central right ascension, central declination and the redshift of each galaxy was used. Using this information together with the user-specified spatial and spectral apertures, sub-cubes were extracted for the galaxies. As in E19, only massive galaxies with $M_* > 10^{10}$ $M_\odot$ were extracted and co-added. This mass limit was chosen so as to compare with the work of E19.

Using the redshift information provided in the O&M14 catalogue, the luminosity distance for each galaxy was calculated in the code written in this work. Each sub-volume was converted into a spectrum by being spatially integrated. The spectra obtained were then presented as HI mass spectra using the following equation from E16:

UNIVERSITY of the
WESTERN CAPE

$$M_{\mathrm{HI},i}\ [M_\odot] = 2.36 \times 10^5\ D_L^2\ S_i\ dv\ \frac{1}{(1+z)}, \tag{3.1}$$

where $M_{\mathrm{HI},i}$ is the HI mass in channel i, D_L is the luminosity distance of the target galaxy (in Mpc), dv is the velocity width of a channel (in km·s^{-1}), S_i is the flux density in channel i in the sub-cube (in Jy) and z is the redshift of the galaxy. dv is redshift-dependent and it was calculated using the following relation from E16:

$$dv = \frac{df \times c(1+z)}{f_\mathrm{emit}}, \tag{3.2}$$

where df is the channel width, c is the speed of light in vacuum, z is the redshift of a galaxy and f_emit is the rest frequency for HI. E19 models its galaxies to have maximum rotation speeds of less than 300 km·s^{-1}, and thus the target galaxy mass is expected to

https://etd.uwc.ac.za

be contained within a velocity range of ± 300 km·s^{-1}. When calculating the total HI mass from the co-added spectra obtained in this work, only the mass in the ± 300 km·s^{-1} velocity range was considered. The aforementioned methods for extracting individual sub-volumes from a large data cube were implemented and they worked efficiently. Using these methods on the version of the E19 noise-free cube, individual galaxy spectra for an isolated as well as a crowded galaxy – using various spatial aperture sizes – were produced. Shown in Figures 3.1 and 3.2 are the results obtained.

The results shown in Figures 3.1 and 3.2 show an ideal case of galaxy spectra extracted in a noise-free cube. As shown by the results, increasing the spatial aperture size results in an increase in the HI mass obtained in each spectrum. This outcome occurs for two reasons: (1) for small values of N, the spatial aperture is significantly smaller than the main lobe of the PSF. It is within the main lobe that most of the galaxy's mass is contained. The amount of mass contained within a spatial aperture will increase as the size of the aperture approaches the size of the main lobe; (2) the extracted spectrum receives additional mass from nearby neighbouring galaxies, which is known as source confusion. For the isolated galaxy shown in Figure 3.1, it can be noted that confusion does not play a crucial role in the total HI mass obtained since there are minor contributions of contaminant flux coming from neighbouring galaxies. These results show a slight increase in the total HI mass obtained with increasing spatial aperture size. For the crowded galaxy shown in Figure 3.2, source confusion has significant contributions and its effects can be noted even when using small spatial aperture sizes. Contamination, which is defined as the ratio of the extracted mass to the true (evaluated) mass from the O&M14 catalogue ($\langle M_{\rm HI}^{\rm cube}\rangle/\langle M_{\rm HI}^{\rm true}\rangle$), increases sharply with increasing spatial aperture size due to source confusion effects from multiple neighbouring galaxies. This result is well-demonstrated by the results shown in the panels of Figures 3.1 and 3.2, respectively. As shown from the obtained results, additional spectra are visible within the ± 300 km·s^{-1} spectral range for all the different spatial aperture sizes used. Figure 3.2 demonstrates a way in which source confusion plays a significant role in this study. It shows that even when extracting

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

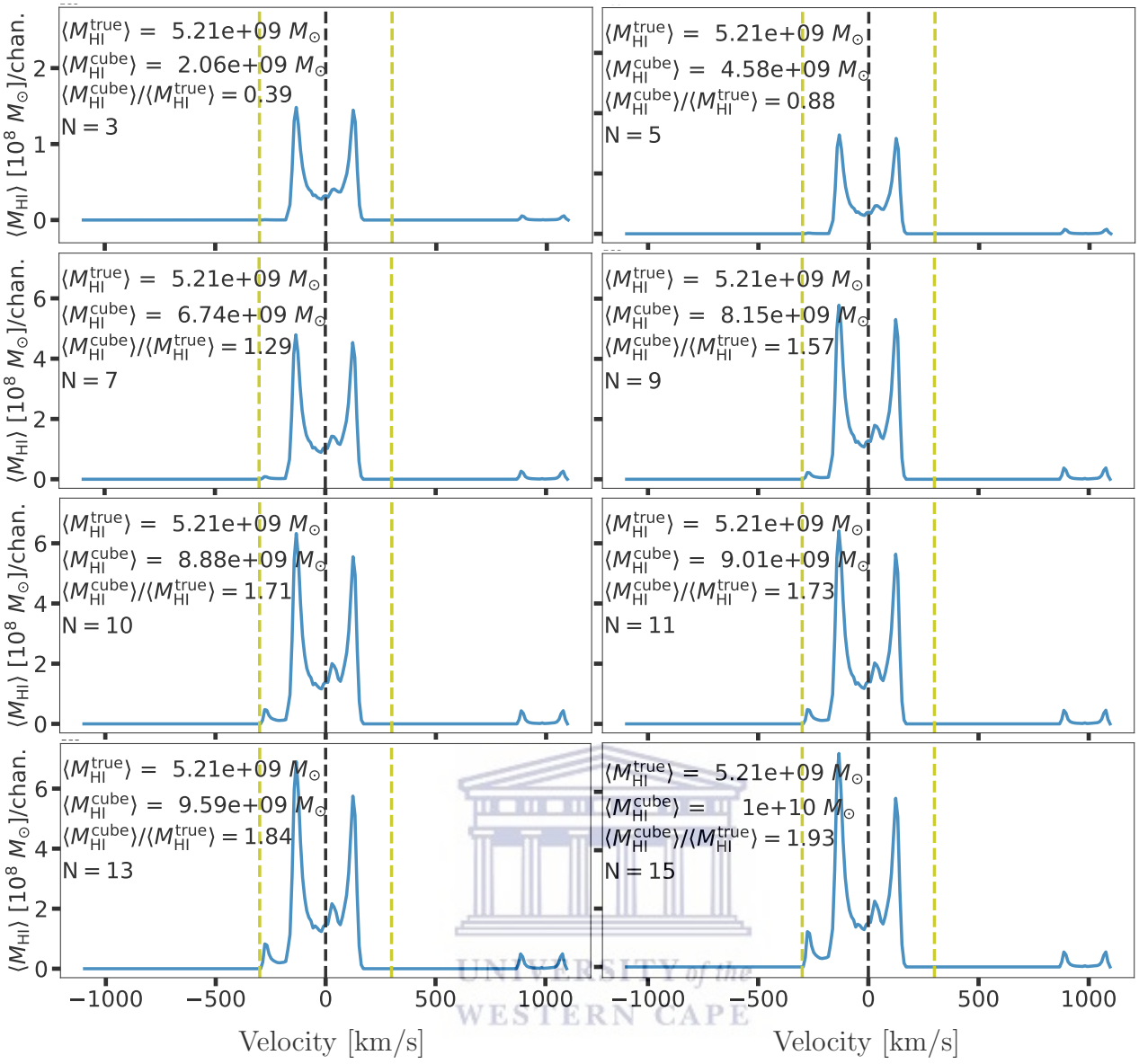


Figure 3.1: Individual HI spectra for an isolated galaxy with ID number 3733297 extracted from the smoothed noise-free E19 cube. The spectra were extracted using a spectral aperture of 2184.31 km·s^{-1} and the spatial aperture sizes (N) – in units of pixels – indicated in each panel. $\langle M_{\rm HI}^{\rm true}\rangle$ indicates the true (evaluated) average HI mass of the target galaxy obtained from the semi-analytic catalogue while $\langle M_{\rm HI}^{\rm cube}\rangle$ indicates the HI mass obtained from the extracted spectrum obtained in this work. $\langle M_{\rm HI}^{\rm cube}\rangle/\langle M_{\rm HI}^{\rm true}\rangle$ shows the ratio of the mass extracted in the cube to the true HI mass obtained from the O&M14 catalogue. The vertical yellow-dashed lines delimit the velocity range ± 300 km·s^{-1}. This is the range over which the stacked spectra were integrated in order to determine the average stacked mass of a galaxy.

individual galaxy spectra using small spatial aperture sizes, most of the regions where target galaxies are located are dominated by contaminant mass from neighbouring galaxies.

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

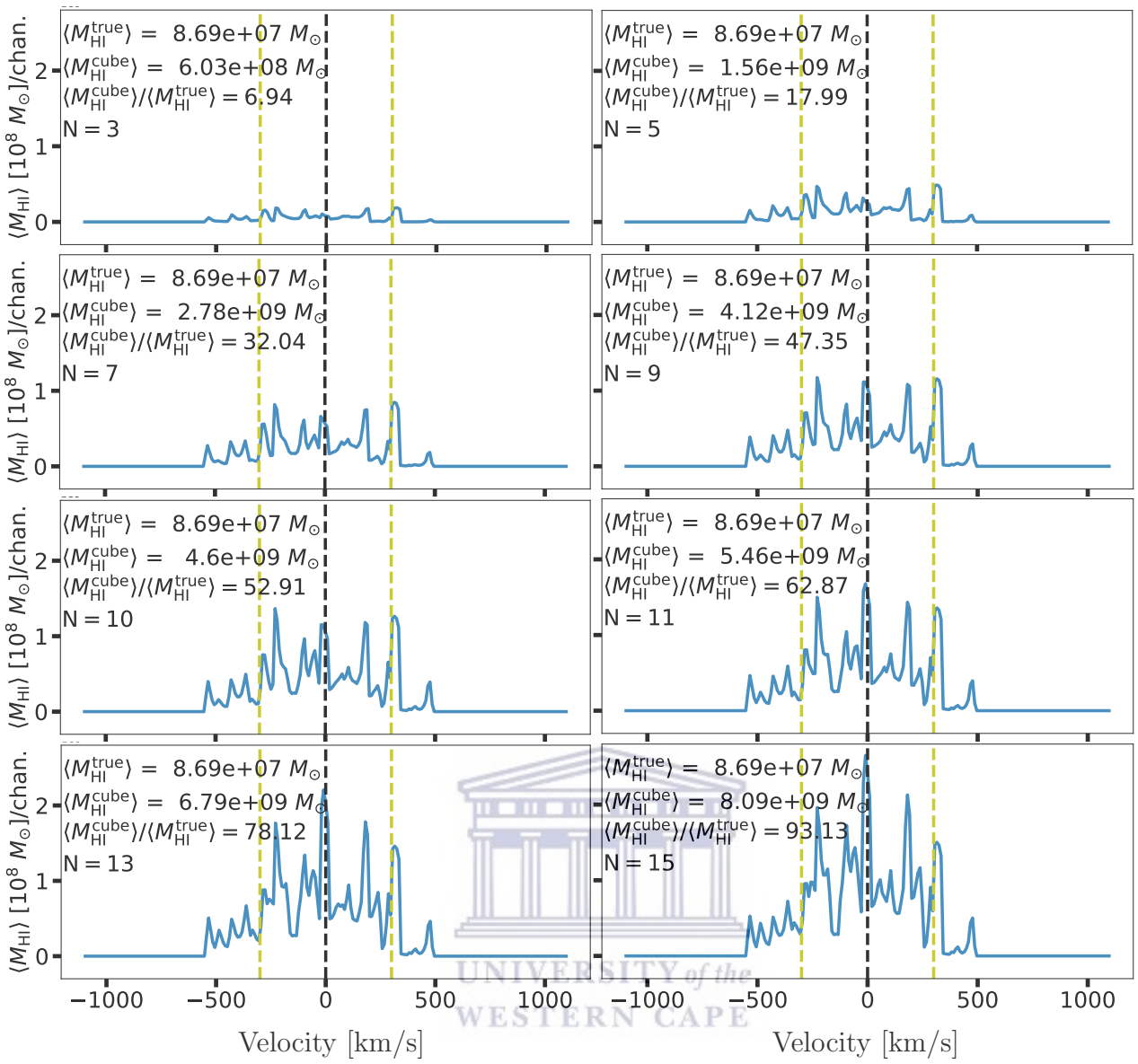


Figure 3.2: Individual HI spectra for a crowded galaxy with ID number 3755931 extracted from the smoothed noise-free E19 cube. Notation is the same as in Figure 3.1.

Figure 3.3 shows the way in which contamination builds up as a function of changing spatial aperture sizes (N) for both the isolated and the crowded galaxies used in this example.

The example in Figure 3.3 presents results for the spectra extracted for two individual galaxies from the E19 cube. After extracting all the spectra for target galaxies with $M_* > 10^{10}$ $M_\odot$ in the cube, methods that would align and co-add the obtained galaxy spectra were included in the codes written in this work. Shown in Figure 3.4 are the stacked spectra produced by co-adding all the galaxies from the E19 cube with $M_* > 10^{10}$

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

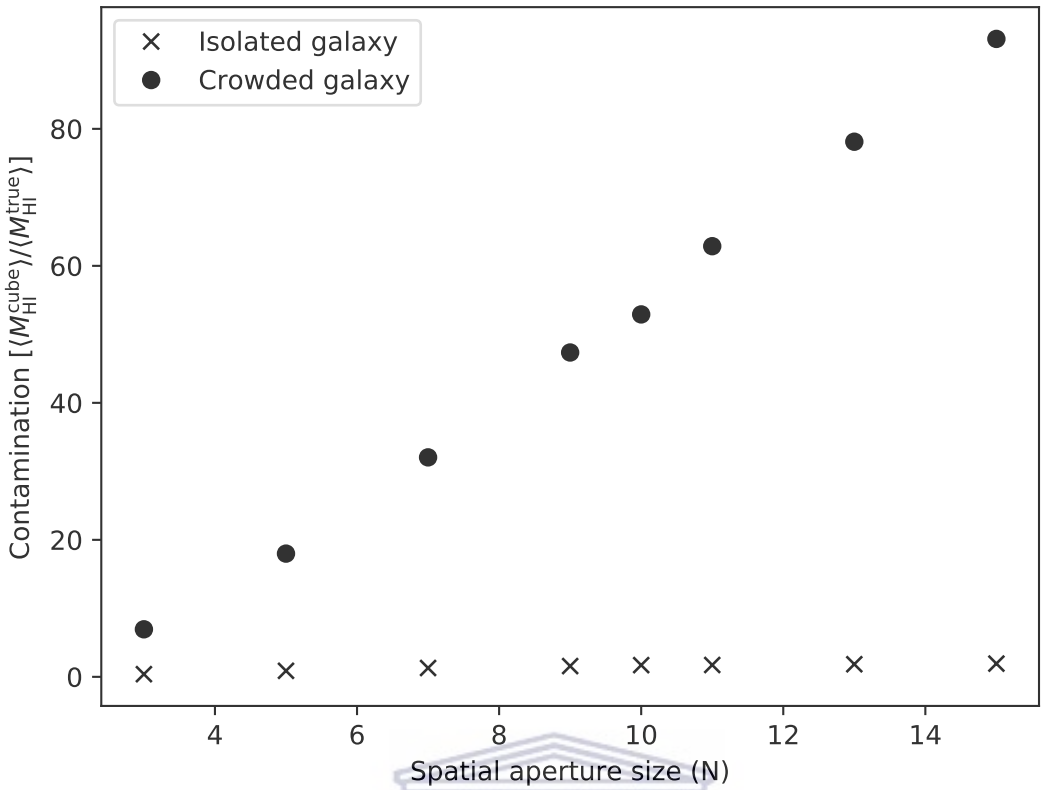


Figure 3.3: Contamination as a function of spatial aperture size for an isolated (shown as crosses) and a crowded galaxy (shown as circles). Clearly visible in the diagram is the manner in which contamination rises up sharply for a crowded galaxy as opposed to an isolated one.

UNIVERSITY of the
WESTERN CAPE

$M_{\odot}$, for the various spatial aperture sizes (N) indicated in each panel.

3.2 Verification of the stacking method

Methods that carry out the stacking technique were generated in this work, as mentioned in Section 3.1. Before applying them on the CASA cubes, their accuracy was confirmed by applying them on the E19 cube using the same spatial and spectral aperture sizes used in the E19 study. E19 uses the size of the Gaussian beam to determine the sizes of the sub-volumes extracted in the cube. They calculated the width of this beam using the central frequency ($\nu = 820$ MHz) of the redshift range $0.7 < z < 0.758$. They considered spatial aperture sizes of 15, 30 and 45 arcsec, which correspond to 1, 2 and 3 half-power widths of the Gaussian beam. These spatial apertures have sizes of 5, 10 and 15 pixels,

https://etd.uwc.ac.za

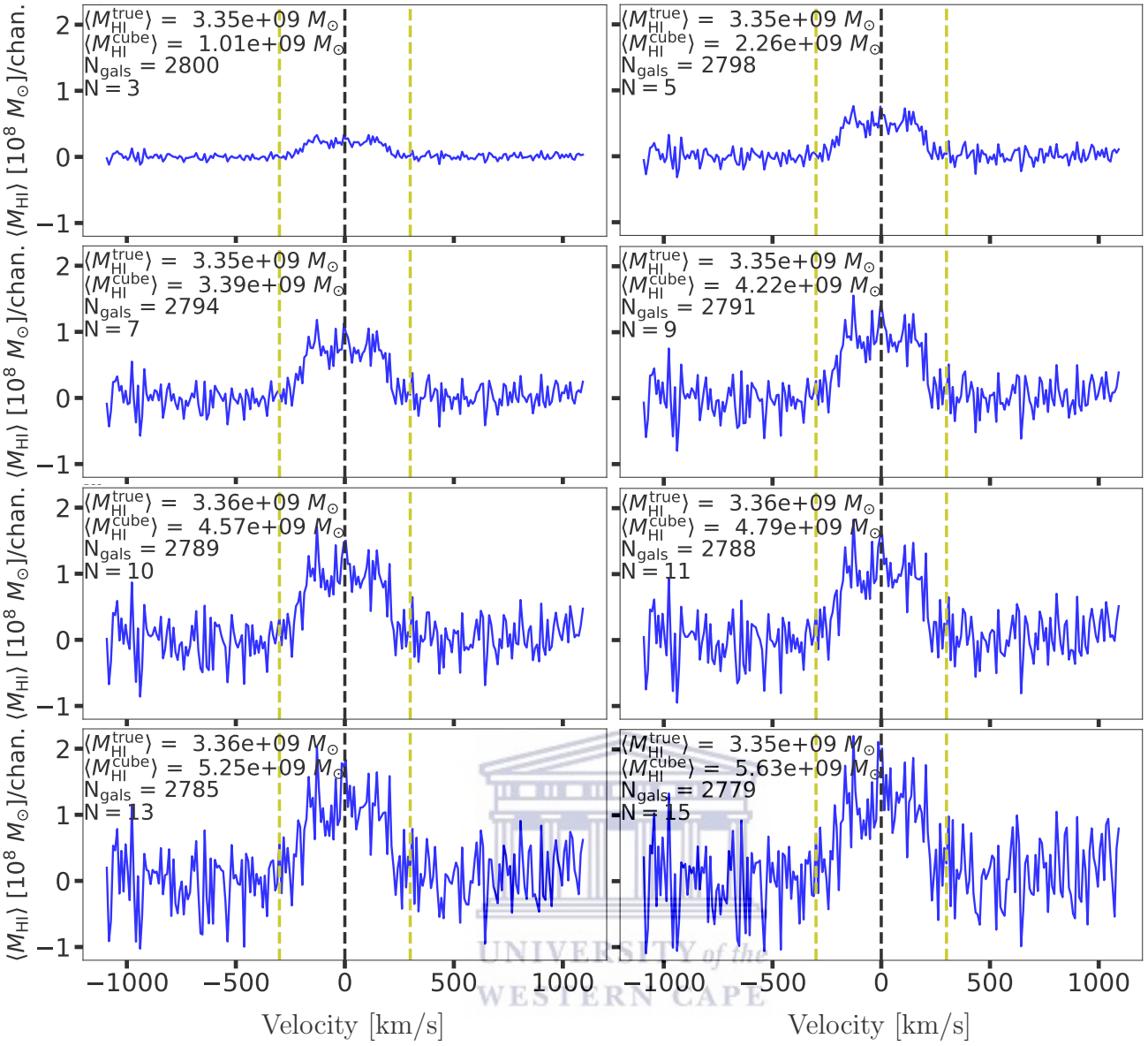


Figure 3.4: Co-added HI spectra obtained for the full-sized E19 cube. The spectral aperture used throughout was 2184.31 km·s^{-1} and the spatial aperture sizes (N) used – in units of pixels – are indicated in each panel. $\langle M_{\rm HI}^{\rm true}\rangle$ indicates the true (evaluated) average co-added HI mass obtained from the semi-analytic catalogue while $\langle M_{\rm HI}^{\rm cube}\rangle$ indicates the HI mass obtained from the co-added spectra obtained in this work. The vertical yellow-dashed lines delimit the velocity range ± 300 km·s^{-1}. This is the range over which the stacked spectra were integrated in order to determine the average stacked mass of a galaxy. $N_{\rm gals}$ represents the total number of galaxies stacked and it decreases with larger spatial aperture sizes. The latter occurs because a larger spatial aperture falls outside the bounds of the cube more often than the smaller apertures and, therefore, more galaxies near the edges are excluded from the stack as the spatial aperture size increases. For their stacked spectra, corresponding to spatial aperture sizes of N = 5, 10 and 15 in pixels, Elson et al. (2019) obtained HI masses of 2.54×10^9 $M_\odot$, 4.87×10^9 $M_\odot$ and 5.78×10^9 $M_\odot$, respectively. These values will be compared to the ones obtained in this work in Section 3.2.

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

given the spatial pixel size of 3 arcsec/pix. They carried out the stacking technique using the spatial aperture sizes mentioned above and a spectral aperture of 2184 km·s^{-1}. The co-added spectra obtained in this work – based on spatial apertures of N = 5, 10 and 15 pixels – may, therefore, be directly compared to the E19 co-adds to assess the degree to which this work recovers the same amount of mass that they do. Shown in Table 3.1 is a comparison of the HI masses obtained in this work and those from the E19 study.

θ [arcsec]	$\langle M_{\rm HI}^{\rm E19} \rangle$ $[10^9 M_\odot]$	$\langle M_{\rm HI}^{\rm This_work} \rangle$ $[10^9 M_\odot]$	% error $= \left\lvert \frac{A-B}{A} \right\rvert \times 100$ [%]
15	2.54×10^9	2.26×10^9	11.02
30	4.87×10^9	4.57×10^9	6.16
45	5.78×10^9	5.63×10^9	2.56

Table 3.1: Comparison of the co-added HI masses obtained in this work and those from E19, using the same spatial and spectral aperture sizes. Column 1 shows the spatial aperture sizes used. Column 2 shows the co-added HI mass from E19. Column 3 shows the co-added HI mass obtained in this work. Column 4 shows the percentage error for the known HI mass value from E19 and the HI mass value obtained in this work. In the equation, A shows the value from E19 (found in column 2) while B shows the HI mass obtained in this work (shown in column 3).

UNIVERSITY of the
WESTERN CAPE

The average HI masses that are inferred for the co-adds from this work (based on spatial apertures of size 15, 30 and 45 arcsec) are similar to the average HI masses generated from E19. The minor differences that arise are mainly due to the methods that were applied in this work when specifying the centres of the spatial apertures, which are most likely different from the ones adopted in E19. This results in central spatial positions that are different by about 1 pixel, which is especially the case for galaxies having an inherent spatial size of 1 pixel. Central positions that are off by just 1 pixel can yield a significantly different total mass, especially in the cases where the spatial aperture size is small (e.g. 15 arcsec, or only 5 pixels). Having used different methods – compared to those from E19 – for specifying the central positions of galaxies, the largest differences were expected for the total HI mass obtained for the co-added spectra obtained in this work and the one from E19 when using small spatial aperture sizes. When increasing the spatial aperture

https://etd.uwc.ac.za

size, a difference in the central spatial position of a single pixel has a much smaller impact on the total extracted mass. The close agreement between the co-added masses from this work and those from E19 leads to the conclusion that the stacking methods produced in this work are reliable. In Chapter 4, these stacking methods are applied to the new CASA cubes.

https://etd.uwc.ac.za

Chapter 4

Mock stacking experiments

This chapter presents the application of the stacking methods – discussed in Chapter 3 – on the data cubes that were produced in order to quantify the different amounts of additional flux in each co-added spectrum. Before presenting the results obtained, Section 4.1 shows the various representations of the Elson et al. (2016, hereafter E16) and the CASA-generated cubes. Section 4.2 then presents the HI stacking results obtained in this work.

UNIVERSITY of the
WESTERN CAPE

4.1 Data cube representations

This section is devoted to showing the various representations of the data cubes used in this work. Figure 4.1 shows the various channel maps for the E16 cube while Figures 4.2 to 4.4 show the various channels for the CASA data cubes produced in this work. Figure 4.5 shows the position-velocity (P-V) slices for the E16 cube while Figures 4.6 to 4.8 show the various P-V slices for the CASA data cubes. Finally, Figure 4.9 shows the various total intensity maps (moment 0 maps) for some of the massive ($M_* > 10^{10}\ M_\odot$) galaxies in the E16 cube.

https://etd.uwc.ac.za

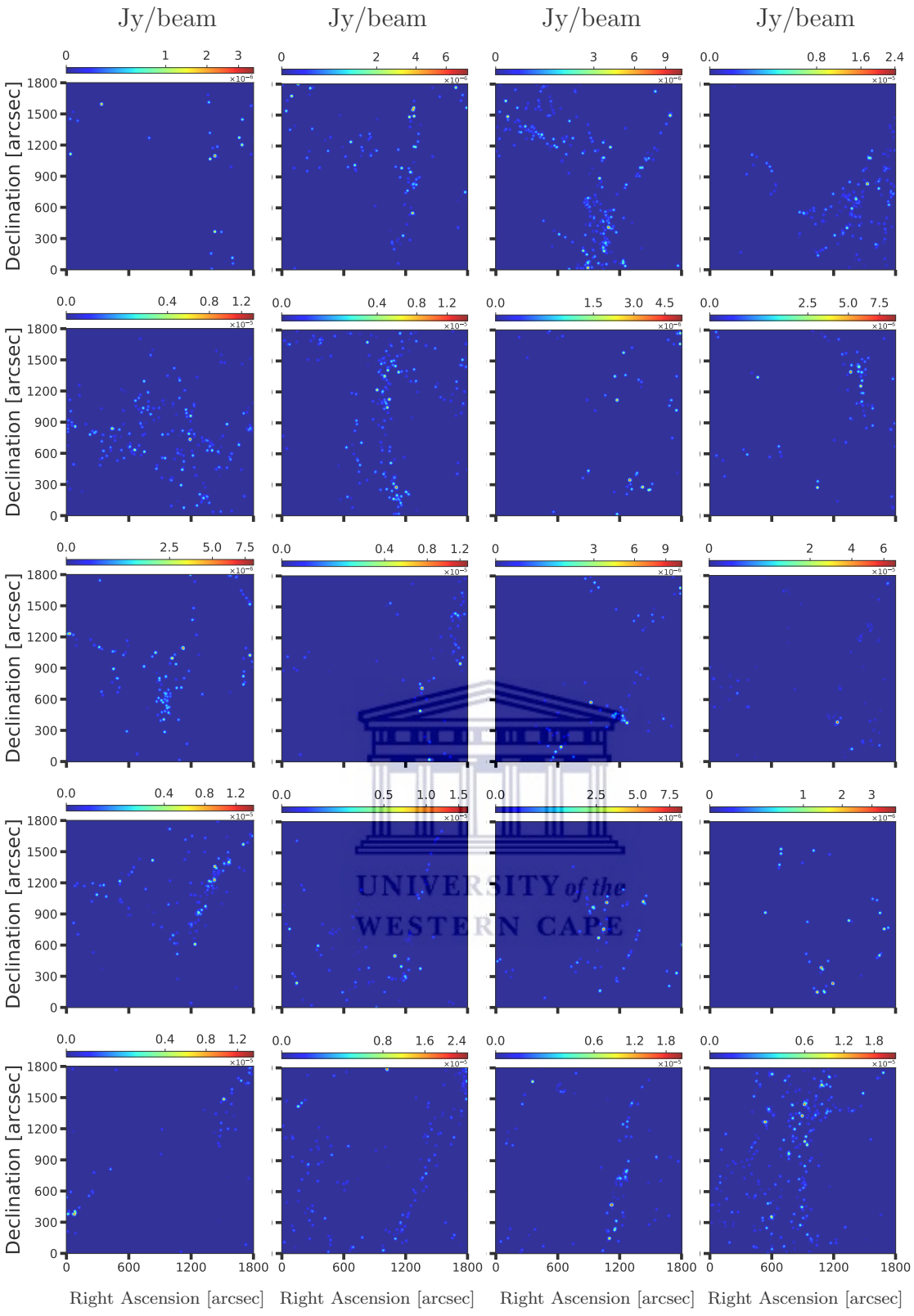


Figure 4.1: Channel maps for the E16 smoothed cube. The data cube consists of 1000 channels and shown in the diagram are sets of 9 adjacent channels averaged together – with the sets being spaced by 50 channels. The colour bars show the HI flux density in units of Jy/beam.

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

4.2 Descriptions of the stacks obtained in this project

After having tested the stacking code with the methods discussed in Chapter 3 and confirming that it works well, it was then applied on the realistic CASA cubes produced in this work. The sizes of the main Gaussian central lobes of each PSF for these CASA cubes are represented by the BMAJ (major axis of the beam) and BMIN (minor axis of the beam) parameters, which are found in the header of each of the cubes. Table 4.1 lists the parameters of each of the three CASA cubes that are based on the different VLA configurations.

	BMAJ [arcsec]	BMIN [arcsec]
VLA-B	10.00	9.89
VLA-C	33.00	31.78
VLA-D	110.00	99.00

Table 4.1: BMAJ and BMIN parameters of the three CASA-generated data cubes.

UNIVERSITY of the WESTERN CAPE

When co-adding individual galaxy spectra for the CASA cubes, the spatial aperture sizes used were integer multiples of the BMAJ values corresponding to each of the three VLA configurations. The sizes of these spatial apertures were N = 1, 2 and 3 $\times$ BMAJ. The spectral aperture used throughout was 151 channels (corresponding to a velocity range of 1434 km$\cdot$s^{-1}). Shown in Figure 4.10 are the results obtained for all the CASA cubes. The co-added spectra constitute some of the main results in this work and they will allow for the determination of the rates of confusion in cubes containing both high levels of source confusion and realistic PSF effects. In Chapter 5, the properties of these co-adds will be discussed in more detail.

https://etd.uwc.ac.za

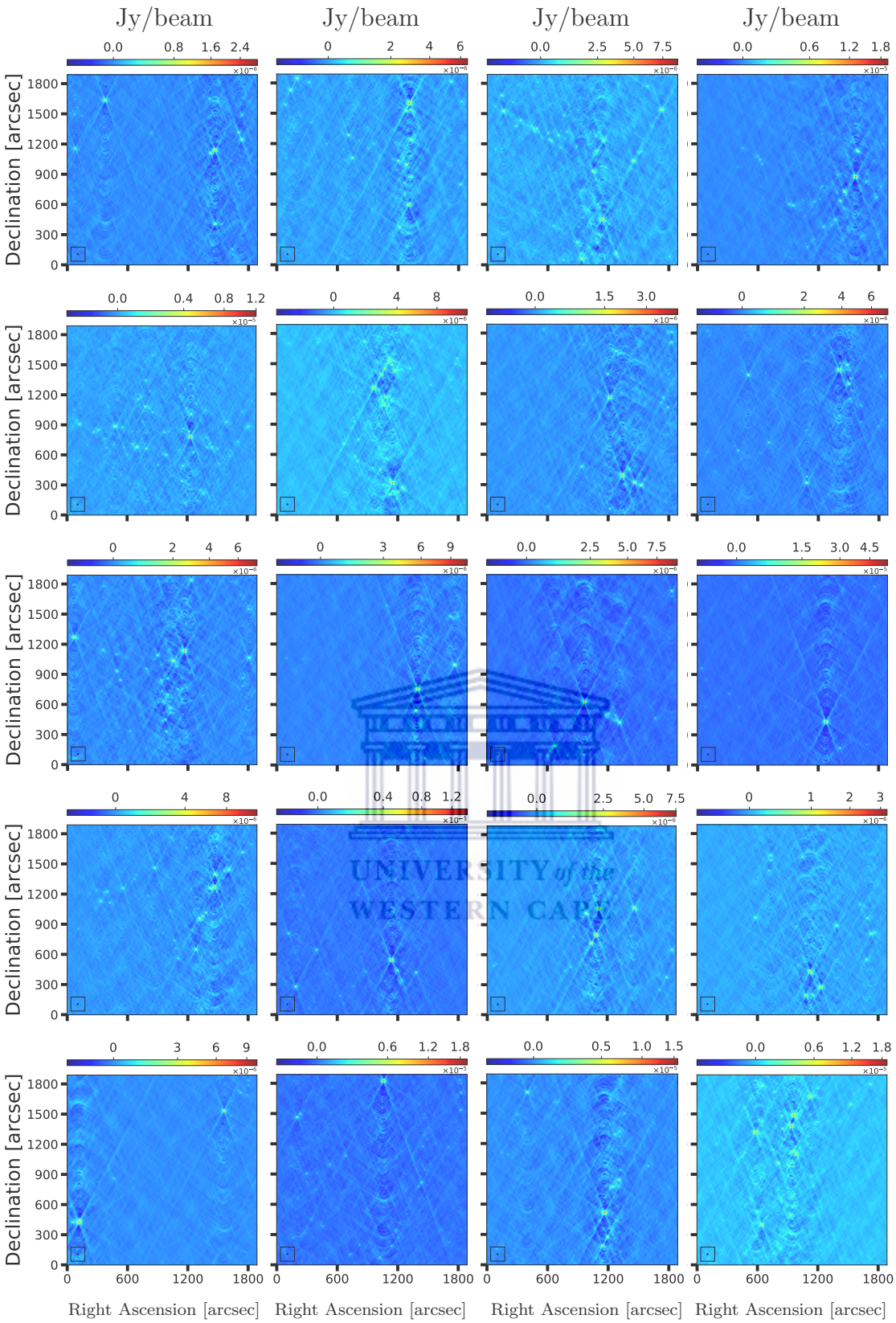


Figure 4.2: Channel maps for the VLA-B CASA cube. The data cube consists of 1000 channels and shown in the diagram are sets of 9 adjacent channels averaged together – with the sets being spaced by 50 channels. The colour bars show the HI flux density in units of Jy/beam. Shown in the bottom left of each panel is the size of the central lobe for this configuration.

https://etd.uwc.ac.za

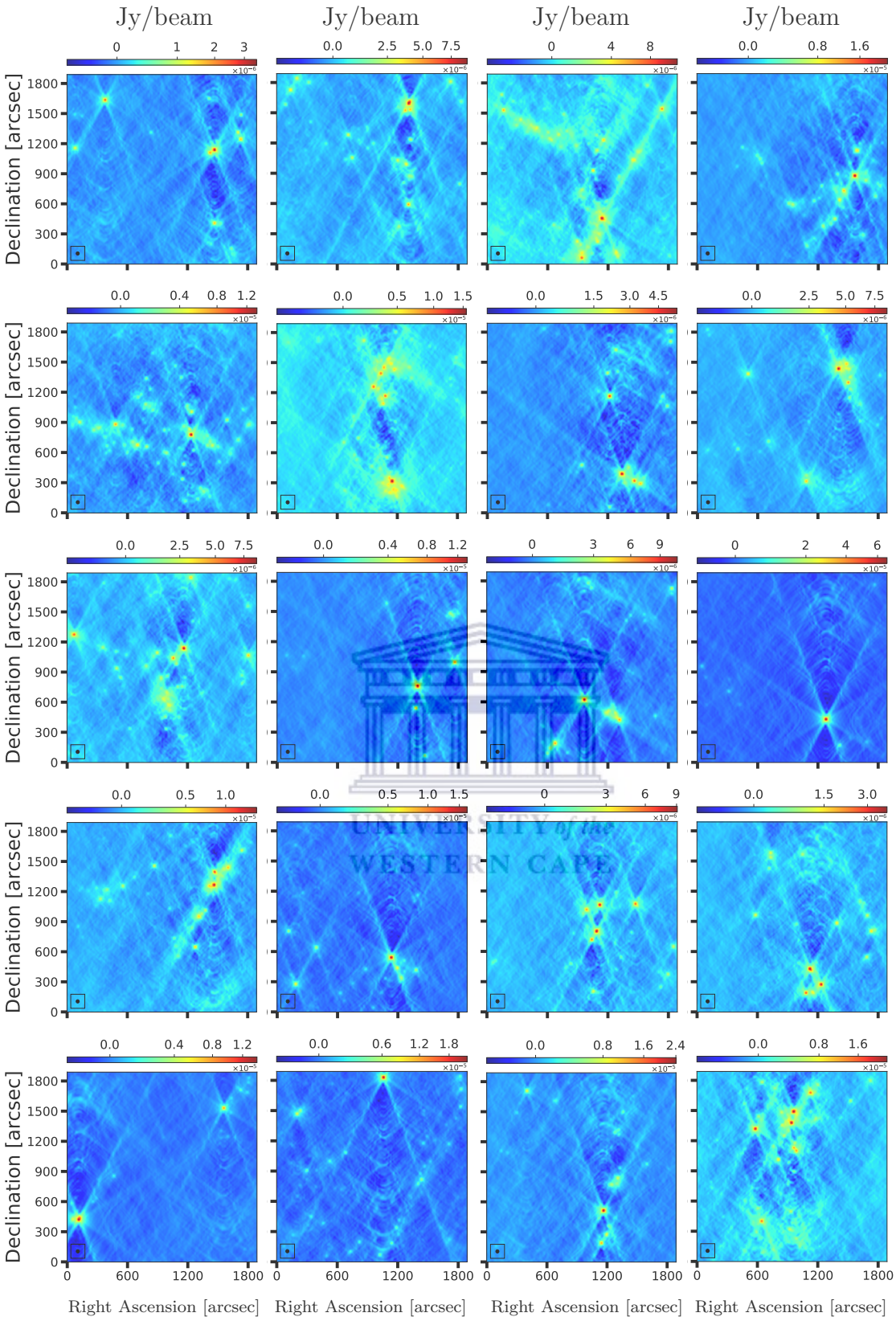


Figure 4.3: Channel maps for the VLA-C CASA cube. Notation is similar to that in Figure 4.2.

https://etd.uwc.ac.za

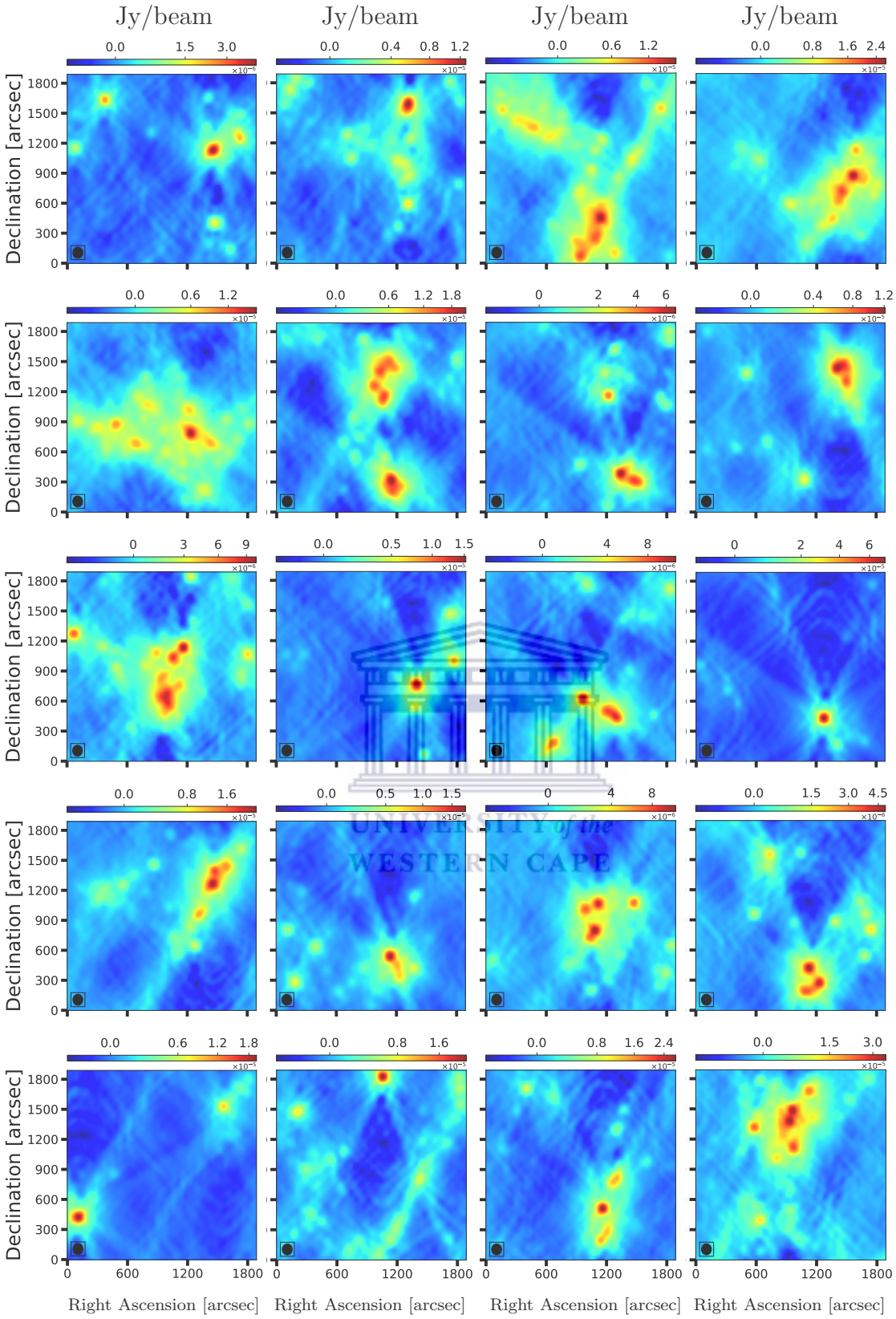


Figure 4.4: Channel maps for the VLA-D CASA cube. Notation is similar to that in Figure 4.2.

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

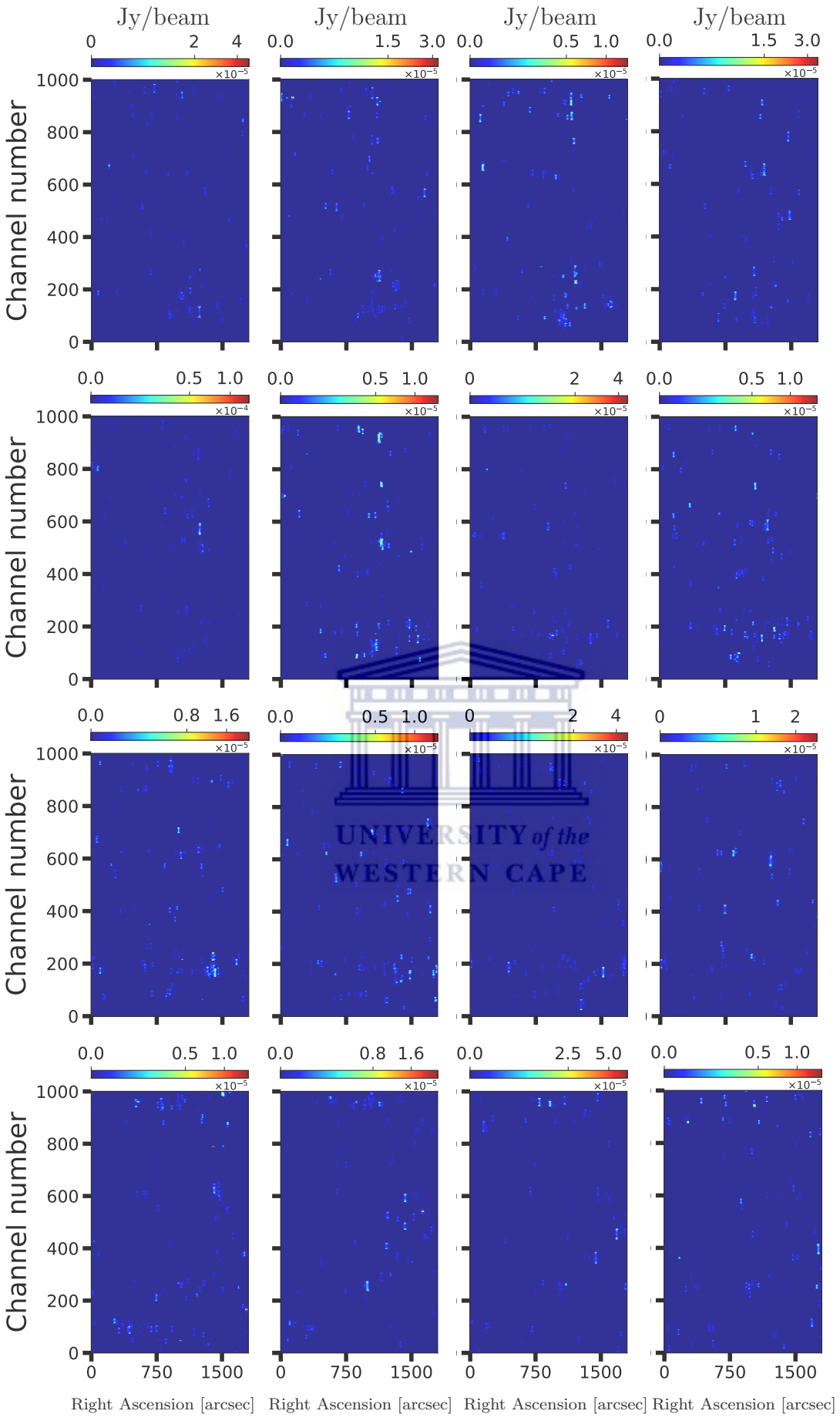


Figure 4.5: Position-velocity (P-V) diagrams for the E16 smoothed cube. The P-V slices were obtained by averaging a set of 6 adjacent declination pixels - with the sets being spaced by 25 pixels. The colour bars show the flux density in units of Jy/beam.

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

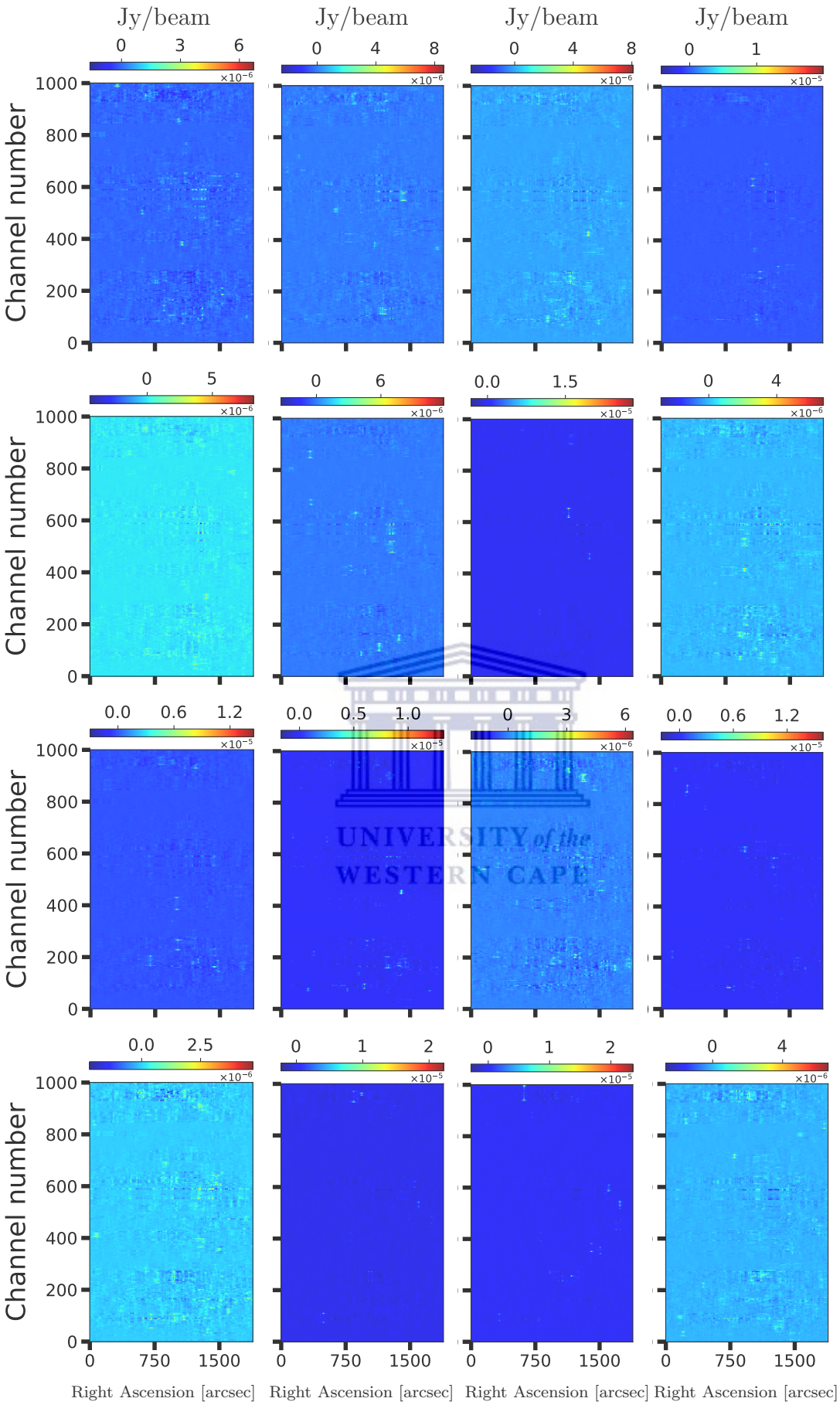


Figure 4.6: P-V diagrams for the VLA-B CASA data cube. Notation is similar to that in Figure 4.5.

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

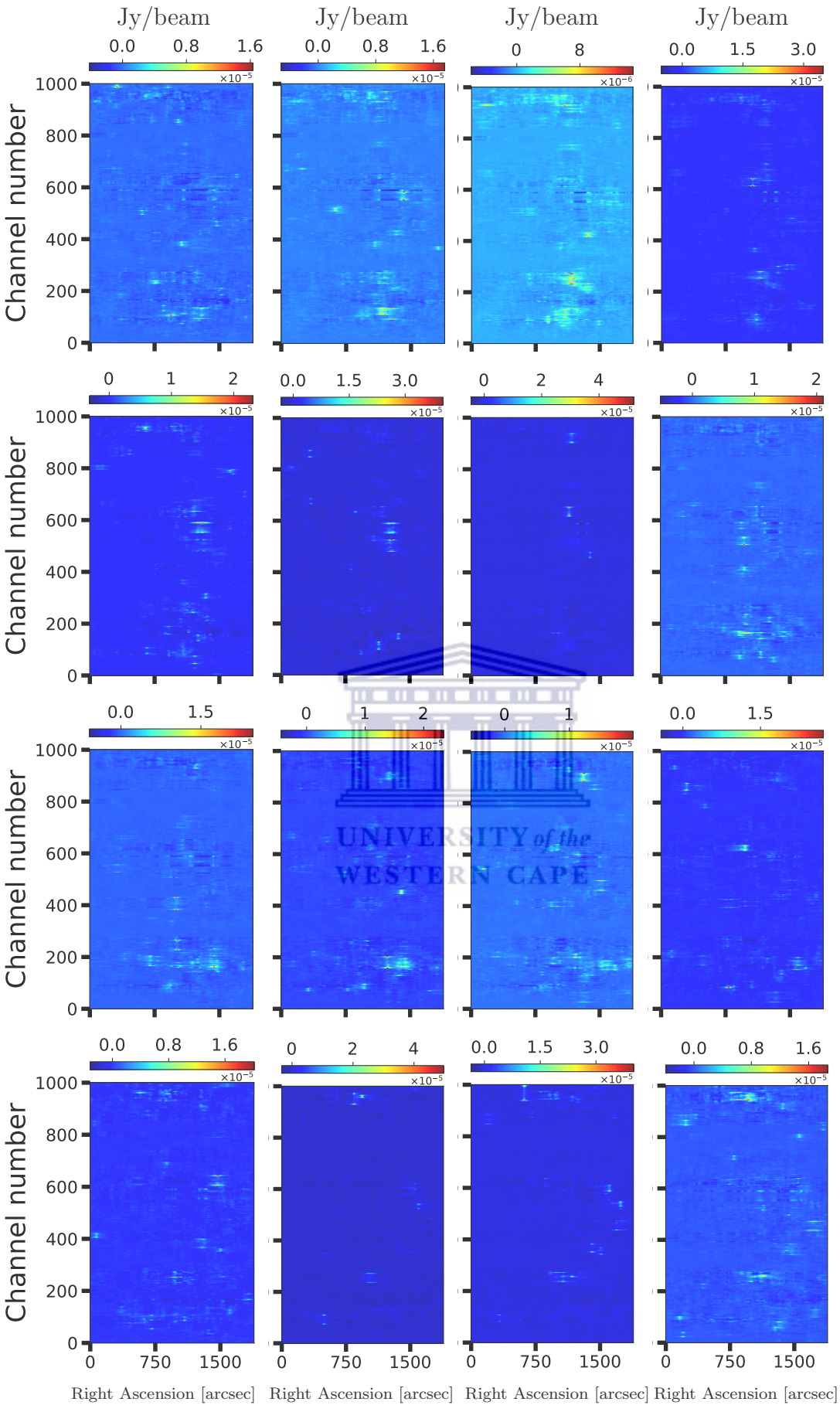


Figure 4.7: P-V diagrams for the VLA-C CASA data cube. Notation is similar to that in Figure 4.5.

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

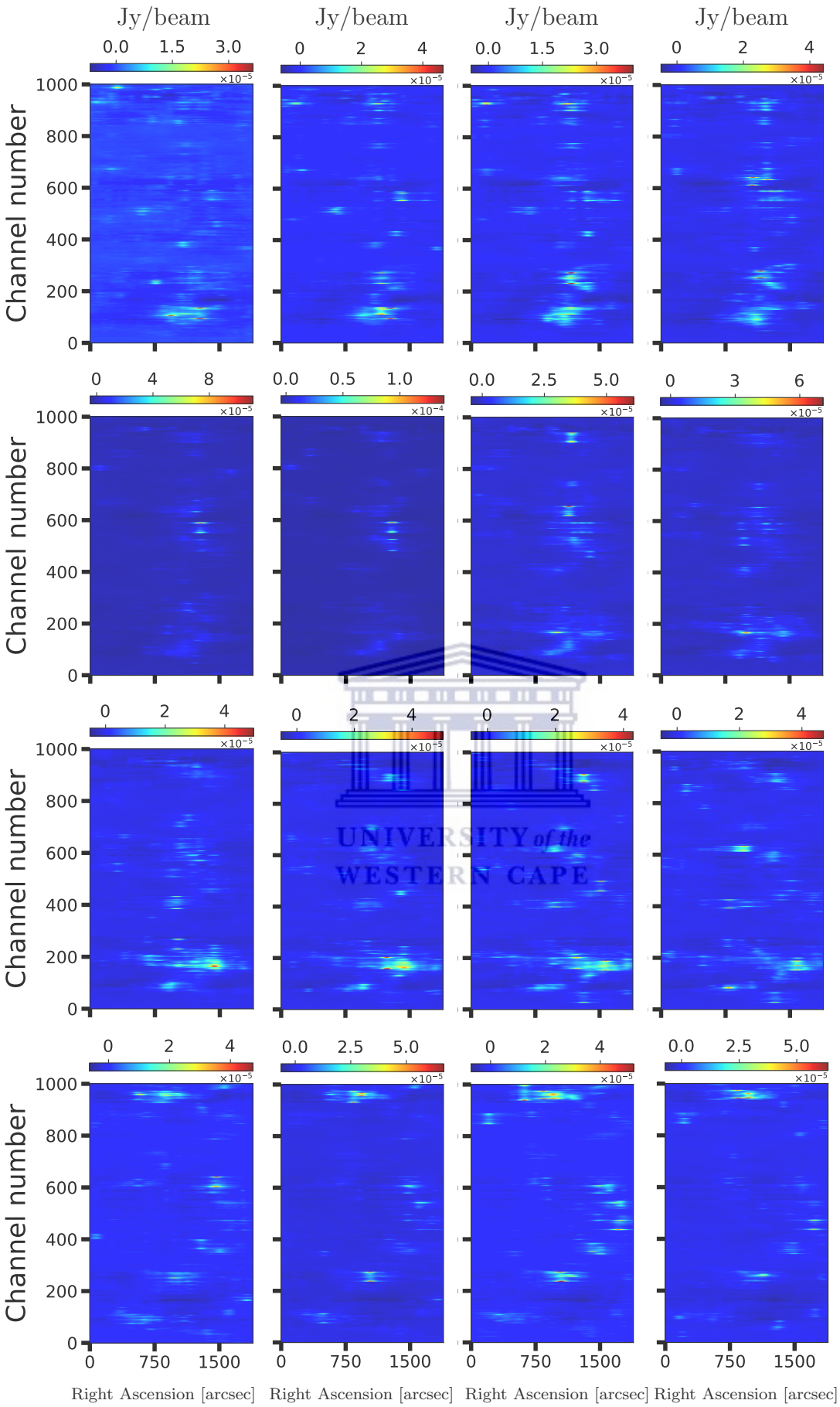


Figure 4.8: P-V diagrams for the VLA-D CASA data cube. Notation is similar to that in Figure 4.5.

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

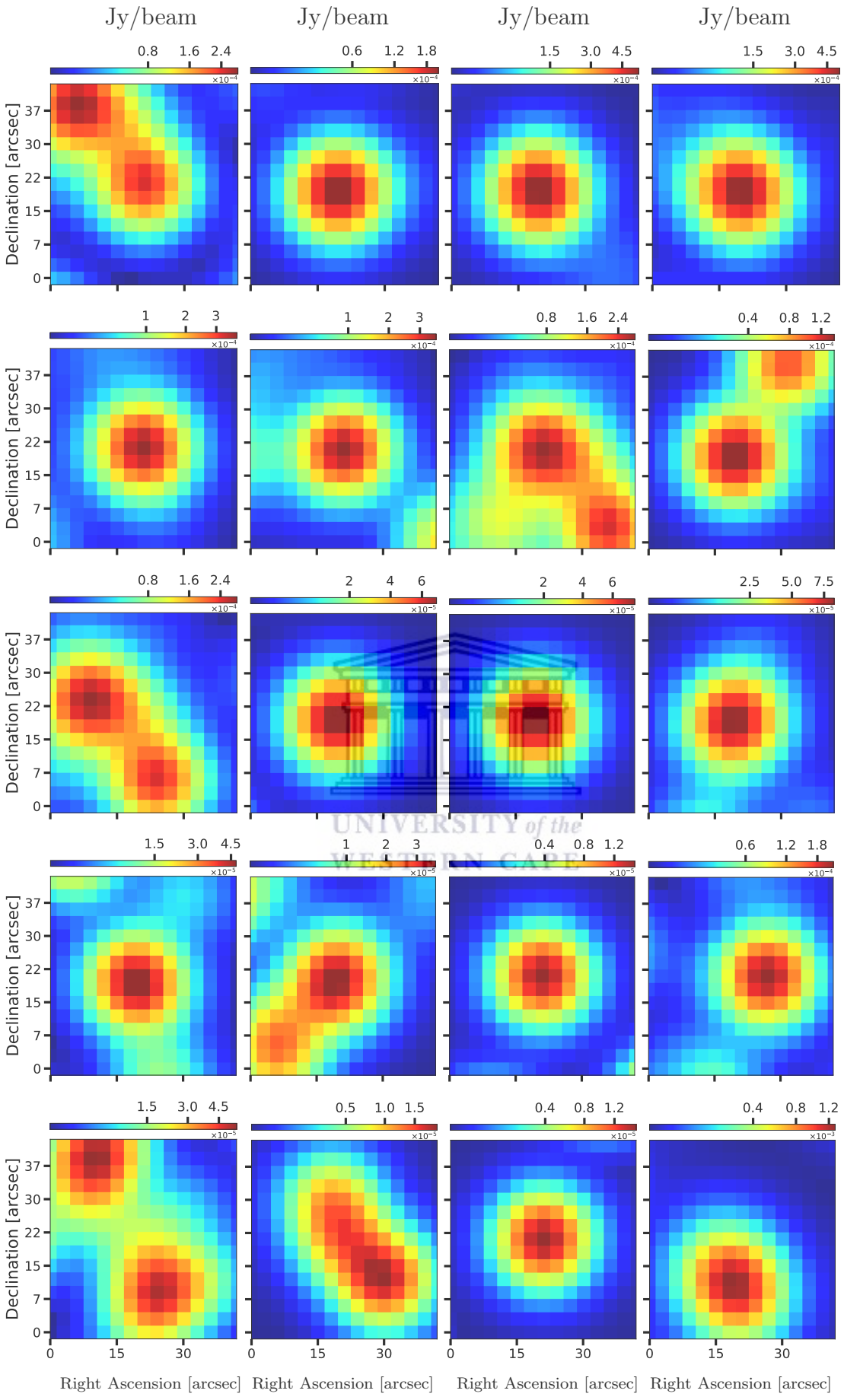


Figure 4.9: Total intensity maps for some of the massive ($M_* > 10^{10}\ M_\odot$) galaxies in the E16 cube. The colour bars show the flux density in units of Jy/beam.

https://etd.uwc.ac.za

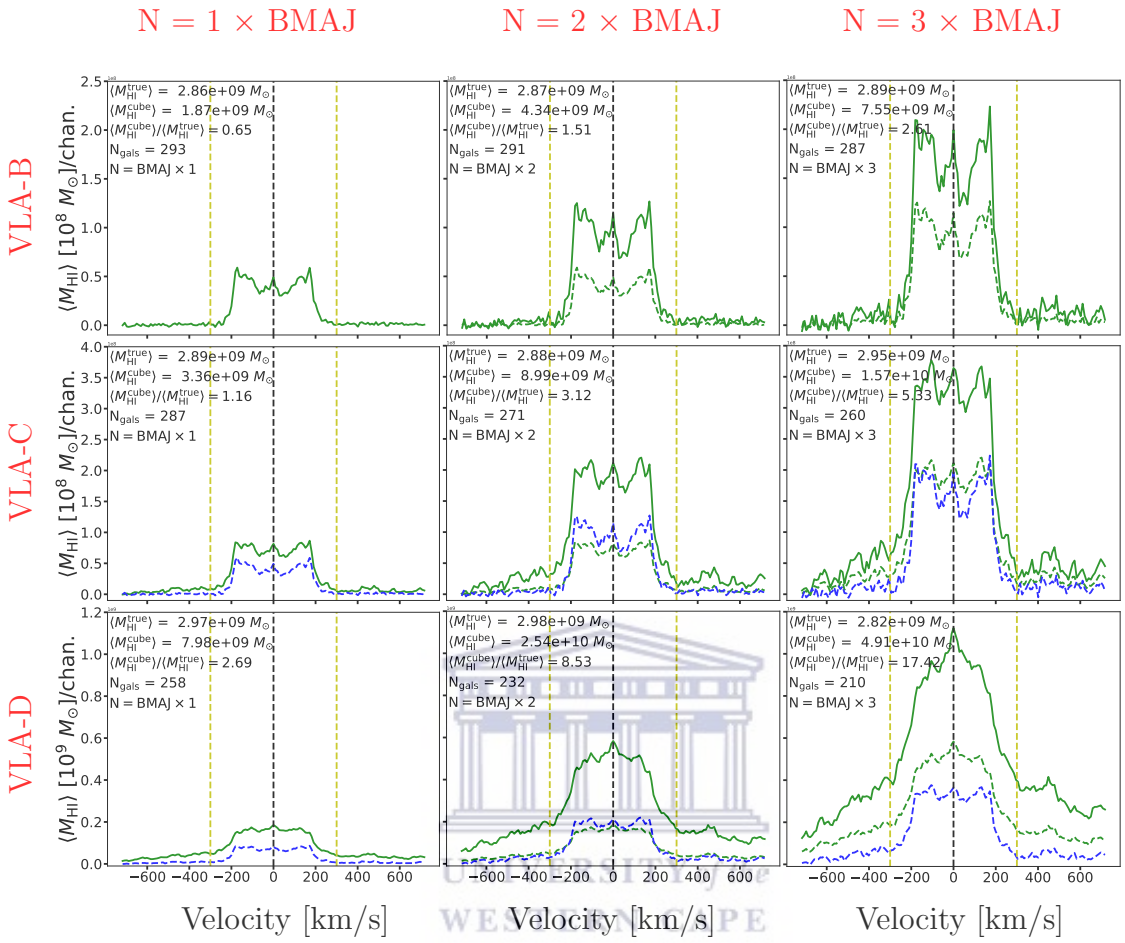


Figure 4.10: Co-added HI spectra for each of the simulated cubes based on the various VLA configurations (top to bottom). From left to right, spatial apertures of sizes N = 1, 2, 3 × BMAJ were used to extract individual galaxy spectra. Solid green spectra represent the stacked spectra based on the specified spatial aperture size. In each of the panels in columns 2 and 3, the green-dotted spectrum is a copy of the solid green spectrum in the left-adjacent panel. For each panel in rows 2 and 3, the blue-dashed spectrum is a copy of the solid green spectrum from the panel immediately above. $\langle M_{\rm HI}^{\rm true}\rangle$, $\langle M_{\rm HI}^{\rm cube}\rangle$ and $\rm N_{gals}$ are defined the same way as in Figure 3.4. $\langle M_{\rm HI}^{\rm cube}\rangle/\langle M_{\rm HI}^{\rm true}\rangle$ shows the ratio of the mass extracted from the CASA cube to the true HI mass obtained from the O&M14 catalogue and it determines the amount of additional mass for each spectrum. The vertical yellow-dashed lines denote the velocity range ± 300 km·s^{-1}. This is the range over which the stacked spectra were integrated in order to determine the average stacked mass of a galaxy.

https://etd.uwc.ac.za

Chapter 5

Discussions

The focus of this chapter is on discussing the extent to which results from HI stacking experiments may be affected by the inclusion of additional flux in a co-added spectrum. The inclusion of this additional flux is a result of source crowding and the spatial redistribution of flux by the point spread function (PSF). For the first time, the latter effect is quantified in this thesis.

UNIVERSITY of the
WESTERN CAPE

5.1 Additional flux in HI co-added spectra

HI stacking experiments assist in determining the total HI content of galaxies at high redshifts. However, the total amount of flux in a co-added spectrum almost always significantly overestimates the total amount of flux from the target galaxies. A large fraction of this additional flux is due to source confusion. This effect is a result of source crowding in a data cube and it occurs when the extracted target galaxy spectrum is contaminated by flux from nearby neighbouring galaxies. Delhaize et al. (2013) made estimations of the contributions of confusion for a stack that was known to be heavily confused and their approach was based on the optical parameters (mainly positions) of the galaxies in the field. Their estimates for the Parkes observations for a redshift range of $0.04 < z < 0.13$ indicated that one galaxy is confused, on average, with seven others in the 2-DFGRS (Two Degree Field Galaxy Redshift Survey) optical sample. The high level of contamination in their results was due to the large angular size of the Parkes beam, which is 15.5

https://etd.uwc.ac.za

arcmin in diameter. The stacked sample of Delhaize et al. (2013) gave a total amount of confusion of $\sim 1.1\times 10^{10}\ h_{70}^{-2}\ M_{\odot}$ per stacked galaxy. Jones et al. (2015) took a different approach where they used the currently available HI correlation function (CF) of the ALFALFA and the measurements of the mean HI density to make predictions of the amount of contaminant flux that will be contained in a co-added spectrum. Before making predictions for upcoming surveys, they tested their model against that of Delhaize et al. (2013). Their model predicted the amount of confused mass to be $\sim 1.3\times 10^{10}$ $h_{70}^{-2}\ M_{\odot}$ per stacked galaxy for a Parkes stacking experiment when using a beam size of 15.5 arcmin. They did not incorporate the beam profile weighting assumed in Delhaize et al. (2013), which was stated as the prime reason for the slight difference in the total amount of confused mass that they obtained compared to that of Delhaize et al. (2013). Elson et al. (2016, hereafter E16) carried out mock stacking experiments on their set of synthetic HI data cubes to quantify the rates of source confusion in both low- and high-redshift galaxy samples. A crucial application of their data cubes is for calculating, with a high degree of accuracy, the amount of contamination present in any stacked spectrum. They produced a suite of simulated cubes with resolutions that are representative of the current and upcoming high-redshift galaxy surveys. For an HI stacking experiment in the redshift range $0.04 < z < 0.13$, they calculated an average contaminant mass per galaxy of $\sim 1.5\times 10^{10}\ h_{70}^{-2}\ M_{\odot}$ per stacked galaxy (for a spatial resolution of ~ 15 arcmin), which is very similar to the estimate from Jones et al. (2015).

UNIVERSITY of the
WESTERN CAPE

As already indicated, source confusion plays a crucial role in HI stacking experiments and its effects must be taken into consideration. Although neighbouring galaxies may contribute a lot of additional flux in addition to the one extracted from the sub-volume containing the target, confusion can be regarded as a local effect since it is dependent only on nearby neighbouring galaxies. The co-adds obtained in this work can be used to measure the level of source confusion for co-adds based on the various VLA configurations. Moreover, because data cubes with realistic PSFs were used in this work, the co-adds that were obtained also contained additional flux due to PSF contamination. Therefore, in the

https://etd.uwc.ac.za

sections that follow, a discussion on the effects of PSF contamination and the methods of decomposing the two additional flux components in the co-adds will be given.
Before decomposing the results into the two aforementioned flux components, a discussion on the total amount of mass in each co-added spectrum is given in Section 5.2.

5.2 Total co-added mass

Section 2.2 discussed the various methods that were applied when creating the CASA-generated cubes. For each data cube produced (corresponding to a specific VLA configuration), co-adds were generated using the methods discussed in Section 3.1. The spatial apertures used had sizes of N = 1, 2 and 3 × BMAJ – where BMAJ is the major axis extent of the central lobe of the PSF. Elson et al. (2019, hereafter E19) model their galaxies to have maximum rotation speeds of less than 300 km·s^{-1}, and thus the target galaxy mass is expected to be contained within a velocity range of ± 300 km·s^{-1}. When calculating the total HI mass from the co-added spectra in this work, only the mass in the ± 300 km·s^{-1} velocity range was considered.
Shown in Figure 5.1 are the co-added spectra based on the CASA-generated cubes. Since these cubes were created using realistic PSFs, the additional mass in each spectrum is expected to consist of both source confusion and PSF contamination components.

UNIVERSITY of the
WESTERN CAPE

Essentially, almost all of the co-adds in Figure 5.1 overestimate the true average galaxy mass. For example, the typical average galaxy mass measured for the VLA-B cube co-add is 4.34×10^9 $M_\odot$ whereas the evaluated (known) average mass is 2.87×10^9 $M_\odot$. For the VLA-C cube co-add, the measured average galaxy mass is 8.99×10^9 $M_\odot$ whereas the evaluated average mass is 2.88×10^9 $M_\odot$. Finally, the average galaxy mass for the VLA-D cube co-add is 2.54×10^{10} $M_\odot$ whereas the evaluated average mass is 2.98×10^9 $M_\odot$. The average co-added and evaluated galaxy masses for the VLA-B, VLA-C and VLA-D cubes are shown in columns 2 and 3 of Tables 5.1 to 5.3 presented in Section 5.4. The obtained results show that the total amounts of additional mass from the spectra of the VLA-B, VLA-C and VLA-D cubes are 1.47×10^9 $M_\odot$, 6.11×10^9 $M_\odot$ and 22.42×10^9 $M_\odot$,

https://etd.uwc.ac.za

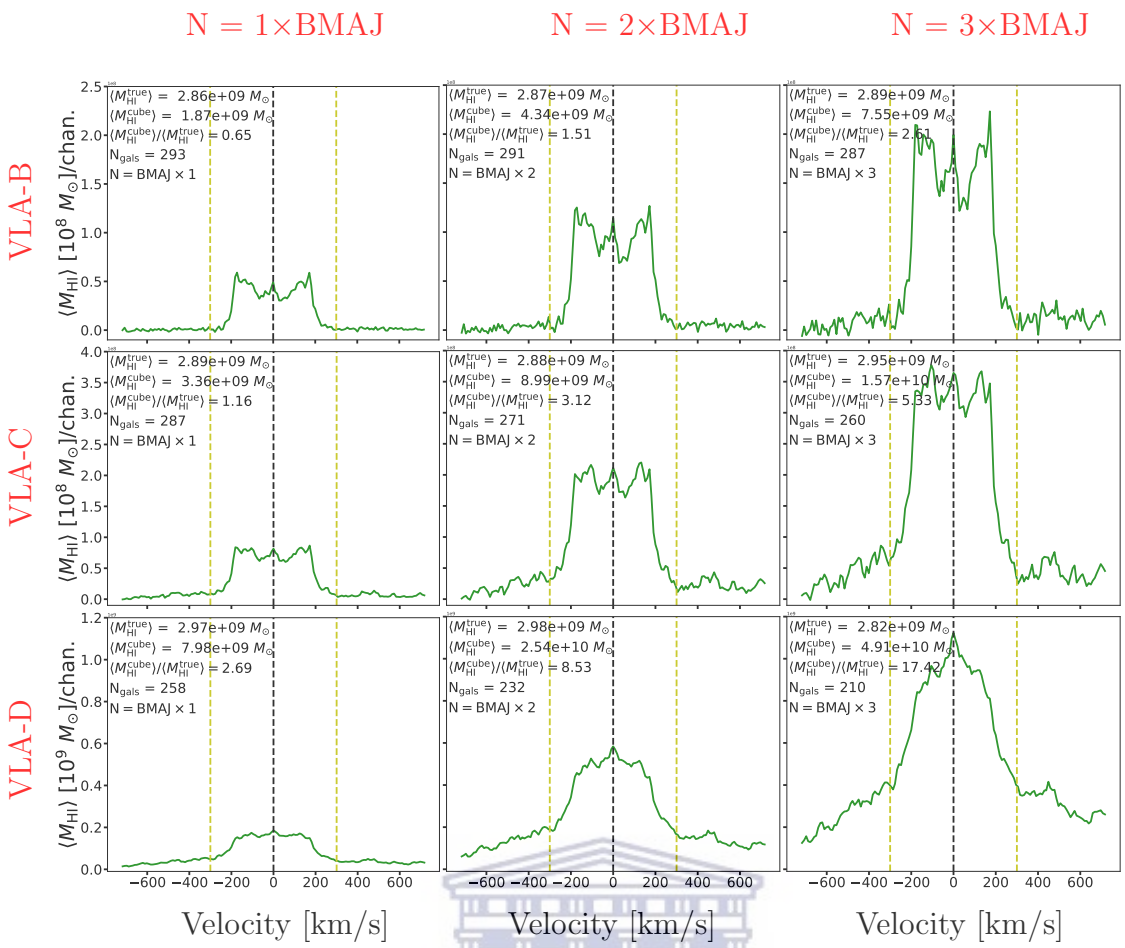


Figure 5.1: Co-added HI spectra for the CASA-generated cubes based on the various VLA configurations (top to bottom). From left to right, spatial apertures of sizes N = 1, 2, 3 × BMAJ were used to extract individual galaxy spectra. $\langle M_{\rm HI}^{\rm true}\rangle$ indicates the true (evaluated) average galaxy HI mass from the semi-analytic catalogue of O&M14 whereas $\langle M_{\rm HI}^{\rm cube}\rangle$ indicates the HI mass obtained from the co-added spectra obtained in this work. $\langle M_{\rm HI}^{\rm cube}\rangle/\langle M_{\rm HI}^{\rm true}\rangle$ shows the ratio of the mass extracted in the cube to the true HI mass and it determines the amount of additional mass contained in each spectrum. $N_{\rm gals}$ represents the total number of galaxies stacked. The vertical yellow-dashed lines delimit the velocity range ± 300 km·s^{-1}. This is the range over which the stacked spectra were integrated in order to determine the average stacked mass of a galaxy.

respectively. This further indicates that the total amount of additional mass for a typical VLA-B, VLA-C and VLA-D cube co-add is ∼ 1.51, ∼ 3.12 and ∼ 8.53 times more than the evaluated average galaxy mass from each of the respective cubes. These results are shown in column 4 of Tables 5.1 to 5.3.

For each of the co-adds presented in Figure 5.1, Figure 5.2 shows (as crosses) the fractional mass (defined as $\langle M_{\rm HI}^{\rm cube}\rangle/\langle M_{\rm HI}^{\rm true}\rangle$) as a function of spatial aperture size (N) for the CASA-

https://etd.uwc.ac.za

generated cubes.

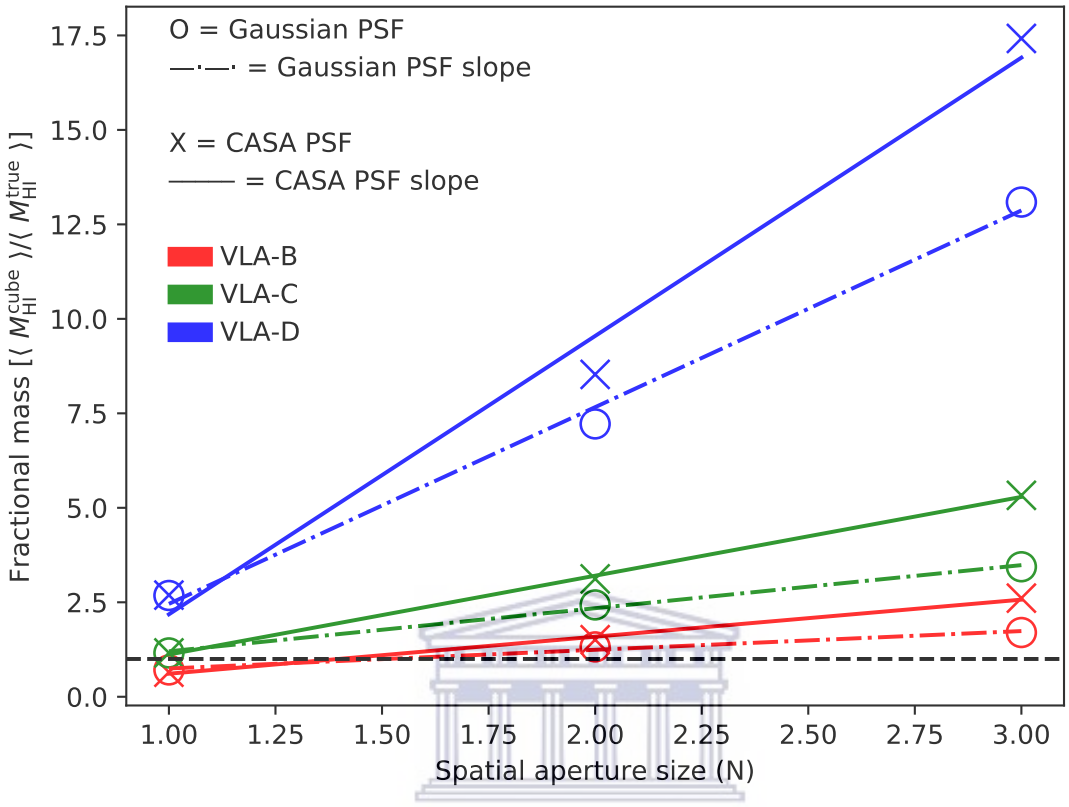


Figure 5.2: Fractional mass as a function of spatial aperture size for the Gaussian-smoothed and the CASA cubes. The circles represent the results based on the Gaussian PSF and the dashed-dotted lines represent the slope for those results. The crosses represent the results based on the CASA PSF and the solid lines represent the slope for those results. The red, green and blue colours represent the VLA-B, VLA-C and VLA-D configurations, respectively. The horizontal dashed line indicates the point where the fractional mass value is equal to 1. The amount of additional mass in a co-add increases approximately linearly due to confusion and PSF contamination.

Clearly evident in Figure 5.2 is the way in which the amount of additional mass in a co-add – relative to the (known) true total mass – increases approximately linearly with N. The rate at which this additional mass increases with N is related to the antenna configuration such that the more compact a configuration is, the more rapid the additional mass increases with N.

https://etd.uwc.ac.za

5.3 Additional flux due to source confusion

As mentioned in Section 5.1, various investigators have tried to estimate the amount of additional flux in a co-added spectrum due to source confusion. For the co-adds based on the CASA cubes generated in this work (shown in Figure 5.1), the additional flux has both source confusion and PSF contamination components. To determine the absolute contribution from source confusion, a corresponding set of cubes were produced by smoothing the Elson et al. (2016) unsmoothed cube using a simple Gaussian approximation of the central lobe of each CASA cube PSF. Smoothing with these Gaussian kernels produced cubes with the same source confusion characteristics as the CASA cubes, yet with zero redistribution of flux due to PSF side lobes. Hence, the amount of additional flux in the resulting co-adds were due to only source confusion. By applying the stacking methods on the various smoothed cubes generated in this work, the solid green spectra shown in Figure 5.3 were obtained.

The total amount of HI mass obtained due to source confusion obtained in this work is shown in column 5 of Tables 5.1 to 5.3. The results obtained have shown that for a typical VLA-B, VLA-C and VLA-D CASA cube co-added HI spectra, the total amount of mass due to source confusion is 0.93×10^9 $M_\odot$, 4.19×10^9 $M_\odot$ and 18.53×10^9 $M_\odot$, respectively. This further indicates that the total amount of additional mass due to source confusion for a typical VLA-B, VLA-C and VLA-D cube co-add is ~ 0.32, ~ 1.45 and ~ 6.22 times more than the evaluated average galaxy mass from each of the respective cubes.

The results presented in Figure 5.2 show – as circles – the results for the fractional mass as a function of spatial aperture size for the smoothed cube results presented as the green spectra in Figure 5.3. Due to the absence of PSF effects in the smoothed cubes, the amount of additional mass for each of these cubes is less than that of the CASA-generated cubes (presented as the blue-dashed spectra in Figure 5.3).

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

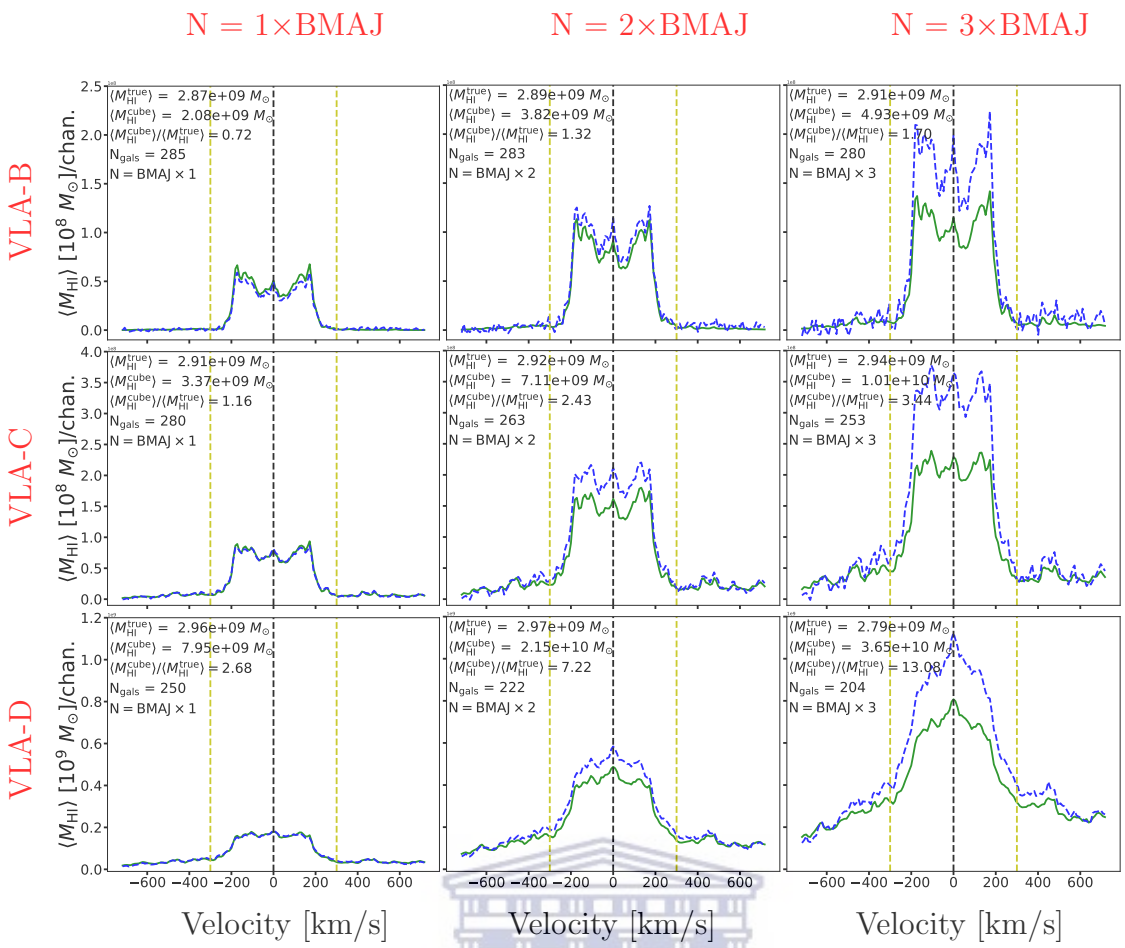


Figure 5.3: Co-added HI spectra for the smoothed cubes (solid green spectra) based on the beam parameters of the various VLA CASA cubes (top to bottom). From left to right, spatial apertures of sizes N = 1, 2, 3 ×BMAJ were used to extract individual galaxy spectra. The blue-dashed spectra show the corresponding results for the CASA-generated cubes shown in Figure 5.1. Notation is similar to the one in Figure 5.1.

5.4 Additional flux due to PSF contamination

As indicated in Section 5.1, source confusion is regarded as a local effect and it is determined by the typical number of sources contained within the beam area. The second source of additional flux in a co-added spectrum is due to an effect called PSF contamination. The effect arises due to the manner in which a PSF with large side lobes and spatially extended features (i.e., spikes) can redistribute the flux of a point source throughout the entire spatial domain of a cube. PSF contamination is thus a global effect in the sense that the flux from two sources that are completely separated physically – which are therefore not nearby neighbours to each other – can have their flux mixed. For the first time, this work has quantified the amount of additional flux due to PSF contamination

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

in a set of co-added spectra.

Before discussing the PSF contamination components of the additional flux in the obtained co-added spectra, an artificial two-dimensional demonstration of the PSF contamination effect (shown in Figure 5.4) is provided. To do this, an empty 2-D array with dimensions of 600 × 600 pixels (which was treated as having a pixel scale of 3″/pixel) was first created. Values of 1 were then assigned to ∼ 500 equally-spaced pixels that were separated by 27 pixels from each other. The separation of 27 pixels corresponds to 81 arcsec. The array was first spatially-smoothed using a Gaussian convolution kernel of size equal to the central lobe of the VLA-B PSF (10.00 arcsec × 9.89 arcsec). Because the angular separation of the non-zero points in the array is 81 arcsec (which is larger than the extent of the convolution kernel), the sources in the smoothed image remain spatially-unconfused (Figure 5.4, left panel). The empty 2-D array was then convolved with the full PSF from the VLA-B CASA cube. The resulting smoothed image thus had sources that were spatially-unconfused, yet which were PSF-contaminated (Figure 5.4, right panel). The two aforementioned smoothed arrays had the same total amount of flux with a percentage difference of ∼ 1.4 %. The presentation of this experiment – in Figure 5.4 – shows 9 sources for the two different cases mentioned above.

UNIVERSITY of the
WESTERN CAPE

Each of the 9 sources (shown in both panels presented in Figure 5.4) spans 13 × 13 pixels and for this reason, an aperture size of 13 pixels was used to extract the total flux for each source. For the case indicated on the left in Figure 5.4, the total flux extracted for a single source produced a value equal to 1.00 arb. units. For the case indicated on the right in Figure 5.4, the total flux for a single source extracted produced a much higher value of ∼ 3.06 arb. units due to PSF effects. Furthermore, for the entire regions indicated in Figure 5.4, the total flux for the whole area spanned in the left panel produced a total amount equal to 9.00 arb. units, whereas the area spanned by the right panel produced a total flux of ∼ 11.80 arb. units. Such a demonstration clearly shows that galaxies that are much further apart, where spatial confusion plays no role, still get significant

https://etd.uwc.ac.za

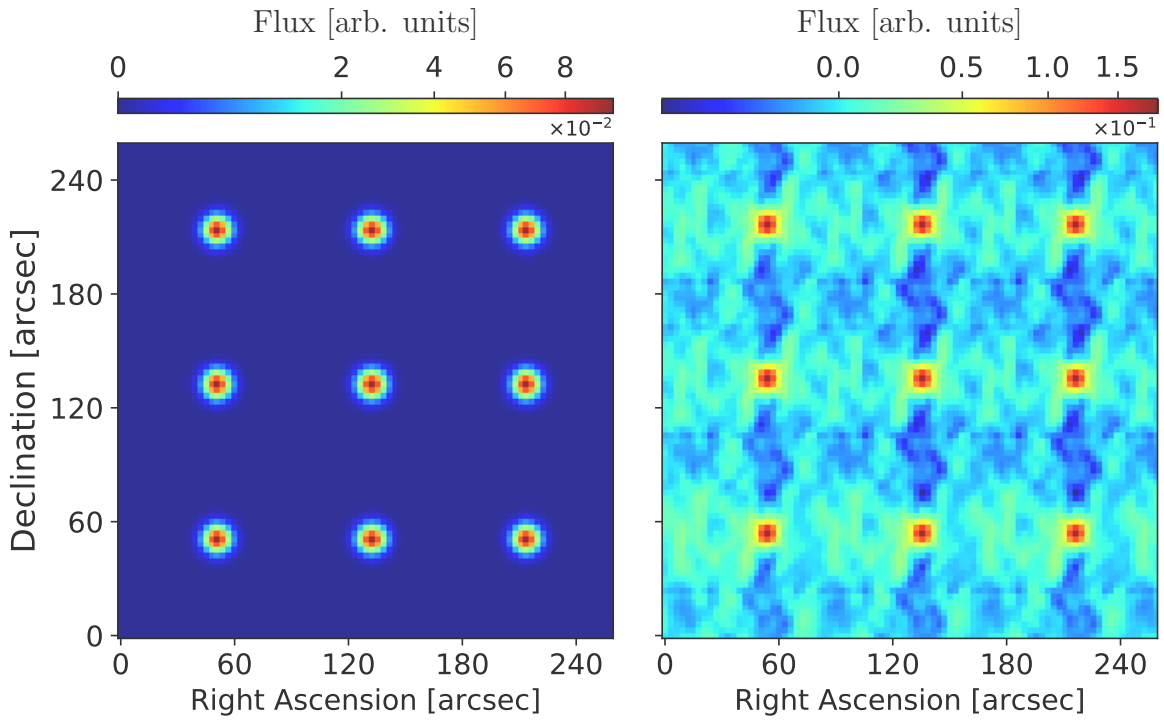


Figure 5.4: Illustration showing a 2-D array convolved with a Gaussian beam (left) and one convolved with a realistic PSF (right). In the two cases presented, the beam parameters of the VLA-B CASA cube were used. Clearly visible in the diagram is the way in which PSF effects (right panel), from sources much further away, contribute towards the total flux in regions where only the target source flux is supposed to be spanned.

additional flux contributions due to PSFs. This clearly shows the significance of realistic PSF effects and it highlights the necessity to consider such effects when carrying out HI stacking experiments.

Using the results from the Gaussian-smoothed cubes (presented in Section 5.3), calculations for the amount of additional flux due to PSF contamination in the co-adds from the CASA cubes were able to be made. This was the case since the results from the CASA cubes (shown in Figure 5.1) were due to both source confusion and PSF contamination. Knowing the contributions of source confusion from the Gaussian-smoothed cubes, this flux could then be subtracted from the flux obtained from the CASA cubes and the remaining amount of additional flux would be due to PSFs.

To clearly distinguish between the results obtained in Figure 5.3, Tables 5.1 to 5.3 were produced.

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

N	$\langle M_{\rm HI}^{\rm cube}\rangle$ $[10^9 M_\odot]$	$\langle M_{\rm HI}^{\rm true}\rangle$ $[10^9 M_\odot]$	$\langle M_{\rm HI}^{\rm cube}\rangle / \langle M_{\rm HI}^{\rm true}\rangle$	$\langle M_{\rm HI}^{\rm conf.}\rangle$ $[10^9 M_\odot]$	$\langle M_{\rm HI}^{\rm PSF}\rangle$ $[10^9 M_\odot]$
1 × BMAJ	1.87	2.86	0.65	0	0
2 × BMAJ	4.34	2.87	1.51	0.93	0.54
3 × BMAJ	7.55	2.89	2.61	2.02	2.64

Table 5.1: VLA-B CASA cube results. Column 1 shows the spatial aperture size used to extract individual galaxy spectra. Column 2 shows the average HI mass from the co-added spectra. Column 3 shows the average evaluated HI mass from the O&M14 catalogue. Column 4 shows the ratio of the mass extracted in the cube to the true HI mass – which indicates the extent to which a co-add under- or over-estimates the true average HI mass. Column 5 shows the HI mass due to source confusion from the co-added spectra. Column 6 shows the HI mass due to PSF contamination from the co-added spectra. The results that generated the results in columns 5 and 6 are discussed in Section 5.3.

N	$\langle M_{\rm HI}^{\rm cube}\rangle$ $[10^9 M_\odot]$	$\langle M_{\rm HI}^{\rm true}\rangle$ $[10^9 M_\odot]$	$\langle M_{\rm HI}^{\rm cube}\rangle / \langle M_{\rm HI}^{\rm true}\rangle$	$\langle M_{\rm HI}^{\rm conf.}\rangle$ $[10^9 M_\odot]$	$\langle M_{\rm HI}^{\rm PSF}\rangle$ $[10^9 M_\odot]$
1 × BMAJ	3.36	2.89	1.16	0.46	0.01
2 × BMAJ	8.99	2.88	3.12	4.19	1.92
3 × BMAJ	15.7	2.95	5.33	7.16	5.59

Table 5.2: VLA-C CASA cube results. Notation is similar to that in Table 5.1.

UNIVERSITY of the WESTERN CAPE

N	$\langle M_{\rm HI}^{\rm cube}\rangle$ $[10^9 M_\odot]$	$\langle M_{\rm HI}^{\rm true}\rangle$ $[10^9 M_\odot]$	$\langle M_{\rm HI}^{\rm cube}\rangle / \langle M_{\rm HI}^{\rm true}\rangle$	$\langle M_{\rm HI}^{\rm conf.}\rangle$ $[10^9 M_\odot]$	$\langle M_{\rm HI}^{\rm PSF}\rangle$ $[10^9 M_\odot]$
1 × BMAJ	7.98	2.97	2.69	4.99	0.02
2 × BMAJ	25.4	2.98	8.53	18.53	3.89
3 × BMAJ	49.1	2.82	17.42	33.71	12.57

Table 5.3: VLA-D CASA cube results. Notation is similar to that in Table 5.1.

For almost all of the VLA B, C and D co-adds, the average galaxy HI mass calculated from the co-added spectra (column 2 of Tables 5.1 to 5.3) overestimates the true average co-added HI mass (column 3 of Tables 5.1 to 5.3). The amount of additional HI mass that was obtained is strongly linked to the spatial aperture size and the angular resolution of the various VLA configurations. The obtained results have shown that for a typical VLA-B, VLA-C and VLA-D cube co-added HI spectra, the typical amount of mass due to PSF contamination (column 6 of Tables 5.1 to 5.3) is 0.54×10^9 $M_\odot$, 1.92×10^9 $M_\odot$ and

https://etd.uwc.ac.za

3.89×10^9 $M_\odot$, respectively. This further indicates that the total amount of additional mass due to PSF contamination for a typical VLA-B, VLA-C and VLA-D cube co-add is ~ 0.19, ~ 0.67 and ~ 1.31 times the evaluated average galaxy mass from each of the respective cubes.

The obtained results (shown in Tables 5.1 to 5.3) have demonstrated the effects of realistic PSFs on HI stacking experiments. For almost all of the stacks that were produced, PSF effects contribute significantly to the additional mass that overestimates the true (expected) HI mass. Such a result indicates the need to consider these effects in forthcoming HI stacking experiments in order to minimize the uncertainties in the results. Upon applying all the correction factors needed (including corrections for PSF contamination effects), future HI stacking experiments will yield efficient and desirable results. This will further allow for the extraction of the most accurate cosmological parameters from these experiments.

https://etd.uwc.ac.za

Chapter 6

Summary and future outlooks

This chapter provides a brief overview of this work and outlines the future outlooks on HI stacking experiments. A further indication is given on the usefulness of the results obtained in this work on the upcoming HI stacking experiments.

UNIVERSITY of the
WESTERN CAPE

6.1 Summary

The first objective in this work was to gain access to various simulated data products that could be used to extract galaxy properties and enable for the study of neutral atomic hydrogen (HI) stacking experiments. This was achieved since access was gained to the semi-analytic models of Obreschkow et al. (2009a,b,c) which present a simulation of the evolution of molecular (H_2) and atomic (HI) phases of cold hydrogen gas in 3×10^7 galaxies – based on the Millennium simulation. These simulations were used by Obreschkow and Meyer (2014) to produce a mock galaxy catalogue. Elson et al. (2016) further went on to take this catalogue and used it to produce synthetic HI data cubes, which play a significant role in this work.

Furthermore, this work aimed to explore the uncertainties in the results from HI stacking experiments due to the inclusion of additional flux from source confusion (the effect that occurs when the extracted target galaxy spectrum is contaminated by flux from nearby neighbouring galaxies) and point spread function (PSF) contamination (the effect that

https://etd.uwc.ac.za

arises due to the manner in which a PSF with large side lobes and spatially extended features, i.e., spikes, can redistribute the flux of a point source throughout the entire spatial domain of a cube), respectively. To explore this, this work intended on introducing more realism to existing data cubes using simulation methods. This was achieved by passing the Elson et al. (2016) cube through the Common Astronomy Software Applications (CASA) software to incorporate PSF effects linked to a specific baseline distribution. These methods produced data cubes containing more realism based on the configurations of the VLA in its B, C and D configurations.

One of the key objectives was to write stacking codes in Python and to further verify their correctness. A stacking software was successfully implemented and its accuracy was tested by applying it on the Elson et al. (2019) data cube and making comparisons of the results from this work with those from Elson et al. (2019). The stacking methods make use of positions from the Obreschkow and Meyer (2014) catalogue together with the user-specified spatial and spectral information to extract individual galaxy spectra from a data cube. The obtained individual galaxy spectra are then stacked to yield the final average HI spectrum representative of the whole sample.

UNIVERSITY of the
WESTERN CAPE

Upon confirming the correctness of the stacking methods generated in this work, they were then applied on the CASA-generated cubes to understand how realistic observing effects, e.g., baseline distribution, interferometric noise and resolution, affect the results from HI stacking experiments.

Further aims for this work were to produce a set of data cubes with no PSF effects. These cubes were produced by smoothing the Elson et al. (2016) unsmoothed data cube using Gaussian beams equal in size to the central lobes of the PSFs from the CASA-generated cubes. These cubes were representative of the perfect case such that they contained the same source confusion characteristics as the CASA cubes, yet with zero redistribution of flux due to PSF side lobes. Comparing the stacked results for the Gaussian-smoothed and the CASA-generated cubes allowed for the proper quantification of the additional

https://etd.uwc.ac.za

flux contributions that are solely due to PSFs.

The results obtained in this work indicated that for a typical VLA-B, VLA-C and VLA-D CASA cube co-added HI spectrum, the total amount of mass due to source confusion is 0.93×10^9 $M_\odot$, 4.19×10^9 $M_\odot$ and 18.53×10^9 $M_\odot$, respectively. This further indicated that the total amount of additional mass due to source confusion for a typical VLA-B, VLA-C and VLA-D cube co-add is ~ 0.32, ~ 1.45 and ~ 6.22 times more than the evaluated average galaxy mass from each of the respective cubes.

This study further showed that for a typical VLA-B, VLA-C and VLA-D cube co-added HI spectra, the typical amount of mass due to PSF contamination is 0.54×10^9 $M_\odot$, 1.92×10^9 $M_\odot$ and 3.89×10^9 $M_\odot$, respectively. This further indicated that the total amount of additional mass due to PSF contamination for a typical VLA-B, VLA-C and VLA-D cube co-add is ~ 0.19, ~ 0.67 and ~ 1.31 times the evaluated average galaxy mass from each of the respective cubes.

A further study was made on the variation of additional mass in a stacked spectrum as a function of changing spatial aperture size (N) for both the perfect and the realistic cubes. A key result noted was that the amount of additional mass in a co-add, relative to the true (known) total mass, increases approximately linearly with N.

UNIVERSITY of the
WESTERN CAPE

All of the aims for this project were achieved and this study was able to quantify, for the first time, the effects of PSFs on HI stacking experiments. This highlighted the significance of considering PSFs when carrying out these experiments. Section 6.2 mentions some of the crucial upcoming surveys that will assist in determining the HI content of galaxies out to high redshifts. The section is further devoted to indicating the importance of the results obtained in this work on future surveys.

6.2 Future outlooks

This section focuses on some of the major upcoming HI surveys and further emphasis is given on the importance of considering additional flux components on these surveys.

https://etd.uwc.ac.za

This section further highlights the importance of the results obtained in this work and the way in which they can be beneficial to the upcoming surveys. First, a discussion on the surveys that will be carried out with the MeerKAT array is given, in subsection 6.2.1. A discussion is then given, in subsection 6.2.2, on the surveys that will be carried out with the ASKAP array. Finally, an overview of this section is given in subsection 6.2.3.

6.2.1 MeerKAT surveys

The first survey considered in this subsection is the LADUMA (Looking At the Distant Universe with MeerKAT Array, Baker et al., 2018) survey, which is one of the key science experiments aimed at making HI detections at high redshifts out to $z = 1$ and beyond (Baker et al., 2018). The survey intends on integrating a single pointing for 5000 hours with the MeerKAT array. The total field of view of the survey corresponds to the beam of a single dish, which is 0.9 deg^2 at $z = 0$. Elson et al. (2016) indicate that HI stacking experiments from surveys such as LADUMA, with spatial resolutions of 10 – 30 arcsec, will not be dominated by source confusion out to redshift $z \approx 0.7$. However, since LADUMA plans on stacking galaxy spectra out to redshifts beyond $z = 1$, these results will be affected significantly by source confusion. Moreover, probing HI at higher redshifts leads to the physical size of the beam growing with redshift and thus enclosing more volume, leading to a large number of objects being contained in the observational beamwidth. Having a large number of objects contained within the observational beamwidth thus indicates that the effects of additional flux will be more prominent when carrying out HI stacking experiments. Elson et al. (2016) further went on to indicate that precise correction factors will need to be applied in order to obtain reliable measurements of the total HI content of galaxies at high redshifts. In addition to this, the additional PSF effects that were investigated in this study will also contribute significantly to the upcoming HI stacking experiments that will be carried out by surveys such as LADUMA. To improve the accuracy of the results obtained from such surveys, proper correction factors for PSFs will also be needed.

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

Another significant survey to be carried out with the MeerKAT telescope is the MIGHTEE (MeerKAT International GHz Tiered Extragalactic Exploration, Jarvis et al., 2017) survey. MIGHTEE will observe deep well-studied fields of the extragalactic sky that are visible from the southern hemisphere – covering a total sky area of 20 deg^2. The survey will be conducted by creating deep images of the Giga-Hertz radio emission in polarisation, continuum and spectral line. Its sensitivity will be ~ 1 μJy for a resolution of $\sim$ 6 arcsec over a full bandwidth of 900 – 1670 MHz. MIGHTEE is over a wider area to LADUMA, and hence will have greater statistics. One of the key objectives of MIGHTEE is to improve an understanding of the evolution of HI in the Universe and to study how this HI turns into stars after having gone through the molecular phase. The survey also intends on studying the evolution of active galactic nuclei (AGN) and star formation over cosmic time. In studying the cosmic evolution of HI, MIGHTEE will form a synergy with LADUMA, where MIGHTEE will conduct observations out to intermediate redshifts ($z \sim 0.5$) while LADUMA will observe at high redshifts ($z < 1.4$) that are inaccessible by MIGHTEE – making the two surveys complementary (e.g., Jarvis et al., 2017). Combining the HI detections from both these surveys will enable for the measurement of the low- and high-mass ends of the HIMF. As already indicated, results from the LADUMA survey will be affected by PSF contamination. Like LADUMA, the MIGHTEE survey will be affected by PSF effects, and the need for correction factors should also be considered for this survey.

UNIVERSITY of the
WESTERN CAPE

6.2.2 ASKAP surveys

This subsection focuses on the two proposed HI emission-line surveys to be conducted using the Australian Square Kilometre Array Pathfinder (ASKAP), namely the WALLABY (Widefield ASKAP L-band Legacy All-sky Blind surveY, Duffy et al., 2012) and DINGO (Deep Investigation of Neutral Gas Origins, Duffy et al., 2012). WALLABY will be a shallow survey – covering the redshift range $0 < z < 0.26$ – which will probe over 6×10^5

https://etd.uwc.ac.za

galaxies. DINGO, on the other hand, will be a much deeper HI survey that will cover the redshift range $0 < z < 0.43$ – probing over 10^5 galaxies (Duffy et al., 2012). Both these surveys have the common aim of studying the cosmic evolution of HI. As indicated in Duffy et al. (2012), source confusion will not be an issue for WALLABY as it will be conducting observations at low redshifts. However, since the confusion rate increases with increasing redshift, DINGO will be highly affected by source confusion since it intends on probing much higher redshifts. Necessary correction factors will thus need to be applied to correct for the amount of confusion for the DINGO survey. Furthermore, as already indicated for LADUMA and MIGHTEE, corrections for PSF effects will also be a necessity for both the WALLABY and DINGO surveys. However, no such corrections should be necessary for galaxies that are directly detected because they will be CLEANed. But, for galaxies below the noise level, PSF effects could be significant.

6.2.3 Overview

UNIVERSITY of the
WESTERN CAPE

The surveys mentioned in subsections 6.2.1 and 6.2.2 will serve very useful in studying the cosmic evolution of HI and will thus improve an understanding of the HI content of galaxies up to redshift $z = 1$ and beyond. The effects of source confusion from several studies were explored and the importance of applying correction factors for these effects was highlighted. This work introduced more realism to existing HI data cubes by incorporating PSF effects and further highlighted – in subsections 6.2.1 and 6.2.2 – the significance of considering these effects on future HI stacking experiments. Having realized the contributions of PSFs on HI stacking experiments, it thus becomes crucial to introduce correction factors for this effect in both the current and upcoming HI surveys. Upon applying all the necessary corrections for both source confusion and PSF effects, the upcoming HI stacking experiments will be able to produce highly reliable results. Generally, detailed modelling will be required to properly quantify the uncertainties and this will allow for the extraction of the most accurate cosmological parameters from these experiments.

https://etd.uwc.ac.za

Bibliography

Baker, A. J., Blyth, S., Holwerda, B. W., Team, L., et al. (2018). In *American Astronomical Society Meeting Abstracts# 231*, volume 231.

Barnes, D. G., Staveley-Smith, L., de Blok, W. e. . a., Oosterloo, T., Stewart, I. M., Wright, A., Banks, G., Bhathal, R., Boyce, P., Calabretta, M., et al. (2001). *Monthly Notices of the Royal Astronomical Society*, 322(3):486–498.

Bouwens, R., Illingworth, G., Oesch, P., Franx, M., Labbé, I., Trenti, M., van Dokkum, P., Carollo, C. M., González, V., Smit, R., et al. (2012a). *The Astrophysical Journal*, 754(2):83.

Bouwens, R. J., Illingworth, G., Oesch, P., Trenti, M., Labbé, I., Franx, M., Stiavelli, M., Carollo, C. M., van Dokkum, P., and Magee, D. (2012b). *The Astrophysical Journal Letters*, 752(1):L5.

Braun, R. (2012). *The Astrophysical Journal*, 749(1):87.

Catinella, B., Giovanelli, R., and Haynes, M. P. (2006). *The Astrophysical Journal*, 640(2):751.

Crighton, N. H., Murphy, M. T., Prochaska, J. X., Worseck, G., Rafelski, M., Becker, G. D., Ellison, S. L., Fumagalli, M., Lopez, S., Meiksin, A., et al. (2015). *Monthly Notices of the Royal Astronomical Society*, 452(1):217–234.

Cucciati, O., Tresse, L., Ilbert, O., Le Fevre, O., Garilli, B., Le Brun, V., Cassata, P., Franzetti, P., Maccagni, D., Scodeggio, M., et al. (2012). *Astronomy & Astrophysics*, 539:A31.

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

Dahlen, T., Mobasher, B., Dickinson, M., Ferguson, H. C., Giavalisco, M., Kretchmer, C., and Ravindranath, S. (2007). *The Astrophysical Journal*, 654(1):172.

Davé, R., Rafieferantsoa, M. H., Thompson, R. J., and Hopkins, P. F. (2017). *Monthly Notices of the Royal Astronomical Society*, 467(1):115–132.

De Lucia, G. and Blaizot, J. (2007). *Monthly Notices of the Royal Astronomical Society*, 375(1):2–14.

Delhaize, J., Meyer, M., Staveley-Smith, L., and Boyle, B. (2013). *Monthly Notices of the Royal Astronomical Society*, 433(2):1398–1410.

Duffy, A. R., Meyer, M. J., Staveley-Smith, L., Bernyk, M., Croton, D. J., Koribalski, B. S., Gerstmann, D., and Westerlund, S. (2012). *Monthly Notices of the Royal Astronomical Society*, 426(4):3385–3402.

Elmegreen, B. (1993). *The Astrophysical Journal*, 411:170–177.

Elson, E., Baker, A., and Blyth, S. (2019). *Monthly Notices of the Royal Astronomical Society*, 486(4):4894–4903.

Elson, E., Blyth, S., and Baker, A. (2016). *Monthly Notices of the Royal Astronomical Society*, 460(4):4366–4381.

Fabello, S., Catinella, B., Giovanelli, R., Kauffmann, G., Haynes, M. P., Heckman, T. M., and Schiminovich, D. (2011). *Monthly Notices of the Royal Astronomical Society*, 411(2):993–1012.

Fernández, X., Gim, H. B., Van Gorkom, J., Yun, M. S., Momjian, E., Popping, A., Chomiuk, L., Hess, K. M., Hunt, L., Kreckel, K., et al. (2016). *The Astrophysical Journal Letters*, 824(1):L1.

Fernández, X., Van Gorkom, J., Hess, K. M., Pisano, D., Kreckel, K., Momjian, E., Popping, A., Oosterloo, T., Chomiuk, L., Verheijen, M., et al. (2013). *The Astrophysical Journal Letters*, 770(2):L29.

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

Geréb, K., Morganti, R., Oosterloo, T., Hoppmann, L., and Staveley-Smith, L. (2015). *Astronomy & Astrophysics*, 580:A43.

Giovanelli, R. and Haynes, M. P. (2002). *The Astrophysical Journal Letters*, 571(2):L107.

Giovanelli, R., Haynes, M. P., Kent, B. R., Perillat, P., Saintonge, A., Brosch, N., Catinella, B., Hoffman, G. L., Stierwalt, S., Spekkens, K., et al. (2005). *The astronomical journal*, 130(6):2598.

Gruppioni, C., Pozzi, F., Rodighiero, G., Delvecchio, I., Berta, S., Pozzetti, L., Zamorani, G., Andreani, P., Cimatti, A., Ilbert, O., et al. (2013). *Monthly Notices of the Royal Astronomical Society*, 432(1):23–52.

Hopkins, A. M., McClure-Griffiths, N., and Gaensler, B. (2008). *The Astrophysical Journal Letters*, 682(1):L13.

Hoppmann, L., Staveley-Smith, L., Freudling, W., Zwaan, M., Minchin, R., and Calabretta, M. (2015). *Monthly Notices of the Royal Astronomical Society*, 452(4):3726–3741.

UNIVERSITY of the
WESTERN CAPE

Hu, W., Hoppmann, L., Staveley-Smith, L., Geréb, K., Oosterloo, T., Morganti, R., Catinella, B., Cortese, L., Lagos, C. d. P., and Meyer, M. (2019). *Monthly Notices of the Royal Astronomical Society*, 489(2):1619–1632.

Jarvis, M. J., Taylor, A., Agudo, I., Allison, J. R., Deane, R., Frank, B., Gupta, N., Heywood, I., Maddox, N., McAlpine, K., et al. (2017). *arXiv preprint arXiv:1709.01901*.

Jones, M. G., Haynes, M. P., Giovanelli, R., and Moorman, C. (2018). *Monthly Notices of the Royal Astronomical Society*, 477(1):2–17.

Jones, M. G., Haynes, M. P., Giovanelli, R., and Papastergis, E. (2015). *Monthly Notices of the Royal Astronomical Society*, 455(2):1574–1583.

Koyama, H. and Ostriker, E. C. (2009). *The Astrophysical Journal*, 693(2):1346.

https://etd.uwc.ac.za

Lagos, C. d. P., Tobar, R. J., Robotham, A. S., Obreschkow, D., Mitchell, P. D., Power, C., and Elahi, P. J. (2018). *Monthly Notices of the Royal Astronomical Society*, 481(3):3573–3603.

Lah, P., Chengalur, J. N., Briggs, F. H., Colless, M., De Propris, R., Pracy, M. B., De Blok, W., Fujita, S. S., Ajiki, M., Shioya, Y., et al. (2007). *Monthly Notices of the Royal Astronomical Society*, 376(3):1357–1366.

Leroy, A. K., Walter, F., Brinks, E., Bigiel, F., De Blok, W., Madore, B., and Thornley, M. (2008). *The Astronomical Journal*, 136(6):2782.

Madau, P. and Dickinson, M. (2014). *Annual Review of Astronomy and Astrophysics*, 52:415–486.

Maddox, N., Hess, K. M., Blyth, S.-L., and Jarvis, M. (2013). *Monthly Notices of the Royal Astronomical Society*, 433(3):2613–2625.

Magnelli, B., Elbaz, D., Chary, R., Dickinson, M., Le Borgne, D., Frayer, D., and Willmer, C. (2011). *Astronomy & Astrophysics*, 528:A35.

Magnelli, B., Popesso, P., Berta, S., Pozzi, F., Elbaz, D., Lutz, D., Dickinson, M., Altieri, B., Andreani, P., Aussel, H., et al. (2013). *Astronomy & Astrophysics*, 553:A132.

McGaugh, S. S., Schombert, J. M., Bothun, G. D., and De Blok, W. (2000). *The Astrophysical Journal Letters*, 533(2):L99.

Obreschkow, D., Croton, D., De Lucia, G., Khochfar, S., and Rawlings, S. (2009a). *The Astrophysical Journal*, 698(2):1467.

Obreschkow, D., Heywood, I., Klöckner, H.-R., and Rawlings, S. (2009b). *The Astrophysical Journal*, 702(2):1321.

Obreschkow, D., Klöckner, H.-R., Heywood, I., Levrier, F., and Rawlings, S. (2009c). *The Astrophysical Journal*, 703(2):1890.

Obreschkow, D. and Meyer, M. (2014). *arXiv preprint arXiv:1406.0966*.

UNIVERSITY of the WESTERN CAPE

https://etd.uwc.ac.za

Obreschkow, D. and Rawlings, S. (2009). *Monthly Notices of the Royal Astronomical Society*, 394(4):1857–1874.

Prochaska, J. X., Herbert-Fort, S., and Wolfe, A. M. (2005). *The Astrophysical Journal*, 635(1):123.

Rao, S. M., Turnshek, D. A., and Nestor, D. B. (2006). *The Astrophysical Journal*, 636(2):610.

Rao, S. M., Turnshek, D. A., Sardane, G. M., and Monier, E. M. (2017). *Monthly Notices of the Royal Astronomical Society*, 471(3):3428–3442.

Reddy, N. A. and Steidel, C. C. (2009). *The Astrophysical Journal*, 692(1):778.

Rhee, J., Lah, P., Briggs, F. H., Chengalur, J. N., Colless, M., Willner, S. P., Ashby, M. L., and Le Fèvre, O. (2017). *Monthly Notices of the Royal Astronomical Society*, 473(2):1879–1894.

Rhee, J., Lah, P., Chengalur, J. N., Briggs, F. H., and Colless, M. (2016). *arXiv preprint arXiv:1605.02006*.

Rhee, J., Zwaan, M. A., Briggs, F. H., Chengalur, J. N., Lah, P., Oosterloo, T., and Hulst, T. v. d. (2013). *Monthly Notices of the Royal Astronomical Society*, 435(3):2693–2706.

Robotham, A. S. and Driver, S. P. (2011). *Monthly Notices of the Royal Astronomical Society*, 413(4):2570–2582.

Sanders, D., Mazzarella, J., Kim, D.-C., Surace, J., and Soifer, B. (2003). *The Astronomical Journal*, 126(4):1607.

Schenker, M. A., Robertson, B. E., Ellis, R. S., Ono, Y., McLure, R. J., Dunlop, J. S., Koekemoer, A., Bowler, R. A., Ouchi, M., Curtis-Lake, E., et al. (2013). *The Astrophysical Journal*, 768(2):196.

Schiminovich, D., Ilbert, O., Arnouts, S., Milliard, B., Tresse, L., Le Fèvre, O., Treyer, M., Wyder, T., Budavári, T., Zucca, E., et al. (2005). *The Astrophysical Journal Letters*, 619(1):L47.

UNIVERSITY of the
WESTERN CAPE

https://etd.uwc.ac.za

Songaila, A. and Cowie, L. L. (2010). *The Astrophysical Journal*, 721(2):1448.

Storrie-Lombardi, L. J. and Wolfe, A. M. (2000). *The Astrophysical Journal*, 543(2):552.

Takeuchi, T. T., Yoshikawa, K., and Ishii, T. T. (2003). *The Astrophysical Journal Letters*, 587(2):L89.

Verheijen, M. A. (2001). *The Astrophysical Journal*, 563(2):694.

Walter, F., Brinks, E., De Blok, W., Bigiel, F., Kennicutt Jr, R. C., Thornley, M. D., and Leroy, A. (2008). *The Astronomical Journal*, 136(6):2563.

Wong, O. I., Ryan-Weber, E. V., Garcia-Appadoo, D. A., Webster, R. L., Staveley-Smith, L., Zwaan, M. A., Meyer, M. J., Barnes, D. G., Kilborn, V. A., Bhathal, R., et al. (2006). *Monthly Notices of the Royal Astronomical Society*, 371(4):1855–1864.

Wyder, T. K., Treyer, M. A., Milliard, B., Schiminovich, D., Arnouts, S., Budavári, T., Barlow, T. A., Bianchi, L., Byun, Y.-I., Donas, J., et al. (2005). *The Astrophysical Journal Letters*, 619(1):L15.

Zafar, T., Péroux, C., Popping, A., Milliard, B., Deharveng, J.-M., and Frank, S. (2013). *Astronomy & Astrophysics*, 556:A141.

Zwaan, M. A., Meyer, M., Staveley-Smith, L., and Webster, R. (2005). *Monthly Notices of the Royal Astronomical Society: Letters*, 359(1):L30–L34.

UNIVERSITY of the
WESTERN CAPE

https://etd.uwc.ac.za