The University of Sydney
Faculty of Science
School of Physics

Young Active Galaxies across the Radio Spectrum

Emily Florence Kerrison

A thesis submitted in fulfilment
of the requirements for the degree of

Doctor of Philosophy

Supervised by
Elaine Sadler (University of Sydney)
&
Vanessa Moss (CSIRO)

2026

Abstract

The lifecycle of active galaxies is one of the puzzles of modern astronomy. Already some pieces of this puzzle have fallen into place; the feedback from active galactic nuclei (AGN) is thought to play a critical role in galaxy quenching and evolution, and the multi-wavelength emissions these AGN put out are known to be the product of an accreting supermassive black hole at their centre. However, not all active galaxies are active *radio* galaxies, in possession of synchrotron-emitting jets, and it is not obvious why.

Observations at radio wavelengths provide a unique window onto the lives of active galaxies; radio waves are unobscured by dust, revealing luminous active galaxies that are shrouded by their dusty galactic environment at other wavelengths. Furthermore, the radio emissions that reach us from extragalactic distances are primarily non-thermal, tracing the synchrotron jets of active galaxies. These jets in turn directly impact their multi-scale environment by propagating outwards from the sub-parsec scales surrounding the supermassive black hole, all the way to the megaparsec scale of galaxy clusters and environments. And, although the radio spectrum is relatively empty, it contains the 21 cm hydrogen line, a critical tracer of the cold, neutral gas throughout the Universe. Taken together, these facts make radio observations of active galaxies critical to our understanding of jet triggering, jet lifecycles, and jet-gas interactions.

This thesis was completed as part of the First Large Absorption Survey in H I , conducted with the Australian SKA Pathfinder (ASKAP-FLASH). Early on, we realised many of our H I detections in ASKAP-FLASH were being made towards so-called 'Peaked Spectrum' AGN, in which the synchrotron jets are still embedded in the dense, nuclear gas of their host galaxy, likely because they are young. This trend in H I-detected AGN has been seen before, but that it persists in an untargeted survey was surprising, and prompted this study into the nature of peaked spectrum sources. We present here a new Bayesian framework for identifying these young active galaxies using pre-existing radio survey data, which we make available to the community as a package, RadioSED. Then, applying RadioSED to a test field, we show we can increase the number of known young radio galaxies by more than an order of magnitude, simply through careful treatment of pre-existing datasets. We investigate the multi-wavelength properties of our new sample of young radio galaxies, and identify that many of them are distant, making such samples interesting probes of the physical conditions at cosmic noon and, potentially, even earlier. We then take early results from the ASKAP-FLASH Pilot Surveys, focusing again on the high detection rate towards compact jets, and use SANGRiA, a custom pipeline we developed to extract mock observations from cosmological simulations to determine whether we can recover this high detection rate under the assumption that it is a purely geometrical effect of the small jets. Finally, this thesis presents a multiwavelength study of a different type of radio AGN detected in H I with ASKAP-FLASH: an AGN which is not young but beamed, its synchrotron jets aligned close to our line of sight. This study of the beamed AGN, PKS 0405−385, demonstrates the gains to be made by mutli-wavelength follow up of ASKAP-FLASH detections, and reveals at least three intervening galaxies along our line of sight, one of which is the likely host of the H I.

Overall, this thesis reinforces the key role of young active galaxies in studies of AGN evolution, jet-gas interactions, and the changing distribution of gas across the different ages of our Universe.

Statement of Originality

This is to certify that to the best of my knowledge, the content of this thesis is all my own work. Where any part of this thesis is not my own work, I have indicated this by acknowledging the source of that part, whether that be written material, formal, or informal discussions.

No part of this thesis has been previously submitted to this or any other institution as part of a degree or for any examination purposes.

Use of Generative AI

During the course of my candidature, generative AI became widely accessible for the first time as a tool for end-users. For this thesis, Large Language Models (LLMs), primarily `ChatGPT` and `Claude`, were used as a coding tool, to aid in development and structuring of the Python package SANGRiA, which is used for analysis in Chapter 4. An LLM was also consulted to help discover appropriate references for some of the historical science discussed in Chapter 1; all references were verified before inclusion in this thesis. To my knowledge, generative AI was not used in any other way for any part of this thesis.

1 June 2026

Emily F. Kerrison — Date

Acknowledgements

The greatest thanks must go to my supervisors, Elaine and Vanessa, for their constant support and encouragement over the course of several years. Elaine, this thesis would not exist without you. I remember sitting on a call discussing thesis applications and you remarked that those peaked spectrum sources seemed interesting in the FLASH context. How right you were, and how far they have taken me. Thank you for guiding me through this new chapter in my academic journey, for teaching me how to become a better scientist, and for showing me how to prioritise what's important both in research and in life. Vanessa, so much of my thesis and beyond has been influenced by the way you approach your work. Thank you for all our conversations about how to handle data, how to handle meetings and how to approach a career. Thank you for introducing me to TFOM and for welcoming me into your world. Thank you also to my co-supervisor, Tara Murphy, for your support and advice, knowing you have been there when needed has meant a lot. I have learnt so much from you all about not only radio astronomy, but also what makes a good researcher, balancing independent work with collaboration, and balancing new projects against existing ones.

To Samuel & Stephanie, team SANGRiA, my home-away-from-home supervisors during my stay at the CCA, thank you for everything. Whatever my expectations were for our time together, you exceeded them. I learnt so much from you both: how 'speak simulator', why cold gas is hard, how to ask the right questions, what it means to be – small and close or big and far away – and also what its like to work in a different environment. I cannot overstate the value of my time with you, and I hope it was at least half as productive for you as it was for me. Samuel, thank you especially for helping me fit in to life at the CCA, and for turning a meeting room into a hack zone to get our project over the line.

They say it takes a village to raise a child, well it has taken a research group – or several – to train this PhD student. So much of my work has been shaped by discussions and meetings with the FLASH team: James Allison, Aditya JNHS, Elizabeth Mahony, Renzhi Su, Matthew Whiting, Simon Weng, Ivy Wong and Hyein Yoon. I have learnt something from all of you and I am proud to call you colleagues. Thanks also to the USyd transients group led by Tara Murphy, who invited me in with open arms when I had questions about variable AGN, then let me stay on so I could learn more about all things radio variable. Crossing continents now, thank you to the Galaxy Formation group at the CCA, who welcomed me in as a lone radio astronomer and taught me about everything from numerical diffusion to energy flows during my five month stay. Working closely with all these groups has been a pleasure and a privilege, and I look forward to continued collaboration.

Beyond these research groups I have benefited immensely from the expertise, guidance and advice of many talented astronomers and scientists. My heartfelt thanks to Ron Ekers and Elizabeth Mahony, who have each gone above and beyond to support me in my science, but also as a scientist, offering insight and guidance on everything from navigating the astronomical profession to an early draft of this thesis. Ron, thank you for distracting me. It is always a pleasure to meet with you and discuss anything and everything, from interferometry to sociology, and it is the very highest privilege to work alongside you. Liz, your openness to discuss each and every weird observing idea that pops into my head is so very much appreciated, as is your support in navigating the end of my PhD. Thanks also to Tara Murphy, Tim Bedding, Scott Croom and Jesse van de Sande for your tid-bits of wisdom, career guidance and advice. To the TFOM crew: Aidan, Elizabeth, Glen, Rika and Vanessa thank you for all the adventures and being together while we're apart.

Many thanks also to Phil Edwards, Tasso Tzioumis, Laura Driessen and Hyein Yoon, you all teach me something new every time we talk and my days are immeasurably better when we have the chance to dive into science together. Thanks also to James Allison for your continued support of my Bayesian journey - I hope I can keep pulling you back to astronomy for a little while longer yet. Thanks to the FLASH-IPS team who taught be about scintillation: Ron Ekers, John Morgan, Rajan Chhetri, and Angie Waszewski. Thanks to Ross Turner, Stas Shabala, and Sophie Young who first introduced me to simulations. Thank you to all the postdocs who lead by example - Liz Iles, Kelly Gourdji, Josh Pritchard, Lucia Perez, Christian Partmann and more. And thank you to Jamie Stevens for answering my 3am calls when ATCA misbehaves and for giving me the confidence to no longer pick up the phone.

None of this work would have been possible without the generous support of several institutes and funding

bodies throughout my candidature. My primary funding came from an Australian Government Research Training Program (RTP) Scholarship and fee offset which supported the backbone of my research. My heartfelt thanks must go to my home institutes, the Sydney Institute for Astronomy (SIfA) at the University of Sydney, and CSIRO Space & Astronomy, which provided me space to work, computational resources to play with, and research groups in which to grow. I would especially like to thank Marie Partridge, Leanne Edwards and Rob Hollow at these institutes, for navigating the administrative waters for me and making my candidacy that much smoother. Thanks must go as well to the CSIRO co-supervised student programme and funding, which allowed me to get deeply involved with radio astronomy at CSIRO, and visit the Australia Telescope Compact Array on several occasions. During my candidacy I also received support from ASTRO3D, and OzGrav 2.0 which allowed me to present my work at international conferences and form stronger connections with astronomers across Australia too through workshops, cross-node meetings and annual retreats. The value of these cross-institutional centres cannot be overstated and I am grateful for the chance to be a part of them. I would also like to thank the Australian Academy of Science and the Lindau Nobel Laureate Meetings Foundation for a truly transformational experience at Lindau, and the Simons Foundation for their generous support and faith in my science as part of the Centre for Computational Astrophysics Pre-Doctoral Program. I have certainly been fortunate in the varied institutes and organisations which have supported me, and I can only hope to do them justice with this thesis, and whatever comes next.

Next, thanks to you! I'm told that the acknowledgements are the first (and only) part of a thesis many people read. So thanks for stopping by, I hope you enjoy the journey. If you're after some background I'd recommend starting at Chapter 1, but otherwise feel free to choose your own adventure. If you're a mentor I hope I did our work together justice. If you're a colleague, I hope you find something useful. If a friend, I hope there's something in here for you too. If you're my student - get back to work! Just kidding, this is academic reading, so it counts. Got any questions about what you read? Ask me, I can't promise I'll remember the answer but I'll try. And if any of you are feeling particularly brave - come find me and tell me I owe you a coffee for reading my thesis, I'll only have myself to blame!

To friends both old and new, thank you for being part of my journey. Rhea, Karen, Kirsten, Shehani and Ricky thank you for over a decade of friendship, I hope there are many more decades to come. Luci, Nic, and Priyanka, I'm the final one to finish but we've done it now! Thank you for sharing this journey with me, I look forward to what comes next for all of us. Katherin and Tariq, thank you for a friendship so strong it withstands the distance between Australia and the UK. Liroy and Sukey thank you for your fabulous dinner parties and even more fantastic conversation. Kovi thanks for the memes and for making me a podcast celebrity. Simon, Ashna, Kavya, Angie, and Tamsyn you've all made the student experience special. Aryanna, Thummim, and Aizhan thank you for exploring New York with me and making lifelong memories. James and Anthony you let me dip my toes back into the Classics. Raphaela, Sara, Bill and Sarah you welcome me back with open arms though my visits are infrequent. Thank you to you all, and to the many others who are a part of this story; if your name is not here it is only because my memory is not well ordered enough to pull together all the names I need. To Pablo, thank you for always asking what's on my mind, for sharing what's on yours, and for so many wonderful moments in so many different places. I am so glad our spacetime coordinates overlap.

Finally, my deepest thanks to my mother, who once more came with me on this adventure after I boldly signed up to do yet another thesis (a postgraduate one this time!). Throughout my academic journey you have always been my biggest supporter, and it is without doubt thanks to you that I have come this far. You first nurtured my love of learning with books on bugs and space long before I met them in a classroom, and you first taught me to be curious by never turning down a question, no matter the place or the subject. Your unwavering support has always been appreciated, and these last four years are no different, except to say that I have noticed more fully just how deep it runs. Thank you for plying me with endless cups of tea, for sitting through my practice talks, and for showing interest in my work.

This thesis was primarily written at the University of Sydney and the Australia Telescope National Facility, which stand on the lands of the Gadigal and Wallumedegal people of the Eora nation.

The work presented in this thesis relies heavily on data from the Australian SKA Pathfinder, and the Murchison Widefield Array, which stand on the land of the Wajarri Yamaji People.

Author Attribution Statement

The contents of this thesis include three published works:

- **Chapter 2 and Appendix C** are published as *RadioSED - I. Bayesian inference of radio SEDs from inhomogeneous surveys.* Kerrison, E. F., Allison, J. R., Moss, V. A., Sadler, E. M., Rees, G. A. 2024. *Monthly Notices of the Royal Astronomical Society* 533, 4248–4267. doi:10.1093/mnras/stae1796.

  The associated code is available on Github and has its own entry in the *Astrophysical Source Code Library* under *RadioSED: Radio SED fitting for AGN*. 2024ascl.soft08012K.

- **Chapter 3 and Appendices D and E** are published as *RadioSED - II. Discovering the peaked spectrum radio sources in Stripe 82.* Kerrison E. F., Sadler E. M., Moss V. A., Mahony E. K., Driessen L., Ross K., Rose K., Dobie D., Murphy T., 2025. *Monthly Notices of the Royal Astronomical Society* 543, 3895–3914. doi:10.1093/mnras/staf1643.

- **Chapter 5** is published as *One sightline, many systems: a FLASH discovery of HI towards scintillating quasar PKS 0405-385.* Kerrison E. F., Yoon, H., Sadler. E. M., Kang, Y., Edwards, P. G., Tuntsov, A., Pritchard, J. P., Moss, V. A., Mahony, E. K., Bignall, H., Aditya, J. N. H. S., Allison, J. R., Curran, S., Ekers, R. D., Glowacki, M., Stevens, J., Su., R, and Whiting, M. *Publications of the Astronomical Society of Australia* 43 e058. doi:10.1017/pasa.2026.10190.

and a further three works which are in various stages of publication:

- **Chapter 4 and Appendices F and G** are under review at the *Astrophysical Journal* under the title *Forward Modelling the ASKAP-FLASH 21cm Absorption Survey with SANGRiA: Geometry Alone Cannot Explain High HI Detection Fractions Towards Compact Radio Sources* with co-authors Kerrison, E. F., Ward, S. R., Tonnesen, S., V. A. Moss, and E. M. Sadler.

- **Appendix A** was accepted for publication in the *Proceedings of the International Astronomical Union*, Vol. S392 in October 2024 under the title *Why so young? A curious connection between the broadband flux and neutral gas content of AGN* with co-authors Kerrison, E. F., Sadler, E. M., Moss, V. A., and Mahony, E. K. doi:10.48550/arXiv.2412.08002.

- **Appendix B** was accepted for publication in the *Proceedings of the International Astronomical Union*, Vol. S394 in December 2024 under the title *From terrestrial weather to space weather through the history of scintillation* with co-authors Kerrison, E. F., Ekers, R. D., Morgan, J., and Chhetri, R. doi:10.48550/arXiv.2412.19816.

The chapters included in this thesis are reproduced in their original form with only minor typographical corrections and formatting adjustments made to ensure a consistent style throughout. In all cases, I performed the analysis contained within the work and wrote the manuscript. More detailed statements of contribution, including specific contributions from co-authors, can be found in the preamble to each individual chapter.

Research conducted for this thesis (specifically radio SEDs of ASKAP-FLASH detections) has also been featured in an appendix to *The first large absorption survey in H I (FLASH): II. Pilot survey data release and first results*. Yoon H., Sadler E. M., Mahony E. K., Aditya J. N. H. S., Allison J. R., Glowacki M., Kerrison E. F., et al., 2025, *Publications of the Astronomical Society of Australia*, 42, e088. doi:10.1017/pasa.2025.10046.

In addition to the authorship attribution statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

1 June 2026

Emily F. Kerrison — Date

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

1 June 2026

Elaine Sadler [Lead Supervisor] — Date

Acronyms, Abbreviations, and Conventions

Given below are useful terms and acronyms that appear at several points throughout this thesis, collected in the one place for easy reference.

ACC	Alma Calibrator Catalogue.
AGN	Active Galactic Nucleus, the dense core of a galaxy exhibiting bright emission.
ALMA	Atacama Large Milimeter Array.
ALMACAL	The ALMA calibrator catalogue (alternative acronym).
ANU	Australian National University.
ARC	Australian Research Council.
ASKAP	Australian SKA Pathfinder.
ASKAPsoft	The default processing pipeline for ASKAP data products.
ASTRO3D	Australian Research Council (ARC) Centre of Excellence for All Sky Astrophysics in 3 Dimensions.
AT20G	The Australia Telescope 20GHz Survey.
ASKAP	Australian SKA Pathfinder.
ATCA	Australia Telescope Compact Array.
ATNF	Australia Telescope National Facility.
ATPMN	Australia Telescope-Parkes-MIT-NRAO Survey.
AURA	Association of Universities for Research in Astronomy.
AusSRC	Australian SKA Regional Centre.
BASS	Beijing-Arizona Sky Survey.
BH	Black Hole.
BLR	Broad Line Region.
CASDA	CSIRO ASKAP Data Science Archive.
C-BASS	C-Band All Sky Survey.
CCA	Centre for Computational Astrophysics.
CDF	Cumulative Distribution Function
CDS	The Strasbourg astronomical Data Center.
CGM	Circumgalactic Medium.
CMB	Cosmic Microwave Background.
CO	Carbon Monoxide.
CSIRO	Commonwealth Scientific and Industrial Research Organisation.
CSO	Compact Symmetric Object.
CSS	Compact Steep Spectrum.
DESI	Dark Energy Spectroscopic Instrument.
DECaLS	The Dark Energy Camera Legacy Survey.
DLA	Damped Lyman Alpha.
DR(#)	Data Release #.

DRAGONS Data Reduction for Astronomy from Gemini Observatory North and South.

DyNesty Dynamic Nested Sampling.

EMU Evolutionary Map of the Universe.

FoV Field of View.

FFA Free-Free Absorption.

FIRST Faint Images of the Radio Sky at Twenty-Centimeters.

FITS Flexible Image Transport System.

FIRE Feedback In Realistic Environments.

FLASH First Large Absorption Survey in HI.

FLASHfinder They Bayesian linefinder tool used by FLASH to detect H I absorption in survey spectra.

Fpk Peak flux density.

Fpwide Peak flux density across the full MWA bandwidth.

Fintwide Integrated flux density across the full MWA bandwidth.

FR 0 Fanaroff-Riley class 0.

FR I Fanaroff-Riley class I, a core-dominant radio AGN.

FR II Fanaroff-Riley class II, a lobe-dominant radio AGN.

FSRQ Flat Spectrum Radio Quasar.

FWHM Full Width Half Maximum.

GADGET A code used for cosmological simulations.

GAMA Galaxy And Mass Assembly (survey).

GLEAM GaLactic and Extragalactic All-Sky MWA Survey.

GLEAM-X GaLactic and Extragalactic All-Sky MWA Survey eXtended.

GMOS Gemini Multi Object Spectrograph.

GMRT Giant Metrewave Radio Telescope.

GPS Gigahertz Peaked Spectrum.

HI Neutral hydrogen.

IAU International Astronomical Union.

IDV Intra-Day Variability.

IPS Interplanetary Scintillation.

IR Infrared.

IRAF Image Reduction and Analysis Facility.

ISM Interstellar Medium.

ISS Interstellar Scintillation.

Jy A Jansky is a non-SI unit which radio astronomers use for spectral flux density, where $1\,\mathrm{Jy} = 10^{-26}\,\mathrm{W\,m^{-2}\,Hz^{-1}}$

JWST James Webb Space Telescope.

KDE Kernel Density Estimate.

KS Kolmogorov-Smirnov.

LARGESS Large Area Radio Galaxy Evolution Spectroscopic Survey.

Λ-CDM	Lambda cold dark matter (cosmology).
LOFAR	Low Frequency Array.
LoTSS	LOFAR Two-metre Sky Survey.
LRD	Little Red Dot.
MALS	MeerKAT Absorption Line Survey.
MARTINI	Mock Array Radio Telescope Interferometry of the Neutral ISM.
MARZ	Manual and automatic redshifting software.
MeerKAT	An SKA precursor telescope.
MWA	Murchison Widefield Array.
MRC	Molonglo Reference Catalogue.
MUSE	Multi Object Spectrograph.
NASA	National Aeronautics and Space Administration.
NED	NASA/IPAC Extragalactic Database.
NEOWISE	Near-Earth Object Wide-field Infrared Survey Explorer.
NLR	Narrow Line Region.
NOIRLab	National Optical-Infrared Astronomy Research Laboratory.
NRAO	National Radio Astronomy Observatory.
NSF	National Science Foundation.
NSI	Normalised Scintillation Index.
NVSS	NRAO VLA Sky Survey.
Pan-STARRS1	Panoramic Survey Telescope and Rapid Response System Phase 1.
PICZL	Photometrically Inferred CNN redshift(Z) Likelihoods.
PS	Peaked Spectrum.
QSO	Quasi-Stellar Object.
RACS	Rapid ASKAP Continuum Survey.
RadioSED	A code for compiling and fitting radio flux densities.
RISS	Refractive Interstellar Scintillation.
RMS	Root Mean Square.
Roma-BZCAT	A catalogue of blazars.
RTP	Research Training Programme.
SANGRiA	Simulating Absorption of Neutral Gas for Radio Astronomy.
SED	Spectral Energy Distribution.
SETI	Search for Extraterrestrial Intelligence.
SDSS	Sloan Digital Sky Survey.
SIMBA	A cosmological simulation.
SIMBAD	SSet of Identifications, Measurements, and Bibliography for Astronomical Data.
SKA	The telescope currently being constructed by an international consortium with sites in Australia and South Africa.
SMB	Supermassive Black Hole.

SNR	Signal-to-Noise Ratio.
SPECFIND	A tool for compiling radio flux densities stored in VizieR.
SPS	Soft Peaked Spectrum.
SSA	Synchrotron Self Absorption.
SUMSS	Sydney University Molonglo Sky Survey.
TIFR	Tata Institute of Fundamental Research.
TGSS	TIFR GMRT Sky Survey.
TXS	A radio survey conducted with the Texas interferometer.
USS	Ultra Steep Spectrum.
VAST	Variables And Slow Transients, an ASKAP survey.
VI	Variability Index.
VizieR	A database and catalogue service.
VLA	Very Large Array.
VLBA	Very Long Baseline Array.
VLBI	Very Long Baseline Interferometry.
VLASS	Very Large Array Sky Survey.
VLSSr	VLA Low-Frequency Sky Survey redux.
VO	Virtual Observatory.
WISE	Wide-field Infrared Survey Explorer.
WISE-PS1	WISE-Pan-STARRS1 combined dataset.
WISE W#	One of four photometric bands in which the WISE instrument observers.
WFC3	Wide-field Camera 3.

Contents

List of Figures

List of Tables

1

An introduction to the radio lives of active galactic nuclei

In the 1960s there was a paradigm shift across astronomy, astrophysics and cosmology, with a perfect storm of advances in instrumentation, mathematical models and cross-disciplinary activity which led to the discovery of the 'Violent Universe' (Bonolis & Furlan, 2025). Penzias & Wilson (1965) detected the cosmic microwave background, interpreted as the afterglow of the Big Bang (Dicke et al., 1965), Jocelyn Bell discovered a rapidly pulsating radio source later identified as the first pulsar (Hewish et al., 1968), and Sandage (1965) identified a new class of extragalactic object far too distant and too luminous to be powered by stellar radiation. It is the last of these, the quasars, and the broader class of active galactic nuclei (AGN) which concern us here.

Even after more than sixty years of study there is much that remains unknown about AGN. What we do know is that they are the central regions of galaxies which exhibit an excess in electromagnetic radiation. However, their triggering mechanism is unclear, and while mergers have often been proposed (e.g. Ellison et al., 2019) there is mixed evidence as to their role (e.g. Sharma et al., 2024). Many AGN have black hole mass estimates too high for secular growth by accretion alone, so it is unclear how they formed, though some may be primordial (Carr et al., 2021), while others may have formed in the very early Universe via direct collapse of massive clouds (Bogdán et al., 2024). AGN formation and triggering are among the key questions which are the focus of the field today; to address them requires a combination of observations across the electromagnetic spectrum, as well as a theoretical framework to measure these observations against.

This chapter provides a short overview of how our understanding of AGN has co-evolved with our instrumentation and our observations, to reach the level of knowledge we have acquired today. It begins with Section 1.1, summarising what is known about AGN structures from sub-parsec, to galactic scales. This includes a discussion of their observational signatures across the full electromagnetic spectrum, though radio frequency observations are particularly useful since the dust which permeates so many galaxies is transparent to radio wavelengths. This fact allows radio waves to reach us where other wavelengths do not. Furthermore, the radio spectrum contains the 21 cm hydrogen line, a tracer of the most abundant element in our Universe and a vital means of understanding the galactic environments in which AGN reside. This thesis therefore focuses on a radio perspective of active galactic nuclei, with a special interest in the earliest evolution of the radio-bright population and their local environment, as discussed in Section 1.2.

Much of the work in this thesis has been done within, or with heavy reliance on, large area surveys, and so Section 1.3 includes a discussion of how these surveys have developed to take advantage of breakthroughs in both computation and in instrumentation. This is complemented by Section 1.4, which outlines a parallel strand of progress in the storage, processing and analysis of what today is known as 'Big Data'. Section 1.5 extends this story of Big Data even further, introducing astrophysical simulations with a discussion of how these can be useful to the observational astronomer, before ending with a summary of the remainder of this thesis in Section 1.6. Without further ado then, let us begin with the science that began this field of research; the discovery of active galactic nuclei.

1.1 Active Galactic Nuclei

The discovery of AGN was firmly rooted in multiwavelength astronomy, requiring both the lunar occultation measurements of Hazard et al. (1963) to identify the extended radio structure of the quasar 3C 273, and the spectroscopic follow up of Schmidt (1963) to observe redshifted Balmer, MgII and [OIII] lines at the enormous (for the time) redshift of $z = 0.158$. Crucially, Schmidt also related these observations at different frequencies to the point-like optical emission at the centre of another galaxy. Soon after that Sandage (1965) identified a sample of radio quiet, optically-selected sources and firmly cemented 'quasi-stellar objects' (QSOs) as a new class of extragalactic objects. This class was soon expanded to include the Seyfert galaxies, discovered almost two decades earlier by Seyfert (1943), thanks to the multiwavelength analysis of Pskovskii (1962) who drew a link between their optical emission and the luminous emissions of radio galaxies. Yet it was still a decade before the *Astrophysical Journal* permitted use of the umbrella term 'quasar' to describe these various objects, and the term AGN (which was originally only used for the low-luminosity counterparts to quasars) came about by equally convoluted means (Kellermann & Bouton, 2023).

But what is an AGN? Put simply, it is the core of a galaxy (a 'galactic nucleus') which exhibits an excess in electromagnetic radiation well beyond that attributable to stars alone, making it 'active'. This emission is thought to be powered by the extraction of energy from a deep gravitational potential well surrounding a supermassive ($\gtrsim 10^6 M_\odot$) black hole (SMBH), which is then released either radiatively or mechanically into its surrounds (Netzer, 2013). Such a transformation of energy must necessarily involve the movement of matter through the potential well, likely as part of accretion onto the SMBH which suggests some level of interaction with the local environment. In fact, observations show these interactions are more than strictly local; it is now well established that central SMBHs co-evolve with their host galaxies, affecting in particular the evolution and stellar dispersion of the galactic bulge (Kormendy & Ho, 2013), and more broadly an AGN can affect everything from star formation rates to gas cooling timescales (Heckman & Best, 2014), and larger-scale radiative and mechanical feedback can even have affects across a local group or cluster, creating cavities in the hot, X-ray emitting gas that permeates these larger systems (Bîrzan et al., 2004; Fabian, 2012; Hlavacek-Larrondo et al., 2022). One way to define an AGN then, is as a driver of multi-scale evolution.

Over cosmic history these feedback processes compound, and seminal papers in computational astrophysics demonstrated that they are necessary for reproducing observed stellar velocity dispersions, star formation rates and galaxy colours in the local Universe (Di Matteo et al., 2005; Croton et al., 2006; Sijacki et al., 2007). AGN are therefore a critical piece in the puzzle of our "violent universe" and how it reached the state in which we find it today. The AGN themselves also remain important laboratories in which to observe extreme physics including strong lensing (Millon et al., 2023), cosmic ray production (IceCube Collaboration et al., 2018), and accretion (Merloni et al., 2003). Furthermore, since they are so intrinsically luminous, we can observe how AGN evolve across large fractions of cosmic history; the most distant AGN identified to date is at a redshift = 10.03 (Bogdán et al., 2024), and there is emerging evidence that they may be related to another class of object in the high redshift Universe known as Little Red Dots (LRDs), discovered in 2024 with JWST (Matthee et al., 2024). Beyond their local, galactic environment, the high intrinsic luminosity of AGN means they illuminate everything in their path, and can be used to peer through the dark Universe, revealing gaseous structures in absorption that would otherwise remain hidden along our line of sight towards them. Such systems are typically referred to as 'intervening absorption' systems when observed at radio wavelengths (detecting 21 cm, H I absorption), or 'transverse absorption' systems at optical wavelengths (detecting a range of atomic transitions, including Mg II, Lyα and many others). When the gas column density is particularly high in any such absorption system ($N_{HI} \geq 2 \times 10^2\ cm^{-2}$), it is thought that it contains a large fraction of neutral (as opposed to ionised) hydrogen, making it an ideal reservoir for star-forming gas. Such systems are commonly referred to 'Damped Lyman-α' systems (DLAs), and are key tracers of chemical evolution across cosmic time (Wolfe et al., 2005).

But to return to our initial question we might ask again, what really is an AGN? Over some 60 years of detailed study we have developed phenomenological models that explain many of the diverse observational signatures associated with AGN, even if certain critical details like their triggering mechanism remain unknown. We outline those key phenomena and the physics thought to produce them below.

1.1.1 The Unification Model and AGN accretion modes

AGN are observationally extremely diverse, yet it is thought that they largely comprise the same basic building blocks described in the Unification Model of Urry & Padovani (1995), and represented schematically in Figure 1.1. The SMBH at the very centre is thought to be surrounded by a geometrically thin, optically thick accretion disk composed of hot ($T \gtrsim 10^4$ K) gas (Bonning et al., 2007, Figure 1.1, left), which may turn into a geometrically thick, advection-dominated disk in the innermost regions of some AGN (Figure 1.1, right). The transition between a thin, radiation-dominated disk and a thick, advection dominated one is strongly linked to the SMBH accretion rate, with efficient accretion above $\sim$ 1 per cent of the Eddington ratio, $\lambda_{\rm Edd} = \frac{L_{bol}}{L_{Edd}}$, corresponding to a radiation-dominated disk, and accretion rates below this to an advection-dominated regime (Heckman & Best, 2014).

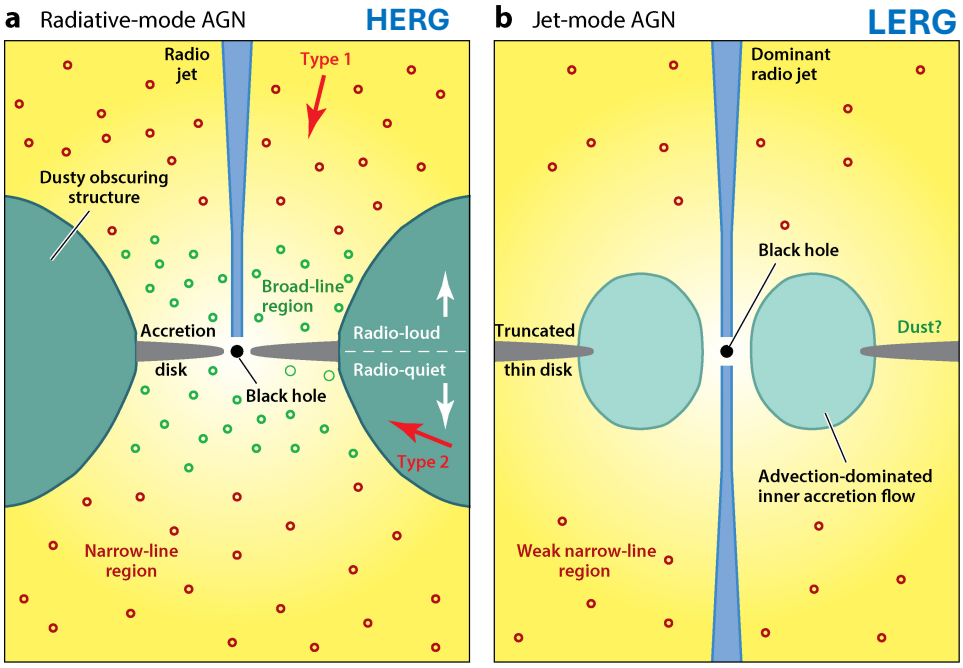


Figure 1.1: A schematic showing the key components of AGN and how they differ based on accretion rates in the innermost regions surrounding the black hole; efficient accretion ($> 0.01\lambda_{Edd}$) is thought to occur in radiative-mode AGN like those on the left, while inefficient accretion ($< 0.01\lambda_{Edd}$) occurs in jet-mode AGN like those on the right. From Heckman & Best (2014)

Along the axis perpendicular to the accretion disk, we sometimes see a jet of relativistic particles produced by interactions between said accretion disk, the SBMH spin, and strong, local magnetic fields. These magnetic fields collimate the jets close to the SMBH, which can terminate anywhere from $<$ 1 parsec to megaparsecs away in strong shocks or voluminous lobes depending on their power (Blandford et al., 2019). The jets are brightest at radio frequencies below a few gigahertz, as discussed below, so when an AGN is observed to possess jets it is often said to be exerting 'radio mode' or 'jet mode' feedback. Both radiation-dominated and advection-dominated accretion disks can produce these jets, as shown in Figure 1.1, but in the case of advection-dominated systems they are often the main observational signature of the AGN, which is why these are sometimes called 'Jet-mode AGN'. When radio-selected samples of AGN are studied in which all of the objects posses jets, those undergoing efficient accretion are classified as 'High Excitation Radio Galaxies' (HERGs), while those accreting inefficiently are considered 'Low Excitation Radio Galaxies' (LERGs; Heckman & Best 2014).

Radiation-dominated AGN, which are often referred to as 'quasars', and are shown in Figure 1.1, left, emit large amounts of quasi-thermal energy into the gas surrounding the accretion disk, which is capable of both ionising this ambient medium and driving powerful winds and outflows (Antonucci, 1993; Netzer, 2015). This ionised, outflowing gas is seen in the Broad Line Region (BLR) less than $\sim$ 1 pc from the accretion disk, while further out on pc–kpc scales we find the Narrow Line Region (NLR). These are detected emitting spectral lines that are, respectively, broad ($1,000 - 25,000\,\mathrm{km\,s^{-1}}$) and narrow ($\lesssim 500\,\mathrm{km\,s^{-1}}$) due to the kinetic energy and

differential bulk motion of the gas at these radii (Peterson, 2006; Heckman & Best, 2014). Outside of this sits the optically thick 'dusty torus' (which may or may not be torus shaped) comprising warm ($T \sim 10^3$ K) molecular dust and gas (Zhao et al., 2021). Advection-dominated AGN, by contrast, do not emit sufficient energy to photo-ionise large amounts of gas, and they may only have a weakly-emitting NLR (Heckman & Best, 2014). Each of these regions emits different electromagnetic signatures, and different combinations of those signatures have historically been given different sub-classifications, as discussed in Padovani et al. (2017). However the unification model posits that this wide variety of observational signatures trace the same structures seen at different viewing angles, as in the case of 'Type 1' and 'Type 2' quasars indicated by red arrows in Figure 1.1, left.

Encircling the entire AGN there is also a corona of hot gas ($\sim 10^5 - 10^6$ K), not shown in in Figure 1.1 since it focuses mainly on the central features, but this hot gas is thought to feed onto the accretion disk via a number of processes including smooth Bondi accretion, and chaotic cold accretion as a result of thermal instabilites (e.g. Gaspari et al., 2013; Yuan & Narayan, 2014). To better understand the role of AGN in their galactic homes though, we must consider this larger-scale environment too.

1.1.2 The Evolution of Active Galactic Nuclei

Just as AGN are thought to exist in two 'modes', so too are there two frameworks for how these modes evolve. The most luminous radiative-mode AGN (which are classified as quasars at optical wavelengths and as HERGs in radio-selected samples) are thought to be triggered by a rare event which deposits large amounts of gas around the central SMBH, such as a merger or interaction with a nearby galaxy (Hopkins et al., 2008). Soon after this interaction the central quasar is thought to be obscured by a large amount of dust and gas which is gradually blown out or photo-ionised to reveal an unobscured quasar, which then eventually switches off, leaving behind a passive elliptical galaxy. This model for quasar evolution is supported by observations of high merger rates in samples of red or obscured quasars (e.g. Kocevski et al., 2015; Fan et al., 2016), observations of quasar-driven outflows (e.g. Harrison et al., 2014; Harrison & Ramos Almeida, 2024), and most recently at radio wavelengths, observations of elevated radio emission in matched samples of red quasars (e.g. Fawcett et al., 2023; Sargent et al., 2026). The powerful, chaotic interactions in this model are thought to be so rare as to only happen once in the lifecycle of an active galaxy.

The jet-mode AGN (LERGs at radio wavelengths) and low-luminosity HERGS are, by contrast, thought to exist in a more balanced, 'maintenance mode', where they are fuelled by the stochastic accretion of gas from minor galaxy-galaxy interactions and other secular processes driving gas repeatedly towards the central SMBH (Hopkins & Hernquist, 2006; Croton et al., 2006). The radio jets which switch on as a result then heat the surrounding gas on nuclear scales, preventing runaway cooling and accretion, and possibly even cutting of their own fuel supply leading to their cessation, before the cycle (potentially) starts again. The LERGs in particular are often seen in massive, quenched galaxies, and since this type of low-level fuelling does not require a rare, explosive event like a merger, it is thought that it could happen continuously in a kind of 'maintenance cycle' (Best et al., 2014). This framework is supported by the fact that the LERG population evolves only mildly with redshift (Best et al., 2014; Kondapally et al., 2022).

1.1.3 Observing Active Galactic Nuclei

The observational signatures of AGN (and indeed of many astronomical objects) come in two primary forms: broadband, continuum emission, and spectral line emission produced by atomic transitions. As mentioned in Section 1.1.1, these spectroscopic signatures may be doppler broadened if they originate in the inner 0.1-1 pc BLR where the gas is moving at high velocities ($\sim 10^4\,\mathrm{km\,s^{-1}}$), or they may be narrower if the emissions originate in the $\sim 10^2$ pc scale NLR where the gas velocity is an order of magnitude lower (Netzer, 1990). The spectral signatures of these regions are shown in Figure 1.2.

The broadband components are often studied using a Spectral Energy Distribution (SED) of the kind shown in Figure 1.3. Across all wavelengths, the continuum emission from an AGN will dominate the light of its galactic host, unless the two can be spatially resolved or disentangled via SED modelling (e.g. Ciesla et al., 2015). To perform such modelling, broadband flux density measurements are collected across some range of wavelengths (typically the more independent measurements the better), and models are then fit to this

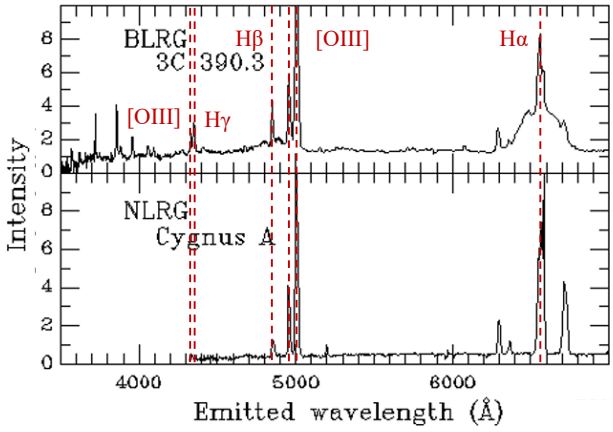


Figure 1.2: Optical spectra of two AGN tracing the BLR (top) and NLR (bottom), modified from Ajello (2007).

dataset which encode the properties of the central AGN and/or its host galaxy. These models may be based on empirical templates from other, well-characterised systems, synthetic observations of theoretical systems, or simple, analytic functions as in the case of Chapter 2, and some form of model comparison is always employed to determine the best fit to the data (for a recent review, see Pacifici et al., 2023).

At X-ray wavelengths, the hot corona (dashed blue line in Figure 1.3) and accretion disk (purple dot-dashed) likely produce some mixture of thermal and inverse Compton emission though the exact mechanism is debated (e.g. the discussion in Waddell et al., 2024). In the case of accretion disk emissions some fraction may be absorbed by the torus and re-emitted at infrared wavelengths (red dashed). The thermal emissions from the accretion disk are dominant at UV and into optical wavelengths, producing a so-called 'Big Blue Bump' in the broadband SED (solid blue line) (e.g Shang et al., 2005). At radio wavelengths (yellow lines), the relativistic jets produce strong synchrotron emission which can be described by a power law:

$$S_\nu = a\nu^\alpha \tag{1.1}$$

where S_ν is the flux density in Janskys[1] at wavelength ν, a is the amplitude of the spectrum and α is the synchrotron spectral index. In AGN this spectral index typically takes a value of $\alpha \approx -0.7$. Where an AGN exhibits some excess of radio emission beyond that attributable to star formation alone, it can be classified as a 'radio AGN' for the purposes of this thesis, and typically this emission is thought to originate in jets, though some may also be the product of AGN-driven winds (Nims et al., 2015; Escott et al., 2025). Furthermore, because the jets are the only element of the AGN to directly propagate out through the galactic host and beyond, they are an important agent in, and tracer of, galaxy-scale interactions.

1.1.4 AGN jets observed 'down the barrel'

Since the synchrotron jets of an AGN are approximately bi-conical, their observational signature varies wildly depending on their inclination with respect to our line of sight. In particular, when the jet axis is parallel to our sightline, the jets have a small angular scale and relativistic beaming boosts the intensity of emissions while compressing the timescales over which variations are observed (Blandford & Rees, 1978). Under these circumstances, the AGN is classified as a blazar. Blazar light curves can vary over timescales as short as a few days, and across the electromagnetic spectrum from radio to $\gamma-$rays (Singh & Meintjes, 2020). For this reason, blazars are often analysed separately to other radio AGN where the jets are observed at greater inclination angles. However, as we will see in Section 1.2.1, their radio variability and small angular size sometimes mean these beamed AGN appear unexpectedly in samples of other radio sources.

[1] A Jansky is a non-SI unit which radio astronomers use for spectral flux density, where 1 Jy = 10^{-26} W m^{-2} Hz^{-1}

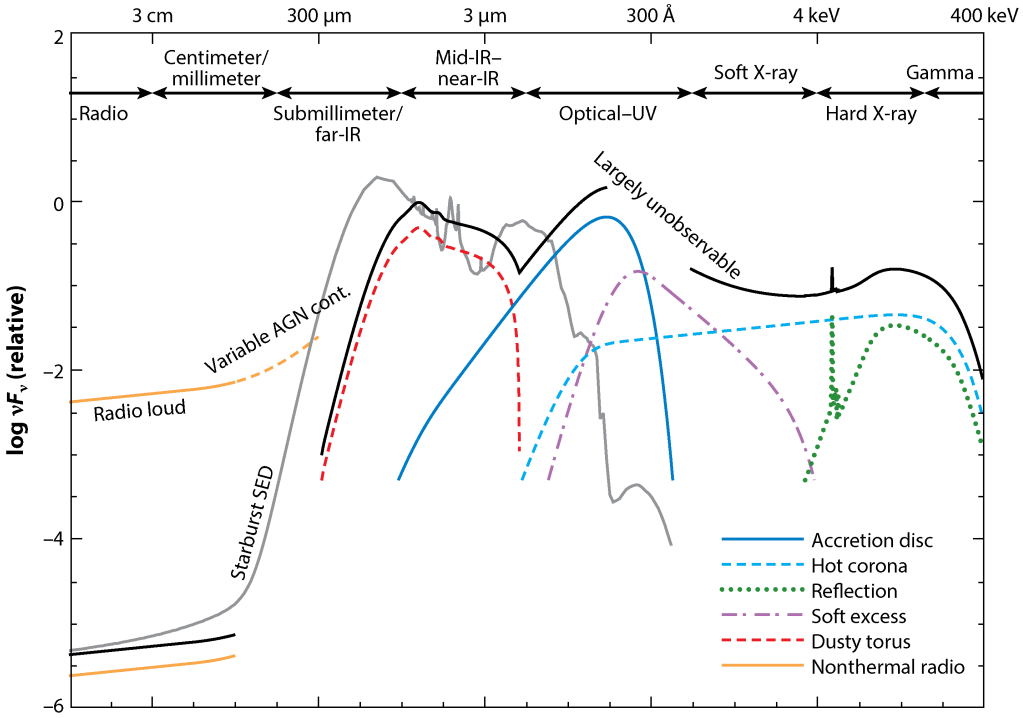


Figure 1.3: The multiwavelength SED of a typical AGN, broken down into components, as presented in (Hickox & Alexander, 2018). Credit: C. M. Harrison

1.1.5 Active Galaxies beyond the Nucleus: a multiphase environment

The synchrotron jets are the first element of an AGN we have described as extending out beyond the nucleus and into the wider host galaxy. This means they are capable of driving feedback across galaxy scales, quenching star formation in some systems by expelling and heating the cold gas it requires (Croton et al., 2006), but inducing it in others by compressing and cooling the interstellar medium (ISM) (Gaibler et al., 2012; Zinn et al., 2013). However, jets are not the only feature of the larger scale 'active galaxy' which set it apart from regular galaxies. Another key element of galaxy-scale interactions referenced already in Section 1.1 is the broader notion of AGN feedback, which is thought to be powered mainly by outflows, and which is responsible for regulating star formation rates and gas cooling (Di Matteo et al., 2005; Croton et al., 2006).

Every galaxy is composed of stars, dust, and a multiphase ISM, all of which slowly co-evolve by means of inflows, outflows and star formation as part of the so-called cosmic baryon cycle (Péroux & Howk, 2020). This large-scale inflow, outflow and recycling of gas between the ISM and the less dense circumgalactic medium (CGM) surrounding it is represented in Figure 1.4. Where a non-active galaxy might have starburst-driven outflows which reach velocities $v \sim 10^2\,\mathrm{km\,s^{-1}}$, AGN-driven outflows are typically much more powerful and reach higher velocities, especially on galactic scales (Förster Schreiber et al., 2019; Veilleux et al., 2020). They will also travel further, extending out to hundreds of kiloparsecs like the outflows shown in Figure 1.4. These AGN-driven outflows can be launched by the propagation of the jets, winds off the accretion disk, or direct thermal radiation pressure, all of which can entrain the multiphase ISM of the host galaxy, effectively removing gas from the inner galactic regions (e.g. King & Pounds, 2015; Ishibashi et al., 2018). Simulations have also shown that the resulting turbulence can cause a backflow onto the AGN (e.g. Wagner et al., 2013; Mukherjee et al., 2018b). These outflows have been detected across virtually all length scales, from the central few parsecs surrounding an AGN out to galaxy-wide scales and beyond (Harrison et al., 2014; Winkel et al., 2023). They are also seen across all phases, from cold molecular ($\sim 10\,\mathrm{K}$), to cool atomic ($\sim 10^2\,\mathrm{K}$), and warm-to-hot ionised gas ($\sim 10^3 - 10^5\,\mathrm{K}$), though at present most outflows have only been observed in one or two phases (Harrison & Ramos Almeida, 2024). Since cold ($T \sim 10\,\mathrm{K}$) gas is critical for star formation, being able to trace the cooler (atomic and molecular) phases of an outflow is, in some sense, most interesting for understanding the future of an active galaxy, many of which show decreased star formation rates relative to galaxies without an AGN (Shimizu et al., 2015; Mountrichas et al., 2022). Furthermore, simulations predict that a significant fraction of the mass is carried by cold gas in multiphase outflows (e.g. Kim et al., 2020; Tanner & Weaver, 2022).

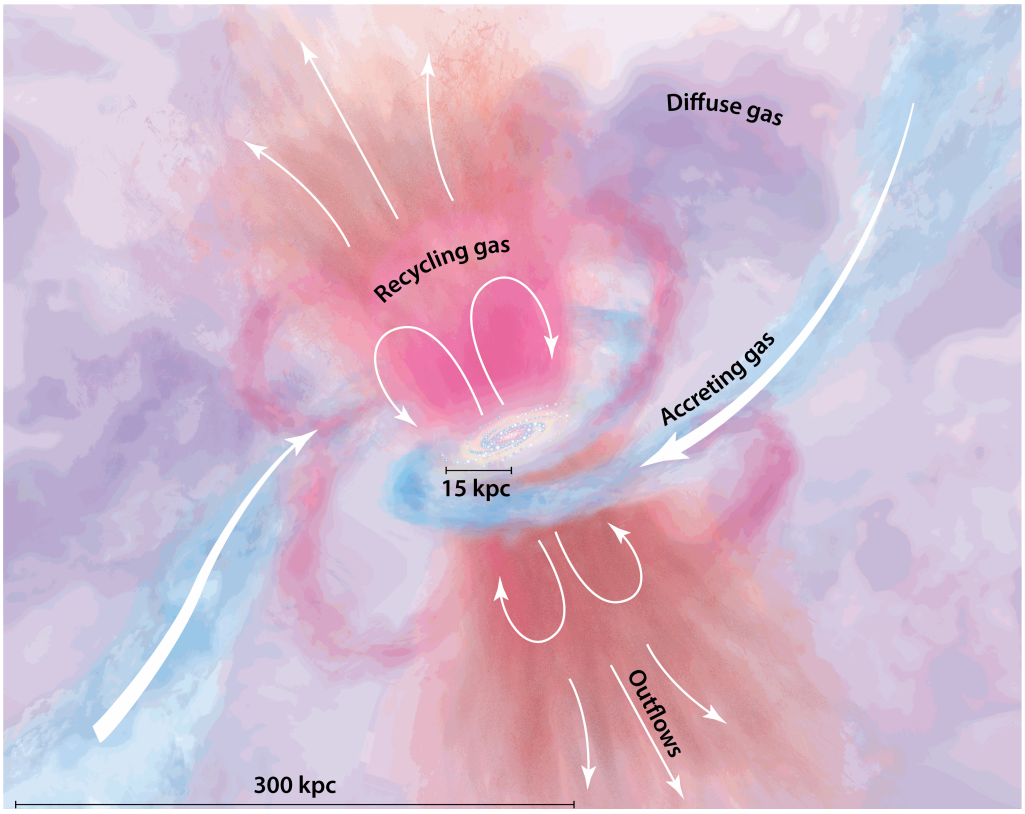


Figure 1.4: An artist's impression of multiphase gas cycling through a galaxy via a series of inflows and outflows (Tumlinson et al., 2017).

There are many atomic transitions spread across the restframe IR-UV spectrum which can be used observationally to trace neutral outflows and the neutral ISM of a galaxy more generally (a good overview can be found in Veilleux et al., 2020), but ultimately they all require assumptions about the metallicity and ionisation state of the gas in order to convert it to the 'fundamental' measure of neutral gas content, the hydrogen column density (N_{HI}). The one exception to this is the Lyα line of hydrogen (λ1216Å), but since it is in the restframe UV it is impossible to detect from the ground at $z \lesssim 1.7$, and space-based spectroscopy is both difficult and expensive. As a result observational samples are small (Rao et al., 2017). Fortunately, another tracer is available to us in the radio domain: the 21 cm H I line, produced by a hyperfine spin-flip transition in the ground state of hydrogen.

While Lyα probes hydrogen that has been locally ionised (either by nearby stars or the background light of an AGN), H I traces neutral hydrogen that must be cold ($T \sim 100$ K), making it a direct probe of the larger-scale gas reservoir responsible for star formation. This means these two spectral signatures are not always correlated (Le Reste et al., 2025), though both are incredibly important to our understanding of the gas in galaxies, active or otherwise. Moreover, H I is in general an excellent tracer of the ISM in active galaxies, since we can exploit the radio brightness of AGN jets to search for H I in absorption out to arbitrarily high redshifts with powerful enough AGN (Morganti & Oosterloo, 2018). It has also been used explicitly to trace cold, jet-driven outflows (e.g. Mahony et al., 2013; Aditya, 2019), and to detect the cold component of ionised outflows (e.g Emonts et al., 2005), reinforcing their multi-phase nature.

Once H I absorption is detected, we can derive a measure of the column density of neutral hydrogen (N_{HI}) directly:

$$N_{HI} = 1.823 \times 10^{18} \frac{T_s}{f} \int \tau(v) \delta v \tag{1.2}$$

where T_s is the spin temperature in Kelvin (a measure of the relative population of hydrogen atoms in each of the two energy levels), f is the covering fraction (how much of the radio continuum is subtended by the gas),

τ is the optical depth (strength) of the line and Δv is its width in $\mathrm{km\,s^{-1}}$. Unless it is possible to constrain the covering factor using high resolution imaging, or the spin temperature using emission (e.g. Nguyen et al., 2024), we typically fix these at 100 K and 1 respectively, noting that there is some degeneracy between the two while they remain unconstrained.

In this way both spectroscopic and continuum observations, and especially radio spectroscopic and continuum observations, reveal important details about not just the Active Galactic Nucleus, but the Active Galaxy as a whole.

1.1.6 Active Galaxies in the radio sky

As mentioned above, the birth of radio astronomy was crucial to the initial discovery and identification of AGN; it revealed both extended, linear structures (now known to be jets and lobes) and compact regions of exceedingly high brightness temperature (now known to be cores or, more precisely, smaller jets). The extended lobes of radio AGN were even observed more than a decade before they earned the name, identified only as 'discrete source[s] imposed upon a smooth continuum' in the absence of higher resolution information (Hey et al., 1946; Bolton et al., 1949).

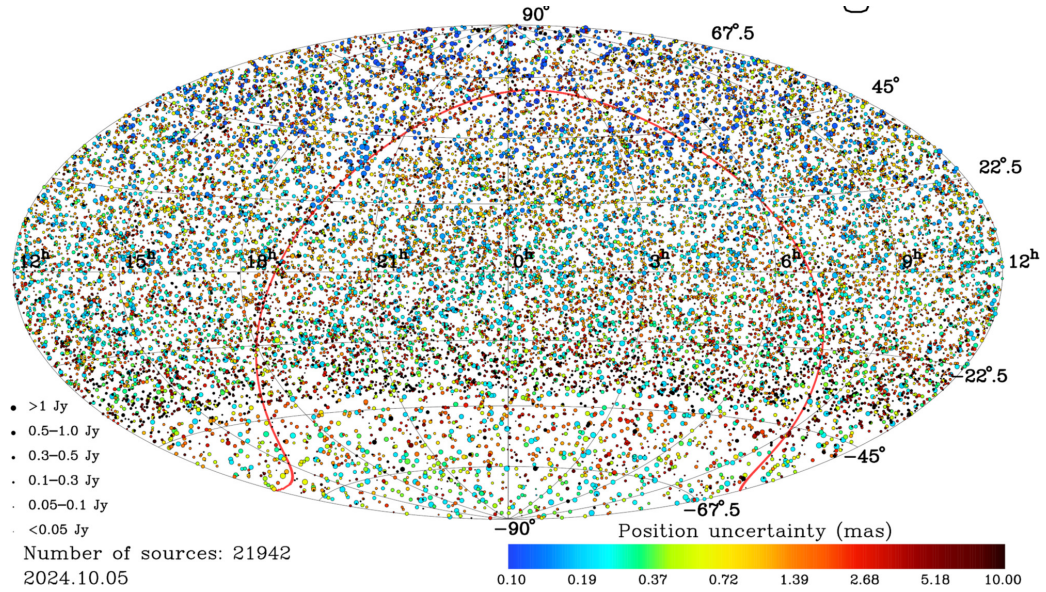


Figure 1.5: The sky distribution of (very) bright radio AGN monitored with Very Long Baseline Interferometry (VLBI) at gigahertz frequencies. From Petrov & Kovalev (2025).

Just as radio astronomy is crucial to the discovery of AGN, radio AGN play a central and critical role in radio astronomy today. Essentially, this is because they are ubiquitous as shown in Figure 1.5; the energy ejected from a SMBH via its synchrotron jets so far exceeds the continuum emissions of most extragalactic objects that AGN dominate the radio sky above $\sim$ 1 mJy at 1.4 GHz (Padovani, 2011). Sources that appear below this threshold tend to be dominated by star formation processes (particularly non-thermal synchrotron from supernovae shocks), though star formation contributions are present even in bright sources and can typically only be separated from AGN contributions with sufficient spatial resolution and brightness temperature arguments (Morabito et al., 2022a). Nevertheless, compact radio AGN are the primary means by which we calibrate our radio telescopes (e.g Bonato et al., 2019; Polisensky et al., 2024), and they form the basis of the International Celestial Reference Frame, by which we track everything from the proper motion of stars to the movement of Earth's tectonic plates (de Witt et al., 2022). This means that not only are radio AGN critical to all other astronomical measurements, but to earth observation, spacecraft navigation and ground-based satellite navigation also.

Beyond their use for calibration, radio AGN are incredibly interesting in their own right as tracers of their multiscale environment. High resolution observations from Very Long Baseline Interferometry (VLBI) can be used to probe the innermost regions of the jet closest to the AGN core which, combined with time domain and multiwavelength monitoring, can be used for inference regarding the accretion rate of the SMBH (Blandford et al., 2019). As discussed in Section 1.1.5, radio jets have been observed driving multiphase outflows of gas,

and their feedback is regularly invoked as a driver of large scale evolution (Hardcastle & Croston, 2020). In particular, Kondapally et al. (2023) show through a careful comparison of observations and simulations, that the jets dominate the AGN energy budget at $z \leq 2$, suggesting their importance in the overall scheme of AGN feedback. At the largest, group to cluster scales, the morphology and orientation of the jets can also be a tracer of ongoing merger activity (e.g. Hardcastle et al., 2019), or even of the cosmic web (Jung et al., 2025). When the mechanism for jet production switches off, radio remnants or relics can persist for $\sim 10^8$ years, and may provide information about the non-thermal intra-cluster medium (Slee et al., 2001).

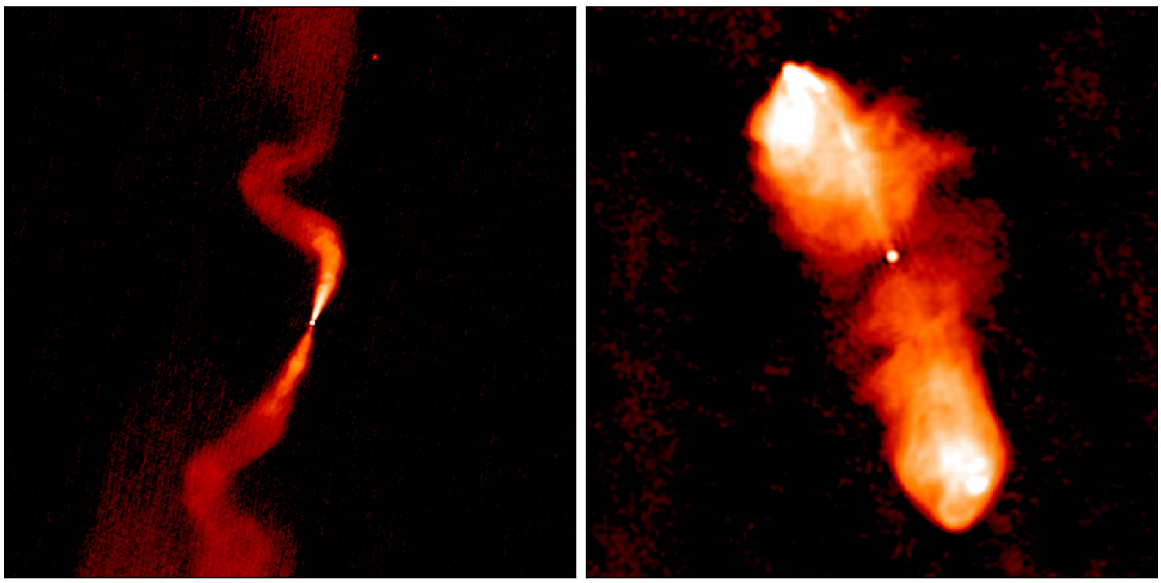


Figure 1.6: Classic core-dominant FR I (3C 31) and FR II (3C 98) type radio AGN. Originally from Robert Laing and the 'Atlas of DRAGNS', reproduced here from Hardcastle & Croston (2020)

When an active galaxy is sufficiently close - or our telescope resolution is sufficiently high - the jet structure can be resolved and is usually grouped under the broad categories of FR I and FR II. Named after Fanaroff & Riley (1974), this classification divides AGN into those with core-dominant, strongly collimated jets (FR I), and those in which the jets are edge-brightened and disperse into diffuse lobes (FR II), as shown in Figure 1.6. In the original sample used for this classification, the radio jets were overall extremely bright (only 2 of the 57 had luminosities $L_{178} \leq 5 \times 10^{24}\,\mathrm{W\,Hz^{-1}}$), and this dichotomy also traced a separation in 178 MHz luminosity, with FRI sources predominantly found at $L_{178} \leq 1.4 \times 10^{26}\,\mathrm{W\,Hz^{-1}}$ and FR II sources above (rescaling the original results to a cosmology where $H_0 = 67.4\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ following Planck Collaboration 2020). However, more recent surveys with higher sensitivity have revealed a population of low luminosity FR IIs thought to be a mixture of low-power radio jets residing in low mass host galaxies, and old jets fading without a fresh injection of relativistic particles (Mingo et al., 2019). Other classes have also emerged to describe more complex (and rare) morphologies such as 'double double' and 'wide-angle tail' radio galaxies (Owen & Rudnick, 1976; Schoenmakers et al., 2000). More recently Baldi et al. (2015) also introduced the 'FR 0' nomenclature to describe the compact sources, typically no more than 1-3 kpc in size, which dominate the radio sky at mJy flux densities. Many of these other classes can generally be thought of as variations on the FR I/II morphology, brought about by jets which are too compact to be resolved, by multiple, discrete epochs of emission, or by structures in the local environment causing deviations of the jet path from a straight axis. The FR I/II dichotomy though, is generally thought to delineate between different jet dynamics. While all radio jets are believed to be launched as relativistic streams of particles, simulations suggest that FR Is decelerate gradually over kiloparsec (i.e. galactic) scales, whereas FR IIs remain relativistic until they terminate in strong shocks which heat the intergalactic medium (Zanni et al., 2005; Perucho et al., 2022; Bhattacharjee et al., 2024). Theoretical frameworks typically predict that both the intrinsic jet power and the local environment density will determine whether a particular AGN exhibits FR I or FR II-type jets (Bicknell, 1995; Tchekhovskoy & Bromberg, 2016). The key takeaway for this thesis though, is that all AGN jets are thought to begin spatially in much the same way, as a relativistic beam of particles. What we are most concerned with is how they begin temporally - or what newly formed AGN jets look like.

1.2 The birth of radio AGN - Peaked Spectrum sources

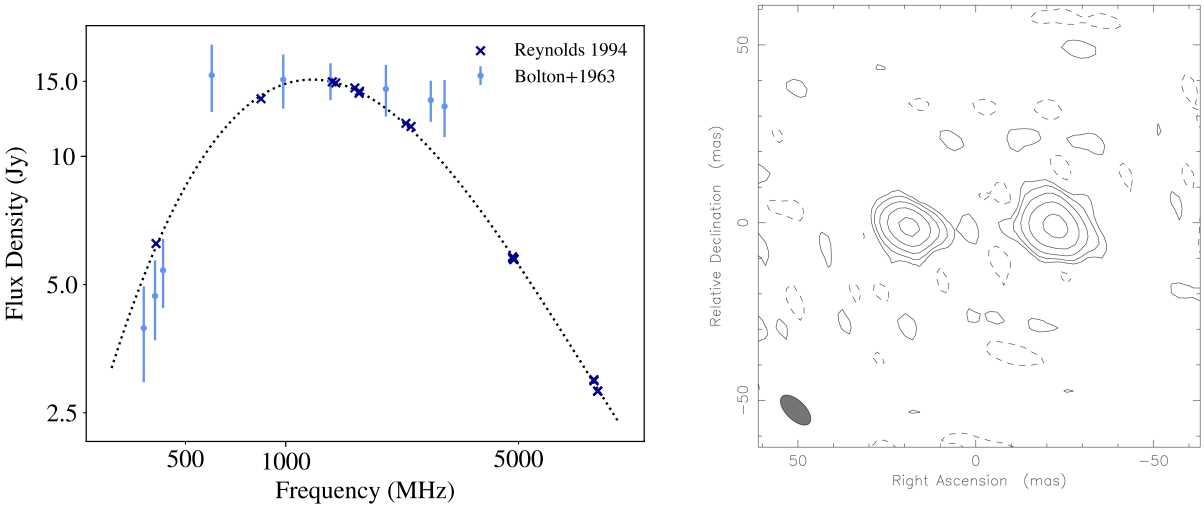


Figure 1.7: Left: broadband radio SED of PKS 1934-638 upon first detection (Bolton et al., 1963) and some thirty years later when it was monitored closely to calibrate the radio flux density scale for the Australia Telescope Compact Array (Reynolds, 1994). The dotted line represents the polynomial fit derived from the 1994 monitoring campaign, which is still in use today. Right: VLBI image showing the two component structure of PKS 1934-638, from Tzioumis et al. (2010).

In 1962, Kellermann et al. (1962) noted eleven sources from amongst the Third Cambridge Catalogue of Radio Sources for which the spectral index "show[ed] a considerable dependence on frequency". From amongst these eleven, two showed a local maximum in their spectra. This was the earliest identification of what are today known as "Peaked Spectrum" (PS) sources, radio AGN which exhibit a flux density that turns over at some characteristic frequency $\nu_{\rm peak}$ proportional to their linear size (Orienti & Dallacasa, 2014; O'Dea & Saikia, 2021). Almost immediately, the importance of PS sources was recognised because of their influence on source counts (and hence, on cosmological tests) at different wavelengths (Bolton et al., 1963) and also because of the potential physics revealed by this spectral curvature (Kellermann, 1965).

The characteristic curvature of a PS source is shown in Figure 1.7, left, which presents the broadband radio spectral energy distribution (SED) of perhaps the most famous PS source in the Southern Hemisphere: PKS 1934-638, an archetypal PS source peaking at $\nu_{\rm peak} = 1.4\,$GHz with an incredibly stable spectrum over time, and the primary calibrator for many southern hemisphere radio telescopes (Heywood et al., 2020). A convex spectrum and its stability over time are, essentially, the sole defining characteristics of a PS source, though as we will see shortly, the picture is more complicated than might first appear.

High resolution imaging of PS sources usually shows them to be compact, confined to scales of no more than a few kiloparsecs, and therefore firmly embedded in the gaseous medium of their host galaxy (O'Dea, 1998; Cui et al., 2010). Indeed PKS 1934-638 is itself only $\sim 120\,$pc in size based on the 40 mas separation between its two components identified in VLBI imaging (Figure 1.7, right) and its redshift of $z = 0.183$ (Penston & Fosbury, 1978). Multiwavelength studies of emission lines have shown that under such circumstances these small radio jets are capable of driving multiphase outflows of gas, directly and powerfully affecting the environment of their host galaxy (Holt et al., 2008; Jarvis et al., 2019). And this influence is bidirectional; the dense environment in turn causes the broadband spectral peak which defines these objects, either by inducing synchrotron self-absorption in the jets themselves, or free-free absorption in the nearby, ambient gas which acts as an absorbing screen (Bicknell et al. 1997; O'Dea et al. 1991; Tingay & de Kool 2003). Their small size, coupled with the slow but well-documented growth of their jets over decades, means that PS are commonly thought to be the young, compact precursors to massive, radio-loud AGN like Centaurus A (e.g. Fanti et al., 1990; Dallacasa et al., 2000).

Furthermore, since these systems are so compact, any galactic absorbing material is typically found in very close to the supermassive black hole at the centre of the AGN, and may also cover the accretion disk,

making many PS not just young but also often obscured AGN (Hickox & Alexander 2018, for the coincidence of obscuration and a PS classification, see Patil et al. 2022 and O'Dea & Saikia 2021). Therefore, a thorough understanding of this particular sub-class of radio AGN has the potential to reveal insight into not only the triggering and evolution of radio AGN more broadly, but also into the gaseous environments in which they reside, particularly the narrow-line region and interstellar medium of their galactic host. In order to study PS sources though, we must first be precise about what types of objects fall into this class, since the boundaries of any phenomenological classification scheme are bound to shift with time and the growth of datasets.

1.2.1 A question of definition

A PS source is identified solely on the basis of its convex, time-invariant spectrum, and other properties like its compact size and young age are an observational consequence of this selection criterion. However the defining broadband peak is typically measured in the observer's frame, and so some (significant) fraction of PS sources could appear to have a steep spectrum without a turnover, due to the confounding effects of redshift (which will draw the peak down to lower frequencies in the observer's frame) and a limited observing bandwidth (so that a peak of 500 MHz would be missed in surveys at and above 1 GHz). Since historical radio surveys had limited bandwidth, a divide formed along these lines, separating 'Gigahertz Peaked Spectrum' sources with, as the name implies, $\nu_{\rm peak} \gtrsim 0.5$ GHz, and 'Compact Steep Spectrum' sources peaking below this (Fanti et al., 1990; O'Dea et al., 1991). This separation was originally motivated by the difference in spectral shape (which we now know can be heavily influenced by observational constraints) and by the fact that Compact Steep Spectrum sources are generally larger ($\sim 1-20$ kpc) than their Gigahertz Peaked Spectrum cousins ($\lesssim 1$ kpc) (Fanti et al., 1985). We now know this second discriminant is a natural and inevitable result of the proportionality between linear size and turnover frequency (Orienti & Dallacasa, 2014), which is itself thought to be a consequence of the density profile of the environment (e.g. Bicknell et al., 2018). Indeed joint analysis has revealed a smooth transition between the two sub-classes in the size-$\nu_{\rm peak}$ plane (Lister, 2003), so it is not surprising that in the most recent and comprehensive review of PS sources by O'Dea & Saikia (2021), these two historically separate subclasses were discussed as one, unified class. We follow this convention in this thesis; if further support for their unification were needed, we might consider the radio AGN phase space of linear size and luminosity as shown in Figure 1.8, where the GPS and CSS classes span an overlapping range of sizes as identified by Lister (2003), and a similar range of luminosities.

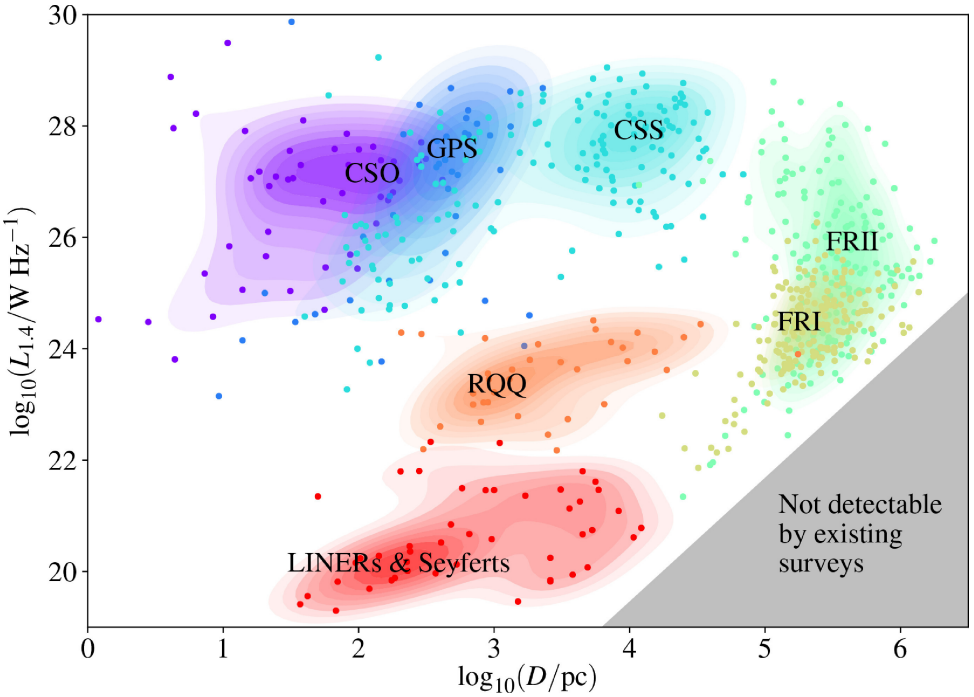


Figure 1.8: The size-luminosity phase space of radio AGN. Our main sources of interest are the PS sources, indicated in this plot by the blue 'GPS' and 'CSS' contours, collectively. The compact symmetric objects (CSOs) are likely related. Radio quiet quasars (RQQs), Low-Ionization Nuclear Emission-line Regions (LINERs) and Seyferts are other classes of AGN not discussed in this thesis. Reproduced from Hardcastle & Croston (2020).

There is another class of source occupying a very similar region of the size-luminosity phase space in Figure 1.8, labelled as 'CSO' for 'Compact Symmetric Objects'. These are defined solely on the basis of two radio components with symmetric flux distributions being present high-resolution VLBI imaging (Phillips & Mutel, 1980). Lister (2003) describes how this classification originally grew out of imaging PS sources, but by the early 2000s it had both diverged from PS sources and itself been diluted by *a*symmetric doubles. Despite these issues around the purity of CSO samples, Kiehlmann et al. (2024a) have sought to revive the classification, arguing that it is "a much more physically motivated classification than PS classifications". I do not wish to dive into the philosophical merits of classification based on SED or morphology here, so I will simply end this section with two remarks. First, over 90% of the bona fide CSO sample meticulously assembled in Kiehlmann et al. (2024a) have a peaked spectrum (the remaining 10% have a steep spectrum with, presumably, a peak below the lowest observing frequency) and second, both CSOs and PS samples suffer from the same contaminant, namely relativistically beamed blazars, which may appear both compact and temporarily peaked due to projection effects, but which are not necessarily true PS sources.

1.2.2 Are all PS sources the progenitors of extended radio AGN?

In short, likely not, though to understand why it is useful to also consider the physical drivers of their characteristic spectra. For as long as PS sources have been studied, there have been two main mechanisms proposed to explain the turnover in their spectrum; synchrotron self absorption (SSA) in the jets themselves, and free free absorption (FFA) in a local, ionised medium, both originally proposed by Kellermann (1965). There is observational evidence for both drivers of spectral turnover derived from different samples of sources, leading to the conclusion that each likely contributes some fraction to the overall population (O'Dea & Saikia, 2021). Running in parallel with this, there is an ongoing debate about whether PS sources are 'young' (i.e. small because the jets have not had time to grow) or 'frustrated' (small because the jets are *prevented from* growing) (e.g Ezeugo & Ubachukwu, 2010; Keim et al., 2019). It is generally argued in observational papers that in the 'youth' scenario, the broadband turnover is the result of synchrotron-self absorption, whereas 'frustration' is related to free-free absorption since it more explicitly requires a dense, ambient environment (Callingham et al., 2015; Zhai et al., 2025, e.g.). However, this recent trend of presenting 'youth' and 'frustration' as mutually exclusive options which map directly onto each absorption mechanism risks confusing two separate, but related, issues.

At a statistical level, Slob et al. (2022) presented evidence that some fraction of the low-luminosity ($L_{144} \lesssim 10^{25}\,\mathrm{W\,Hz^{-1}}$) PS population must never evolve into large-scale radio galaxies, based on an overabundance in their luminosity function relative to that of typical, extended radio sources. Readhead et al. (1996) and An & Baan (2012) independently came to much the same conclusion about a statistical excess of CSOs which do not evolve into larger radio galaxies. However, where samples of CSOs could be kinematically aged by tracking hotspots across multi-epoch images, it was found that even the impeded sources were no more than a few thousand years old (An & Baan, 2012; Kiehlmann et al., 2024b). While the same analysis has not (yet) been applied to PS sources, it seems entirely plausible that it may hold for them too, given the similarities between these populations; PS sources are not young *or* frustrated, but young *and either* able to evolve into classical radio galaxies, or not, in which case they are short-lived. This interpretation is supported by detailed simulations from both Bicknell et al. (2018) and Young et al. (2025), who model their young (and expanding) AGN with free-free absorption to reproduce observed spectral turnover, and secular jet evolution towards extended (~Mpc) structures. Though there are some simulations which allow for confined jets to persist for $\sim 2 \times 10^6$ years (see Mukherjee, 2025, for a good overview), even this is still on the threshold of a 'young' jet (e.g. Morganti et al., 2023). Even in instances where a PS source is embedded in more extended emission, the usual interpretation is that this extended emission is a remnant from a previous epoch of activity which has ceased, and so the small-scale jets producing the broadband spectral turnover are again young and recently triggered (Saikia & Jamrozy, 2009). This point has been obscured in recent observational discussions in the literature and so it is worth reiterating here; *some* PS sources may never evolve into extended radio galaxies like Centaurus A, but they are very likely *all* young, irrespective of whether they exhibit synchrotron-self absorption or free-free absorption.

1.2.3 Peaked spectrum sources probing multiphase gas

Instead of focusing on the youth/frustration dichotomy, it is interesting for our purposes to consider how the broadband spectral turnover in PS sources might inform us about their interactions with the local, multiphase ISM of their host. Section 1.1.5 outlined how radio jets can drive powerful outflows in broad terms, but what changes when the jets are small?

Generally speaking, the typical size range of PS sources (less than a few kpc) suggests they probe a particularly interesting region of the ISM which undergoes dramatic variations in density, structure and composition. Moreover, it has long been thought that many PS sources live in gas-rich environments conducive to jet-gas interactions which may influence these properties (van Breugel et al., 1984). Indeed, detailed spectroscopic study of PS source PKS 1718-649 suggested that secular accretion of cold gas (observed in H I absorption, and H_2 in emission) might have triggered the radio jets in that source (Maccagni et al., 2014, 2016). Moving from fuelling to feedback, observations of PS source PKS 1549-79 revealed the presence of cold, kiloparsec-scale inflows, a nuclear ($\sim$ 120 pc-scale) outflow and a circumnuclear disk seen in CO, where the outflow was thought to be jet-driven due to its compact size (Oosterloo et al., 2019). Similar, cool, molecular outflows have been identified even in PS sources with intrinsically low power jets ($L_{1.4\,\mathrm{GHz}} = 2.1 \times 10^{23}\,\mathrm{W\,Hz^{-1}}$ Murthy et al., 2022). Complex [O III] emission line profiles in the nuclear regions of PS sources have also been interpreted as evidence for jet-driven, ionised outflows (e.g. Holt et al., 2008). Furthermore, small-scale jets have been shown capable of driving outflows in simulations, such as those of Mukherjee et al. (2018b) shown in Figure 1.9. This certainly suggests that small jets have a particular influence on the multi-phase ISM, and vice versa.

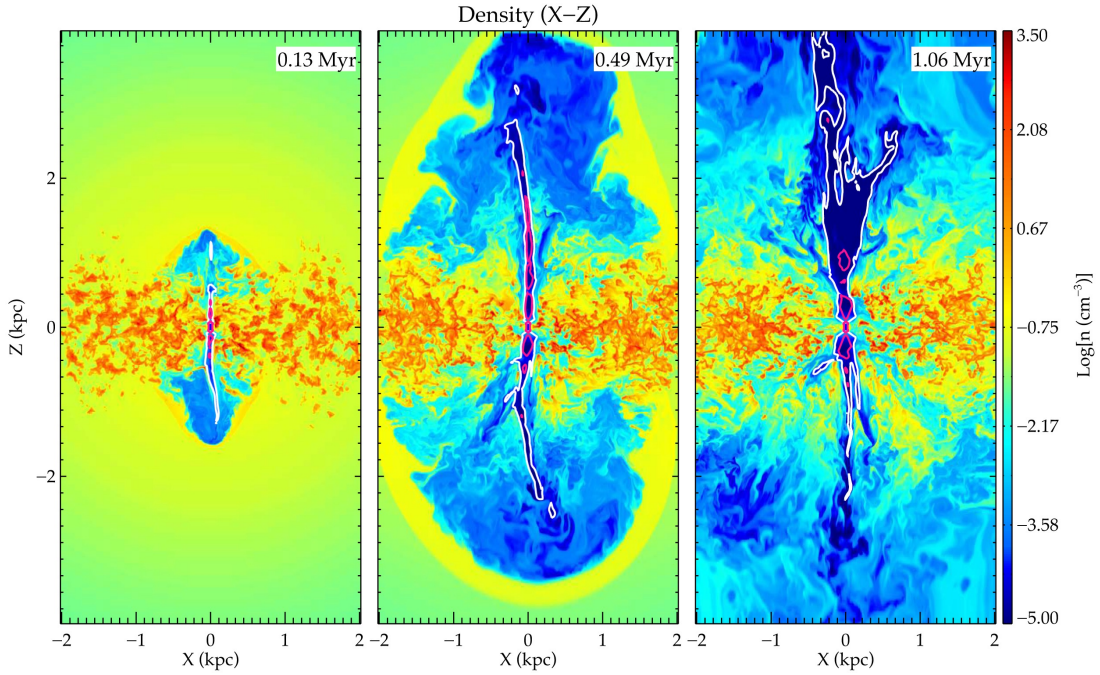


Figure 1.9: Sub-kiloparsec jets driving outflows from a thin gas disk in hydrodynamical simulations. Reproduced from Mukherjee et al. (2018b).

At a statistical level though, there is conflicting evidence as to the strength of jet-gas interactions in PS sources. They are clearly capable of driving strong, cold outflows in H I (Morganti et al., 2005) and CO (Morganti et al., 2023), and Santoro et al. (2020) find that PS sources drive warm, ionised outflows which are more energetic and efficient than those in control samples, evidence for stronger jet-gas coupling when the jets are small. By contrast, Miranda Marques et al. (2025) find evidence of ionised, [O III] outflows at a lower rate than in control samples of extended radio galaxies or radio-weak AGN, and the outflows they do find are no more energetic. In the cold phase, PS sources are a favourite target to search for H I absorption, since they tend to have a higher detection rate than the general population of radio AGN (Vermeulen et al., 2003; Geréb et al., 2014). This is thought again to be due to the rich, multiphase ISM in which they are embedded (e.g. Morganti et al., 2026), though we still lack a detailed study of the H I properties of PS sources (including the possible presence of cold, H I outflows) compared to a matched sample of extended radio galaxies, as has been done in the optical for warm outflows.

One further consideration ought to be incorporated into jet-gas interaction studies, and that is the spectral shape of the PS sources. It seems plausible that a PS source driving multiphase outflows of the ambient ISM might be absorbed by that same medium, leading to free-free absorption being the dominant mechanism responsible for the broadband spectral turnover. Macquart & Tingay (2016) proposed a theoretical framework in which variability in both the broadband radio, and H I 21 cm absorption could be studied to infer properties about a multiphase medium common to both signatures. However, observational tests on individual sources are yet to show significant results, as this hypothesis requires both spectroscopic observations (to identify the multiphase outflow) and multifrequency, single epoch radio continuum observations (to carefully constrain the broadband SED shape). Realistically, large observational samples may be needed to infer the likelihood of such a scenario occurring across the PS population. In order to make progress in this direction then, we must work out how to assemble samples of PS sources that are well sampled in all required parameter spaces.

1.2.4 How we study peaked spectrum sources

As with so many astrophysical objects and phenomena, the first examples of PS sources were unexpected discoveries made as part of a larger observational project (Kellermann et al., 1962). The samples were small (no more than a dozen or so sources) for at least the first decade of work on these objects, and as can be seen in the top row of Figure 1.10, fits made to the data were generally made using "a smooth curve drawn by hand" (Reynolds, 1994). To be clear, that is not to say the astrophysical models were simplistic, in fact Kellermann (1965) identified SSA and FFA as potential drivers of spectral turnover shortly after the initial discovery of PKS 1934-638, and Shklovsky (1965) proposed elegant, analytical formulae to predict the rate at which ν_{peak} would shift downwards in frequency as the source aged, based on the local magnetic field and ambient density. However, fitting these models to the data with enough precision to distinguish between, e.g. SSA and FFA was not yet tractable. This was a shortcoming of computational power, rather than physical understanding, and is something we revisit again in Section 1.4.

Even some two decades after their initial discovery, studies of PS sources grew heterogeneously out of surveys at a range of frequencies and sensitivities, but this led to a disparate collection of samples with seemingly discrepant properties due to selection effects (as discussed in e.g. Fanti et al., 1990; de Vries et al., 1997; Edwards & Tingay, 2004). Therefore neither a combination of these samples, nor an investigation of them in isolation is suitable for unbiased, statistical studies of the young radio AGN population.

It is really only in the last decade that we have seen significant advances in the way that we collect and study these (and indeed many other) radio sources. Advances in computational power and physical models capable of producing broadband absorption have allowed for more rigorous and reproducible modelling of the broadband spectral shape, as exemplified in Callingham et al. (2015) (shown in Figure 1.10, lower left). More flexibility in telescope configurations and an increase in instantaneous bandwidth brought about by improved digital signal processing has allowed for multifrequency and multiscale monitoring of peaked spectrum sources to confirm their time-invariance (Edwards & Tingay, 2004; Tingay & Edwards, 2015) or, in the case of some 1.5 per cent of sources, prove them an exception to the rule (Ross et al., 2021). Perhaps most importantly though, we now have an unprecedented coverage of the radio sky thanks to the latest generation of widefield surveys. This promises breakthroughs in the study of these and many other objects just by virtue of the sheer volume of sources we can now identify and examine. The first such widefield sample was presented in Callingham et al. (2017) (and SEDs of a few of these sources are shown in Figure 1.10, lower right). Five years later, a northern sky sample was presented by Slob et al. (2022) and no doubt many more will follow. For the first time, our samples can be both large ($> 10^4$ sources) and uniform - a prime chance to harness the power of statistics to better understand the physics of these PS AGN.

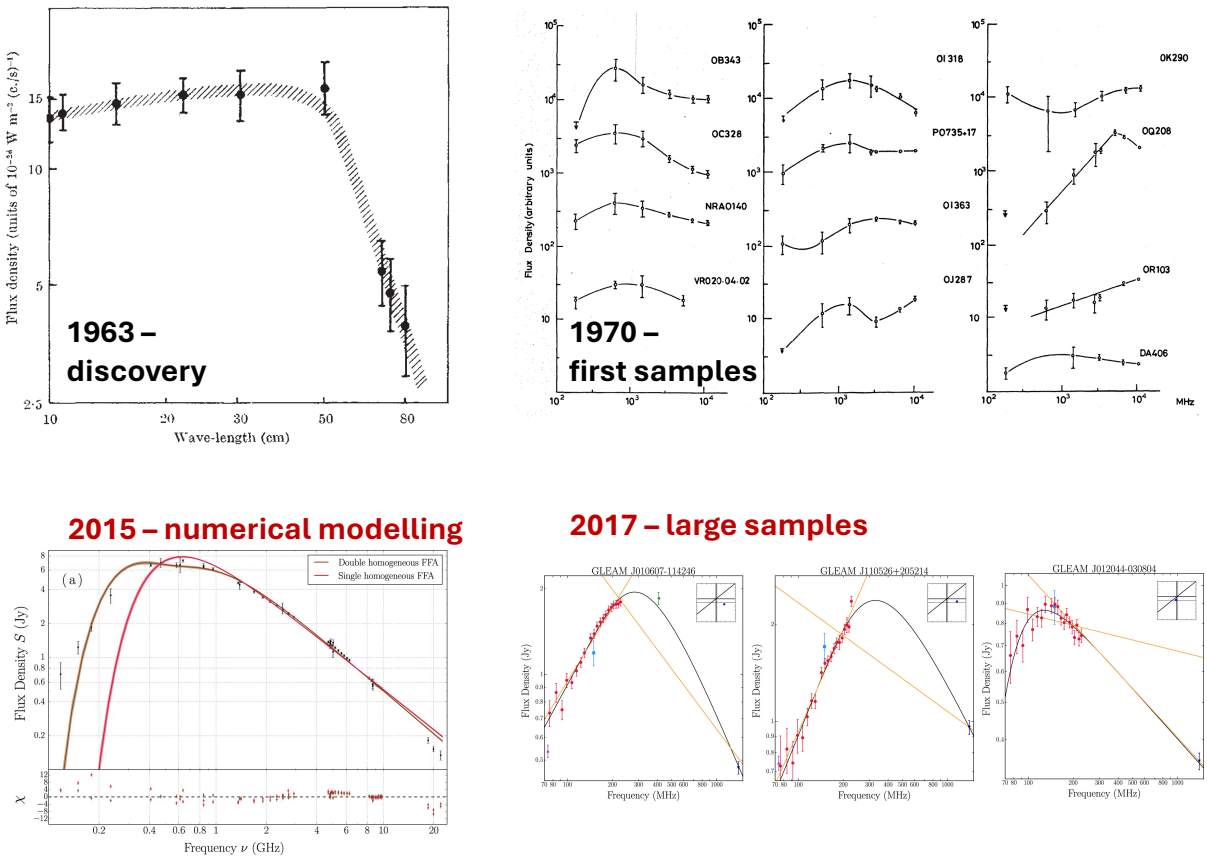


Figure 1.10: The improvements to our identification of PS sources. Top left is Bolton et al. (1963), top right is Blake (1970). Bottom left is Callingham et al. (2015), bottom right Callingham et al. (2017).

1.3 Survey science in radio astronomy

In discussing the gains made possible by large, uniform samples of objects we ought to take a moment to appreciate the considerable time and effort it has taken to reach this point, as we head into what many radio astronomers have a fondness for calling the SKA era, but which is really emblematic of a broader shift towards all-sky surveys across all wavelengths (and now, all messengers too).

The methods by which we do radio astronomy have moved in step with technology, allowing for exponential increase in resolution and sensitivity over the course of many years (Parijskij, 1992). From the inception of the field until about the 1970s, surveys were the gold standard by which new discoveries were made; the radio sky was a new frontier, and like the terrestrial cartographers of old, astronomers sought to map this uncharted territory in order to understand it, cataloguing what they found along the way. The first radio surveys in the early 1950s contained a modest number if sources ($\lesssim$ 100) and were observed at a single frequency with no spectral information (Goss et al., 2023), but they progressed rapidly to well over a thousand across large fractions of sky (Shakeshaft et al., 1955; Mills et al., 1958). Initially, there was debate about the nature of discrete radio sources, whether they were Galactic 'radio stars' or extragalactic in origin, since their optical counterparts were often difficult to identify. This confusion was compounded by seemingly contradictory results presented by the leading radio astronomy groups of the period, and by subtle assumptions built into their instrument design, a good overview of which is provided in Goss et al. (2023), Chapters 35-36. But once it was realised some large fraction of radio sources were extragalactic, they were used as a novel test of cosmological models, with some evidence that the spatial distribution of radio bright AGN was incompatible with the Steady-state theory of the Universe (Ryle, 1955), though interpretations of the results remained uncertain for some time (Kellermann, 1972; von Hoerner, 1973; Goss et al., 2023). In support of this science, radio telescopes evolved from single antennas, to fixed aperture arrays built on the principles of a Michelson interferometer (Ryle & Hewish, 1960), to steerable dishes like the Parkes Radio Telescope (recently renamed *Murriyang* in acknowledgement of the traditional custodians of the region), and the Very Large Array, a 27-dish steerable and reconfigurable interferometer (Thompson et al., 1980; Perley et al., 2009).

With this new generation of steerable dishes and reconfigurable arrays came the ability to survey the sky at higher resolution by switching to a higher observing frequency. Not only that, but a single, steerable dish made multifrequency observations on short timescales tractable for the first time. As a result, Parkes produced the first large area, multifrequency survey and hence the first catalogue to contain radio SED information (Ekers, 1969). It was from this dataset that PKS 1934–638, our archetypal PS source introduced in Section 1.2, was first identified. A follow up survey at 2.7 GHz with Parkes also revealed, for the first time, the diversity of spectral shapes in the radio sky (Wall et al., 1971), reinforcing the idea that PS sources were not merely outliers, but part of a diverse population of extragalactic radio sources. Contemporaneous advances in processing and storage also allowed for the first joint map of the full radio sky at 4.85 GHz using the Green Bank and Parkes radio telescopes, with the NRAO multibeam receiver shuttled between the two (Griffith & Wright, 1993; Condon et al., 1994).

Soon after, surveys at 1.4 GHz and 0.843 GHz were conducted in the northern and southern hemispheres, respectively, exploiting further advances in processing power and deconvolution algorithms that allowed astronomers to make full use of the resolution afforded by an interferometer (NVSS; Condon et al. 1998 and SUMSS; Mauch et al. 2003. Almost a decade later, Murphy et al. (2010) conducted an all-sky survey at 20 GHz with follow-up at 4.85 and 8 GHz using the Australia Telescope Compact Array by exploiting a custom analogue correlator, which is still the only survey of its kind at these frequencies. For the first time, this systematically extended the frequency coverage in the southern hemisphere above 10 GHz, which allowed for the identification of smaller PS sources with higher frequency turnovers (Hancock et al., 2010).

Recently, several new surveys have once again reshaped our view of the radio Universe, improving on both sensitivity and coverage thanks to a new generation of instruments working at the limits of computational and storage capacities. Radio astronomy is returning to its roots, with aperture arrays once more at the forefront of scientific progress thanks to both the Murchison Widefield Array (MWA; Tingay et al., 2013) and the LOw-Frequency ARray (LOFAR; van Haarlem et al., 2013) producing high-resolution, widefield surveys of the radio sky at $\sim$ 200 MHz. These surveys are revealing new classes of objects (e.g. Hurley-Walker et al., 2022b), detecting solar-like radio bursts from other stars (Callingham et al., 2025), and revolutionising our understanding of everything from stars (Driessen et al., 2024) to Peaked Spectrum sources, the focus of this thesis (Callingham et al., 2017; Slob et al., 2022). Recent software upgrades at both the MWA and LOFAR have also vastly improved their processing capabilities without changes to the underlying infrastructure, with the result that even higher resolution surveys are now underway (Hurley-Walker et al., 2022a; Morabito et al., 2022b). In the southern hemisphere, the Australian SKA Pathfinder (ASKAP) is conducting a series of all sky surveys between 0.7-1.8 GHz which are delivered to the community as science ready data products (Hotan et al., 2021). These are already being used for a wide array of scientific pursuits, from identifying fast radio bursts (Wang et al., 2025) and Galactic scintillation arcs (Wang et al., 2021a), to testing the cosmological principle (Oayda et al., 2024; Land-Strykowski et al., 2025) and identifying $z > 5$ radio bright AGN (Ighina et al. 2024, with a comparable search using LOFAR by Gloudemans et al. 2022).

Across the radio spectrum, surveys have been conducted in the continuum, polarisation and spectral line domains for decades, with each new survey bringing a fresh perspective to the field. Already the next generation of instruments is on the horizon, with the SKA-low and -mid in commissioning (Braun et al., 2015) alongside the Deep Synoptic Array (DSA Hallinan et al., 2019), and the next generation VLA (ngVLA) in development. Yet there is still so much science to be done with current instruments and surveys; below I cover in a little more detail those surveys and instruments most relevant to this thesis in the realm of both continuum and spectral line science.

1.3.1 The Australian SKA Pathfinder

The Australian SKA Pathfinder (ASKAP, Figure 1.11) is an SKA precursor situated at Inyarrimanha Ilgari Bundara, the CSIRO Murchison Radio-astronomy Observatory in Western Australia, where its main focus is on efficient, widefield surveys of the southern sky at ~1 GHz in continuum, spectral line, and polarisation modes (Hotan et al., 2021). The telescope became fully operational in 2019, and since then it has been observing continuously under nine separate Survey Science Projects, as well as producing data for observatory-run projects, and guest science proposals.[2]

[2] see https://www.atnf.csiro.au/projects/askap/index.html for a list of current projects

Figure 1.11: The Australian SKA Pathfinder. Credit: CSIRO.

ASKAP comprises 36×12 m antennas with a maximum baseline of ~ 6 km. Each antenna is fitted with a Phased Array Feed (PAF) which produces a $\sim$30 deg^2 field of view at a spatial resolution of 6-15 arcsec depending on the observing frequency (to put this in context this is about 3 times the FoV of the Rubin Observatory,[3] and 10,000 times the FoV of WFC3 on the Hubble Space Telescope: Ryon 2021). The telescope operates within three fixed frequency bands centred on 887.5 MHz (band 1), 1367.5 MHz (band 2) or 1655.5 MHz (band 3), with a maximum bandwidth of 288 MHz and standard spectral resolution of 18.5 kHz. The wide field of view, excellent spectral resolution, sensitivity below 1.4 GHz and radio-quiet location all make ASKAP an excellent instrument for probing the continuum structure of PS sources, and their gaseous properties via HI absorption searches.

1.3.2 Continuum surveys used in this thesis

Thanks to its wide FoV and excellent instantaneous sensitivity, ASKAP can easily live up to its brief as a survey telescope. Its time is currently divided between nine major science surveys, alongside additional fast, month-long surveys of the entire southern sky across each continuum bands, in an observatory-led project known as the Rapid ASKAP Continuum Survey (RACS; McConnell et al., 2020). The first pass of this survey in the low ASKAP band ($\nu_{\rm centre}$ = 888 MHz) achieved an RMS sensitivity of < 0.6 mJy/beam, almost a factor of two improvement over SUMSS in a fraction of the time, and provides the first complete census of southern sky radio sources since that survey, cataloguing over 2.1 million discrete sources. It is therefore critical to the identification of our active galaxies of interest; Peaked Spectrum sources.

In addition to RACS, this thesis also makes use of Variable And Slow Transients survey from ASKAP (VAST Murphy et al., 2013), which provides radio light curves for all continuum sources within its footprint. These will be key to our understanding of the variability in the PS sources we study. Beyond ASKAP, this thesis benefits from the recent advances in MWA processing discussed in Section 1.3. As we will see in Chapter 3, the The GaLactic and Extragalactic All-Sky MWA Survey - eXtended (GLEAM-X; Hurley-Walker et al., 2022a) will be critical to the way in which we constrain the turnover and low frequency absorption in our PS sources. In addition to GLEAM-X and once more in the time domain, this thesis makes use of the MWA interplanetary scintillation (IPS) catalogues (Morgan et al., 2022). IPS is variability on short, second-long timescales brought

[3] https://rubinobservatory.org/for-scientists/rubin-101/key-numbers

about by density variations in the solar wind. These density variations cause phase shifts in light passing through the solar wind, shifts which are most obvious when the background source is compact. For this reason, IPS measurements can be used as a powerful probe of the structure of radio sources beyond the resolution limit of the observing telescope, making them interesting for us to help constrain the size of the emitting region in PS sources.

1.3.3 Spectral line surveys used in this thesis

At radio wavelengths, the 21 cm hydrogen line is by far the most common atomic transition studied, especially at extragalactic distances. It was also a primary motivator for the SKA when it was proposed in 1990 (Ekers, 2012). Produced by a classically forbidden spin-flip transition, it has an exceedingly low spontaneous transition probability ($A_{10} = 2.85 \times 10^{-15}\,\mathrm{s}^{-1}$), but hydrogen is so ubiquitous throughout the Universe that it serves as an excellent tracer of both Galactic (e.g. McClure-Griffiths et al., 2023) and extragalactic gas (e.g. Dutta, 2019). Indeed there is even hope that it will be used to detect the elusive epoch of reionisation, when the first stars reionised the cool, dark intergalactic medium at $z \sim 6 - 25$ (Furlanetto et al., 2006). Furthermore, several ASKAP surveys are searching for H I across different cuts of redshift space, from our own Milky Way (Dickey et al., 2013), all the way out to $z = 1$ (Allison et al., 2022). It is the highest redshift survey of these, the First Large Absorption Survey in HI (ASKAP-FLASH), which concerns us here.

ASKAP-FLASH is an untargeted, all-sky ($|b| > 8.5°$) search for H I in absorption at $z = 0.4 - 1$ (Allison et al., 2022; Yoon et al., 2025). Each pointing is observed continuously for 2 hours, reaching a median rms of 90 μJy/beam in the continuum images, and ∼6 mJy/beam/channel in the spectral cube. At this distance, searching for H I in emission would be prohibitively expensive with ASKAP, requiring many hundreds of hours per pointing to reach any appreciable sensitivity. Instead, ASKAP-FLASH is looking, as the name suggests, for absorption towards radio bright continuum sources, which must be at $z > 0.4$, and are therefore exclusively radio AGN. These radio AGN act as 'backlights', illuminating foreground neutral gas either localised in their own host galaxy (and hence, 'associated' H I), or in some intervening region along the line of sight. So, for each of the 300 sources at $S > 30$ mJy in a 30 deg^2 ASKAP field, a 1-dimensional spectrum is extracted, which can then be searched for H I absorption features using the automated, Bayesian FLASHfinder tool (Allison et al., 2012). As discussed in Section 1.2.3, PS sources show a high incidence of both jet-driven outflows, and H I absorption specifically in targeted searches. ASKAP-FLASH will therefore be critical to understanding how widespread this phenomenon truly is, and whether there is any selection bias at play in current samples. In turn, PS sources will, as we shall see, be critical to understanding the neutral gas content of the Universe at these redshifts, through the lens of ASKAP-FLASH.

1.4 Archival science and the shift towards big data

Radio astronomy is no longer a young science; we have archives stretching back over 80 years and they are growing exponentially. When used judiciously, they can still provide important new insights, such as a long-period transient that has been emitting sporadic pulses since 1988 (Hurley-Walker et al., 2023), or a PS source which lost its curvature due to evolution (Ross et al., 2021). But with the ever-growing data volumes of the latest generation of telescopes (ASKAP averages ∼10s of petabytes a year) and the absolute deluge promised by the next (current SKA estimates sit at ∼100s of petabytes a year), we need to start planning now for how to not only manage this data, but *continue doing science with it*. In many ways we are lucky that the first half of this quandary has been anticipated for some time; our file formats were standardised in 2001 with the Flexible Image Transport System (FITS; Pence et al., 2010), and we have a robust Virtual Observatory framework that standardises access protocols (Djorgovski & Williams, 2005). What remains is the question of how we continue to use the data we already have whilst making the most of the data yet to come.

Already in the study of PS sources, the ability to numerically fit models capturing different physically-motivated absorption scenarios has changed the way we think about SSA and FFA in these sources, as discussed in Section 1.2.4, and the ability to quickly fit functional models to many thousands of sources is what enabled the first, large sample of Callingham et al. (2017) alongside the new data from the MWA. More broadly across astronomy and indeed all physical sciences, advances in computation power and storage have made new numerical methods tractable, opening up the 'big data' space where discoveries are made by collating and analysing $\gg 10^4$ data points, thanks largely to the power of random sampling and regression, which power

many complex machine learning methods in use today (as explained in e.g. the introduction to deep learning by Goodfellow et al., 2016). To these we might add Bayesian inference, which benefits greatly from random sampling as well, and provides a framework for encoding prior information in statistical analyses (von Toussaint, 2011). This last tool is in frequent use within the ASKAP-FLASH team thanks to the Bayesian FLASHfinder, which is used to filter all incoming ASKAP-FLASH spectra and identify candidate absorption lines, each one reported with a robust measure of significance (Allison et al., 2012).

In a recent white paper Hogg (2026) described the 'data-scientification' of astronomy, whereby a good astronomer might know as much about database queries and code optimisation as she does about conducting an observation and calibrating her data. This is, I think, tied closely to the great oceans of data that we have collected and maintained in readily-accessible archives. In order to get the most out of our datasets, both archival and fresh off the fibre optic cables, we need to be mindful of what already exists, and how these datastreams new and old can be combined in novel and, most importantly, reproducible ways. This philosophy, as you will see, informs large portions of this thesis.

1.5 The role of simulations in an observer's toolkit

Observational radio astronomy is not the only field to benefit from the exponential growth of both computing capability and storage, far from it. But where this progress has led to (vast) improvements in our imaging and analysis, it has fundamentally changed the tools available to a theorist. By the 1990s, it was possible to study dark matter structure and evolution in large scale (~megaparsec) boxes with N-body simulations (Kravtsov et al., 1997), a feat impossible analytically. Just a few years later, cosmological simulations revolutionised theoretical astrophysics thanks largely to the GADGET code, capable of running hydrodynamical simulations that evolve both collisionless dark matter, and collisional baryonic structures (Springel et al., 2001). This had its own kind of feedback on observational astronomy; it allowed theorists to identify the necessity of AGN feedback in quenching star formation (Di Matteo et al., 2005; Croton et al., 2006; Sijacki et al., 2007), test different models of exotic dark matter (e.g. Mocz et al., 2019), and reproduce dozens of observational phenomena (Vogelsberger et al., 2020).

But what does this mean for an observational astronomer? Evidently insights from simulations feed into how we design our observational experiments, and vice versa, but this process is not as straightforward as one might hope. The limitations inherent in computational methods can manifest in subtle effects not always obvious to an observer (e.g. Pontzen et al., 2021; Lancaster et al., 2025). So care must be taken when observers (like this author) attempt to interpret the results from simulations as we would the data coming from our telescopes. Furthermore, simulations deal in different fundamental quantities to those considered by an observer, so even relying strictly on physical quantities can be challenging; where a theorist might directly measure an outflow in terms of an energy flow rate, dE/dt, or perhaps mass dM/dt, an observer will measure velocity v from a redshifted spectral line, and make assumptions about the geometry of the gas and its multiphase composition (recalling that individual atomic transitions will only give us information about one phase) to *derive* a mass outflow rate dM/dt, then make additional assumptions about the relative velocity contributions from turbulence and radial motion to finally derive a comparable energy outflow rate, dE/dt (e.g Holt et al., 2006). Similar assumptions would need to be made by the simulator to reach a 1-D velocity spectrum from their fundamental quantities of energy and mass. Of course this can be done, and has been to great success (e.g. Chisholm et al., 2017; Ward et al., 2024), but it does mean that all assumptions that go into comparisons between simulated and observed datasets should be rigorously examined and systematically documented alongside any conclusions derived from the analysis.

One useful and productive avenue for more direct comparison between simulations and observations is through the use of mock observations, or "mocks". These mocks are produced by post-processing simulated datasets of any kind, from cosmological to idealised simulations and semi-analytic models, to produce observable quantities (Haworth et al., 2018). Often, mocks will be tailored to match a specific instrument by folding in instrumental noise, systematics or other known effects, to allow for fair comparison between datasets. This can uncover systematic biases in observational techniques (Torrey et al., 2015), and analysis of mocks themselves can reveal correlations between observables and the underlying physics, which can be applied back to new observations (Sommovigo et al., 2025). It is with this careful approach that this thesis attempts to incorporate a meaningful analysis of simulations, in the context of our new, untargeted surveys.

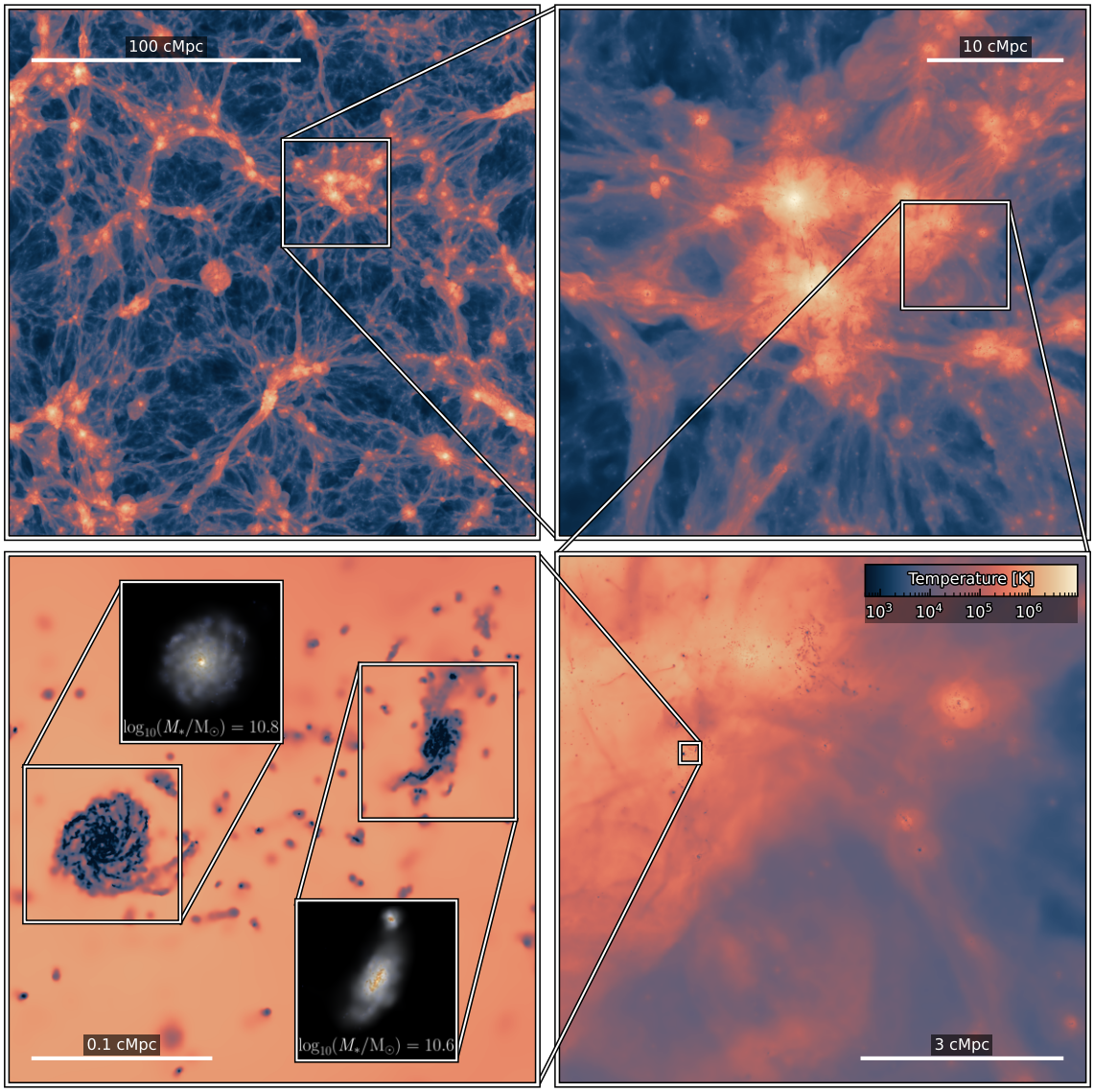


Figure 1.12: The dynamic range achieved in the COLIBRE cosmological simulation, with all panels coloured by the mass-weighted mean of the gas temperature in a slice 25 cMpc thick. The images in the lower left panel show post-processed mocks designed to match the *Euclid* photometric bands, and accounting for dust attenuation. Reproduced from Schaye et al. (2026).

1.5.1 Cosmological simulations

In the context of large-area observational surveys, perhaps one of the most useful theoretical comparisons to be made is with cosmological simulations. These simulations are typically either hydrodynamical (treating baryons as a collisional fluid) or semi-analytic (N-body dark matter onto which baryonic structure is later affixed), and they have large box sizes of $(\sim 100\,\mathrm{cMpc})^3$. With a box of this size, they can track $\sim 10^4$ galaxies from inception in a primordial density field through to some later redshift, capturing mergers, interactions (e.g. Jhee et al., 2026), and evolving large scale structures within the cosmic web (e.g. Cui et al., 2018), as seen in Figure 1.12.

These simulations are so complex and computationally expensive that they require large, international teams to produce, much like the very biggest observational surveys. Over the last ten years several cosmological hydrodynamical simulations have been released including Horizon-AGN (Dubois et al., 2014), EAGLE (Schaye et al., 2015), IllustrisTNG (Pillepich et al., 2018), SIMBA (Davé et al., 2019) and COLIBRE (Schaye et al., 2026), each with their own specific implementations of cosmology and physics. Even with recent advances in computational power though, some physical processes will always be below the resolution limit of a big-box hydrodynamical simulation; these typically include star formation, black hole growth, stellar or AGN feedback, and cold gas physics, all of which must be implemented as sub-grid models. Such sub-grid models must be

chosen carefully, as they can have profound impacts on the large-scale results of a simulation (for a good review of the challenges involved see Crain & van de Voort 2023, and for cold gas specifically Butsky et al. 2024). Nevertheless, cosmological simulations are incredibly useful tools provided we remain aware of these subtleties.

The large teams required for cosmological simulations and observational surveys are not their only similarity, the two are also particularly well suited for comparing population-scale physics and relationships. Since cosmological simulations evolve so many galaxies simultaneously, large samples of different kinds can be extracted from them to compare with the stellar mass - infrared luminosity relation for galaxies (Aoyama et al., 2019) or the radio luminosity function of AGN (Thomas et al., 2021) for example. Furthermore, since the evolution of these galaxies is tracked across cosmic time, it is even possible to create mock light cones which mimic a full field observation (e.g. Merson et al., 2013; Yang et al., 2021; Thomas et al., 2025). Overall then, cosmological simulations and in particular mock observations of them, are an incredibly useful tool for understanding observational trends seen at the population level.

1.5.2 Simulating radio AGN

Given the focus of this thesis, it is useful to consider here how well these simulations are able to match observations of radio AGN in particular. There are a number of idealised simulations of jets evolving through a multiphase ISM. These have demonstrated how jets excavate a central cavity in the gas (Dugan et al., 2017; Mukherjee et al., 2018b), inducing star formation in ring-like structures at larger radii (Gaibler et al. 2012, matched observationally by Zinn et al. 2013). These idealised simulations also show jets can ionise and perturb the ambient ISM in a manner consistent with observations (Mukherjee et al., 2018a; Meenakshi et al., 2022), suggesting they are capable of driving multi-phase outflows. Typically these simulations are (magneto)hydrodynamical, with varying levels of detail for the implementation of cosmic ray transport and radiative transfer (for a review see Mukherjee 2025), but overall idealised simulations are capable of reproducing many observed jet-gas interactions.

In terms of radio signatures specifically, idealised simulations of individual systems can again be well-matched to observations, often with the aid of semi-analytic models for synchrotron emissivity and broadband absorption (such as Turner et al., 2018). Of particular relevance to this thesis, Bicknell et al. (2018) and Young et al. (2025) have both produced young radio jets fully embedded in an ISM which exhibit a broadband spectral turnover due to free-free absorption, the characteristic signature of a PS source, and have reproduced the anti-correlation between turnover frequency and linear size as these jets propagate out towards an eventual breakout stage. The SED of one such simulation, drawn from Young et al. (2025), is shown in Figure 1.13, left. Furthermore, Young et al. (2025) have shown that some double-double sources (usually interpreted observationally as 'restarted') can be produced by one epoch of jet activity within a clumpy ISM. In addition to these, simulation-based semi-analytic models which build upon idealised simulations can reproduce the observed radio signatures of jets at a range of different lifetimes and powers, and in a range of different large-scale environments (e.g. Hardcastle, 2018; Turner et al., 2023).

Unfortunately cosmological simulations do not offer the same careful treatment of the jets seen in idealised simulations. Since Croton et al. (2006) it has become standard to implement two modes of AGN feedback corresponding broadly to 'quasar-driven' and '(radio) jet-driven' feedback; in that work one mode accretes local, cold gas for the 'quasar' component, and one heats the hot gas for the 'jet' component. However, there the jet-mode feedback was only switched on at late times in massive, passive galaxies, and even in modern cosmological simulations, jet-mode feedback is not always included; EAGLE does not implement it at all (Schaye et al., 2015), nor does the flagship run of COLIBRE, though smaller runs do (Schaye et al., 2026). Most crucially though, even where jet-mode feedback is included (often as an injection of kinetic energy e.g. in IllustrisTNG and SIMBA), the radio emission from these 'jets' is not provided as part of the auxiliary datasets, which typically only include UV-optical-IR SEDs for the simulated galaxies (see Trčka et al. 2022; Gebek et al. 2024 for IllustrisTNG, Davé et al. 2019 for SIMBA). Nevertheless, Yates-Jones et al. (2023) have used cluster-scale environments from cosmological simulations as a starting point for zoom-in simulations of radio jets, demonstrating the strong impact of environment on observable radio relics (Stewart et al., 2025, 2026). Furthermore, Thomas et al. (2021) were able to successfully reproduce the total 1.4 GHz radio luminosity function out to $z \sim 3$, and the radio luminosity function for LERGs at $z \sim 0$ in SIMBA by adopting a sensible prescription for radio luminosity. This was based on the properties of jet-mode feedback tracked within the

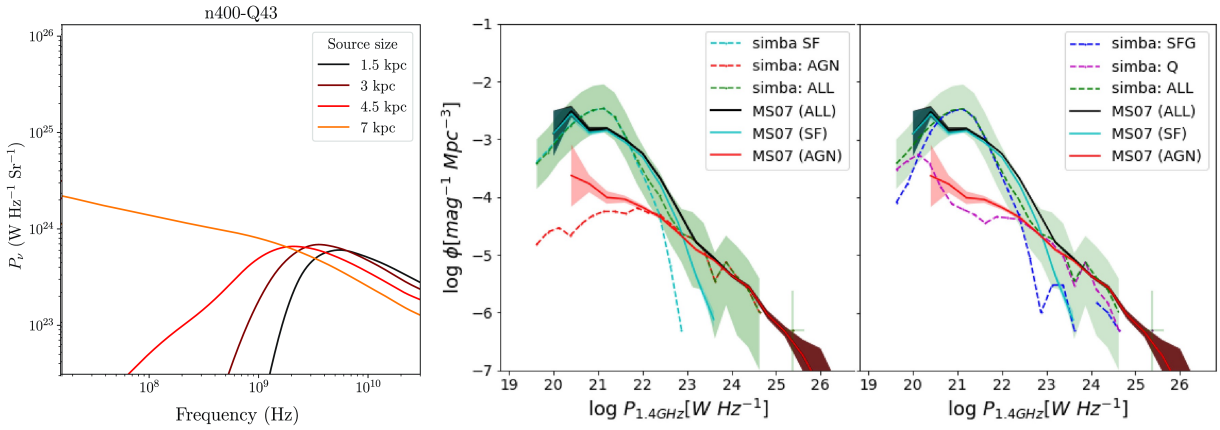


Figure 1.13: Left: The SEDs of a simulated radio AGN as it expands out to different spatial scales over time, from Young et al. (2025) simulation n400-Q43. This should be directly compared to observational SEDs for PS sources, like those in Figure 1.10. Right: the RLF derived from SIMBA AGN compared to that observed in the local Universe. The observational data indicated by 'M07' is from Mauch & Sadler (2007). Figure from Thomas et al. (2021).

SIMBA catalogues, and is shown in Figure 1.13, right. They were also able to use this framework to predict the presence of a population of faint, local HERGs which were later confirmed in deep radio observations (Thomas et al., 2025). Thus while there has been some promising recent work on radio AGN as seen in cosmological simulations, this field remains largely under-explored, and it is an area to which we will return in Chapter 4.

1.6 This thesis

So far, we have taken a tour through the history of active galactic nuclei (Section 1.1), what they look like (Section 1.1.3), the environments they inhabit (Section 1.1.5), and their importance to radio astronomers (Section 1.1.6). We have considered current evolutionary models for radio AGN through the lens of peaked spectrum sources (Section 1.2), as well as the importance of PS sources to studies of jet-gas interactions (Section 1.2.3). In the second half of this chapter, we have seen how survey science has evolved in step with advances in instrumentation and computation (Section 1.3), and what this means for the future of archival science (Section 1.4). Finally, we considered, in broad strokes, the ways in which simulations can and should be useful to observational astronomers (Section 1.5). All of these ingredients are mixed together to inform the chapters which follow.

In Chapter 2, I establish a new, Bayesian framework for identifying PS sources from pre-existing radio surveys, and test it on a sample of sources with known broadband SED shapes. This is heavily informed by the ideas raised in Section 1.4 surrounding the use of archival data, and my intention is to outline a method which will remain useful well into SKA operations and beyond.

When Chapter 2 was published, the full ASKAP-FLASH survey had only just begun and our sample of H I detected sources remained modest. This prompted the work in Chapter 3, where I asked not "how many H I detected sources are PS?" but "how many PS sources could ASKAP-FLASH search for H I?", applying my Bayesian framework from Chapter 2 to my own, untargeted search of PS sources in the equatorial, Stripe 82 field.

Again, while the sample of H I-detections from ASKAP-FLASH was growing, I turned to simulations in Chapter 4, where I used a new tool, SANGRiA, which I co-developed for the project to produce mock H I-absorption observations from the SIMBA cosmological simulations. Here, I asked the question "how much of the high H I-detection rate towards PS sources can be explained by geometry alone?". Based on detection rates from the ASKAP-FLASH Pilot Survey, these simulations tell us that geometry alone is not sufficient to account for the detection of H I towards PS sources, it must be due to their gas-rich environment also.

Of course, the process of doing science is not always linear, and often discoveries come at unexpected

times. In Chapter 5, we outline one such discovery, made during the course of validating ASKAP-FLASH data and developed in collaboration with colleagues across a number of areas of expertise. This chapter presents a multiwavelength study of a quasar that is neither a peaked spectrum, nor a host of H I gas, but a strong backlight for intervening gas. We also discuss the physics of the intervening systems detected in absorption.

In Appendix A, I outline a short piece of analysis on the ASKAP-FLASH Pilot Survey detections and the significance of the broadband radio SED in these sources. Appendix B is a short work on the history of scintillation studies, which arose from a series of discussions with members of the ASKAP-FLASH team who work closely with interplanetary scintillation datasets.

I end this thesis with a summary of the work presented, and some thoughts on the future direction of peaked spectrum studies, large area surveys and our use of our ever growing archives of data.

2

Bayesian inference of radio SEDs from inhomogeneous survey data

Emily F. Kerrison[1,2,3], James R. Allison[4], Vanessa A. Moss[1,3], Elaine M. Sadler[1,2,3], and Glen A. Rees

[1]Sydney Institute for Astronomy, School of Physics A28, University of Sydney, NSW, 2006, Australia
[2]ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia
[3]ATNF, CSIRO Space and Astronomy, PO Box 76, Epping, NSW 1719, Australia
[4]First Light Fusion Ltd, Unit 9/10, Oxford Pioneer Park, Mead Road, Yarnton, Kidlington OX5 1QU, UK

This chapter is published as *RadioSED - I. Bayesian inference of radio SEDs from inhomogeneous surveys.* Kerrison, E. F., Allison, J. R., Moss, V. A., Sadler, E. M., Rees, G. A. 2024. *Monthly Notices of the Royal Astronomical Society* 533, 4248–4267. doi:10.1093/mnras/stae1796.

We are thus facing an embarrassment of richness: a situation where we cannot effectively use the tremendous – and ever growing – amounts of valuable data already in hand.

Djorgovski & Williams (2005), on the motivation for the VO

This chapter in context

The release of the Callingham et al. (2017) sample of 1,483 PS sources, more than an order of magnitude larger than any previous uniform sample of these AGN, brought a wave of fresh enthusiasm to the field. Yet the radio archives already contain terabytes, if not petabytes of data on millions more continuum sources, some ~ 10 per cent of which are likely to be peaked spectrum (Callingham et al., 2017; O'Dea, 1998).

Many individuals and groups have their own code for compiling flux densities and modelling the radio SEDs of their favourite sources, but while these methods remain unpublished they are difficult (or at least, time consuming) to reproduce. This leads to a duplication of effort across teams, and makes the barrier to entry higher than it need be for those who come with expertise from other wavelengths and are thus unfamiliar with some of the nuances of radio catalogues. With the wealth of data already sitting in archives, this seemed an easy problem to address.

This chapter outlines a new framework, built on Bayesian principles and written in Python, for compiling flux densities and performing radio SED modelling for any (unresolved) radio continuum source at Dec $\leq +30^{\circ}$ (the northern declination limit of the RACS-low survey; Hale et al. 2021). Our intention is for this framework to be flexible enough that it can be continually and easily updated with more complex SED models, or more data as new, large-area surveys are made public (as we will see in Chapter 3). It is also designed so that non-expert users can obtain a modelled radio SED for any source, either by making use of the pre-matched flux densities from large surveys as it was used in Yoon et al. (2025) and Weng et al. (2025), or by bringing their own measurements for their favourite sources, as it was used in Behiri et al. (2025). The code is available on Github.[1]

Statement of contribution

This chapter faithfully reproduces Kerrison et al. (2024c), which was accepted for publication in the *Monthly Notices of the Royal Astronomical Society*, Vol. 533, Issue 4, pp. 4248 – 4267 (October 2024), with minor stylistic adaptations to fit the thesis style. I am the first author, with James Allison, Vanessa Moss, Elaine Sadler and Glen Rees as co-authors. The research problem was conceived of by me in the first few months of my thesis, building on earlier work completed by J. Allison using a frequentist approach to modelling. J. Allison provided guidance as to the interpretation of Bayes factors and the implementation of model selection, and G. Rees aided in some technical aspects of the Python implementation and code profiling, with advice and feedback provided by E. Sadler and V. Moss. I developed and packaged the code, wrote the paper and interpreted the SEDs in Section 4.3. Co-authors reviewed the manuscript and assisted with minor improvements to the text.

[1] `https://github.com/ekerrison/RadioSED`

2.1 RadioSED: Abstract

We present here RadioSED, a Bayesian inference framework tailored to modelling and classifying broadband radio spectral energy distributions (SEDs) using only data from publicly-released, large-area surveys. We outline the functionality of RadioSED, with its focus on broadband radio emissions which can trace kiloparsec-scale absorption within both the radio jets and the circumgalactic medium of Active Galactic Nuclei (AGN). In particular, we discuss the capability of RadioSED to advance our understanding of AGN physics and composition within the youngest and most compact sources, for which high resolution imaging is often unavailable. These young radio AGN typically manifest as peaked spectrum (PS) sources which, before RadioSED, were difficult to identify owing to the large, broadband frequency coverage typically required, and yet they provide an invaluable environment for understanding AGN evolution and feedback. We discuss the implementation details of RadioSED, and we validate our approach against both synthetic and observational data. Since the surveys used are drawn from multiple epochs of observation, we also consider the output from RadioSED in the context of AGN variability. Finally, we show that RadioSED recovers the expected SED shapes for a selection of well-characterised radio sources from the literature, and we discuss avenues for further study of these and other sources using radio SED fitting as a starting point. The scalability and modularity of this framework make it an exciting tool for multiwavelength astronomers as next-generation telescopes begin several all-sky surveys. Accordingly, we make the code for RadioSED, which is written in python, available on Github.

2.2 Introduction

Over the last few decades, advances in both observational and theoretical astronomy have brought great progress in our understanding of galaxy structure and evolution, as well as in our understanding of complex AGN-galaxy interactions. On the theoretical side, ever-increasing computational efficiency has allowed for more physically realistic simulations from the cosmological scale down to the hydro-dynamics of individual AGN jets (e.g. Somerville & Davé 2015). On the observational side and amongst many other advances, large, uniform surveys spanning both optical (e.g. GAMA: Driver et al. 2009, SDSS: York et al. 2000, DESI: Dey et al. 2019) and infrared wavelengths (e.g. WISE: Wright et al. 2010) have provided millions of low-resolution, photometric measurements of galaxies across cosmic time. One place in which the two fields meet is spectral energy distribution (SED) fitting, where these broadband, photometric measurements are mapped onto theoretical models of galaxy structure and evolution to tell us about everything from star formation history, to the state and quantity of the dust and gas of a galaxy (Conroy, 2013). However in the radio regime, where AGN emissions dominate the sky above a few milliJansky (Windhorst et al., 1999; Richards et al., 1999), SED fitting and analysis is still somewhat in its infancy.

There are many SED fitting tools on offer for the interested astronomer, and Pacifici et al. (2023) provide an excellent summary of these to date. Yet while the library of publicly-available SED fitting codes has now grown to more than a dozen, fewer than half are capable of handling AGN emissions (as opposed to those from the host galaxy), and fewer still extend their SED fitting down into the radio regime. Amongst the most popular fitting codes, both CIGALE (Burgarella et al., 2005; Boquien et al., 2019) and MAGPHYS (da Cunha et al., 2008) implement radio SED fitting, but only CIGALE implements a radio AGN component, which appears there in the form of a power law relationship between radio luminosity L_{AGN} and radio frequency ν, and serves mainly to correct for an apparent excess in the radio flux predicted due to star formation (see Yang et al. 2022, Section 5 for further discussion).

Part of this lack of radio SED fitting tools is undoubtedly due to the relative sparsity of radio flux density measurements compared to those in the optical and infrared. Indeed until very recently, large-area radio surveys were limited to only a handful of frequencies; the NRAO VLA Sky Survey (NVSS; Condon et al. 1998) at 1.4 GHz in the north, and the Sydney University Molonglo Sky Survey (SUMSS; Mauch & Sadler 2007) at 843 MHz in the south are perhaps two of the most well-known.

In recent years though this has begun to change, starting with the low frequency GaLactic and Extragalactic All-sky Murchison Widefield Array (GLEAM) survey (Hurley-Walker et al., 2017) which surveyed the full southern equatorial sky with a 160 MHz bandwidth in a previously under-explored frequency space below 300 MHz. This led to the first systematic studies of radio spectral shape at those frequencies by Callingham

et al. (2017) and Ross et al. (2021), where the former produced, amongst other things, the largest sample of "Peaked Spectrum" (PS) AGN to date, a particularly interesting class thought to represent the very youngest radio AGN (O'Dea & Saikia, 2021), and the latter studied their spectral variability. Indeed these works reinforced earlier findings (e.g. O'Dea 1998; Snellen et al. 2000; Edwards & Tingay 2004 to name but a few), that a small but scientifically significant fraction of radio AGN exhibit diverse radio spectral shapes that cannot be captured by a simple power law model alone.

Fortunately, we are on the cusp of many releases from extremely sensitive all-sky surveys at higher frequencies from new and upgraded instruments alike that will make such analyses possible at higher frequencies, including the Very Large Array (VLA, with the VLA Sky Survey VLASS; Gordon et al. 2021), the Giant Metrewave Radio Telescope (GMRT, producing the TIFR GMRT Sky Survey TGSS; Intema et al. 2017), the Australian SKA Pathfinder (both the Rapid ASKAP Continuum Survey RACS; McConnell et al. 2020; Hale et al. 2021; Duchesne et al. 2023 and Evolutionary Map of the Universe EMU; Norris et al. 2011) and LOFAR (with the LOFAR Two-metre Sky Survey LoTSS; Shimwell et al. 2017). Yet despite the large bandwidth of many of these instruments, no single survey alone will be enough to characterise broadband radio spectra spanning 700 MHz through to tens or even hundreds of gigahertz. This is especially so since the standard practice for radio catalogues is to compress the observed bandwidth into a single flux density measurement at some central frequency. Therefore any approach to radio SED fitting on population-scales will still require a careful combination of several, independent radio surveys.

It is timely, therefore, that we present here RadioSED[2], a framework for extending the low-frequency, all-sky AGN spectral modelling of Callingham et al. (2017) by compiling SEDs from all suitable, large-area radio surveys to date. This framework grew out of a desire to identify and characterise a large, uniform sample of PS AGN with spectral peaks spanning a few megahertz through to tens or even hundreds of gigahertz, a task which was identified as a key area for future research in the review of O'Dea & Saikia (2021). However by construction, the model classification performed by RadioSED allows for the characterisation of both the young, PS AGN which were the main focus of Callingham et al. (2017), as well other radio AGN with flat, steep, and inverted SEDs, representing beamed blazar-like emission, larger (likely evolved PS) radio sources, and extremely young PS sources respectively, so that this framework has the potential to better inform us about the broadband spectral properties of the radio AGN population as a whole. For this reason our SED construction is complementary to that provided by existing tools like SPECFIND which perform more complex, algorithmic cross-matching between surveys but focus on sources with a power law spectral shape (Vollmer et al., 2005; Stein et al., 2021). Indeed at present, the models implemented in RadioSED are designed to differentiate between spectral shapes whilst being broadly applicable to different radio AGN environments, with a particular focus on identifying the young, PS sources. However, the modularity of our framework means that additional models incorporating more physical subtleties, such as differentiating between synchrotron self-absorption and free-free absorption of the radio jets, could be easily incorporated in the future. Not only that, but the automated and modular nature of this framework means it can be easily extended to include new surveys as they are released to the public, covering a wider sky area, a different frequency range or both at a greater sensitivity than ever before. The output from RadioSED, when considered alongside the output from semi-analytical models of radio jets like those of Turner & Shabala (2015), further developed in Turner et al. (2023), will also allow for a deeper analysis of broadband radio spectral properties akin to what is already possible with optical and infrared SED analysis.

In Section 2.3 we outline the process of compiling individual SEDs from large area surveys observed at different epochs. Section 2.4 presents our Bayesian approach to modelling and classifying the SEDs of individual radio sources, and in Section 2.5 we discuss the accuracy and reliability of this modelling with reference to both synthetic data, and observational data for a selection of sources with well-known SED shapes. A summary of this work, and plans for future science applications of RadioSED are discussed in Section 2.6, while a second, forthcoming paper in this series will present a deeper scientific analysis of the sources in our chosen pilot field, Stripe 82 (Abazajian et al., 2009).

[2]The code is made available to the user on Github:`https://github.com/ekerrison/RadioSED/`

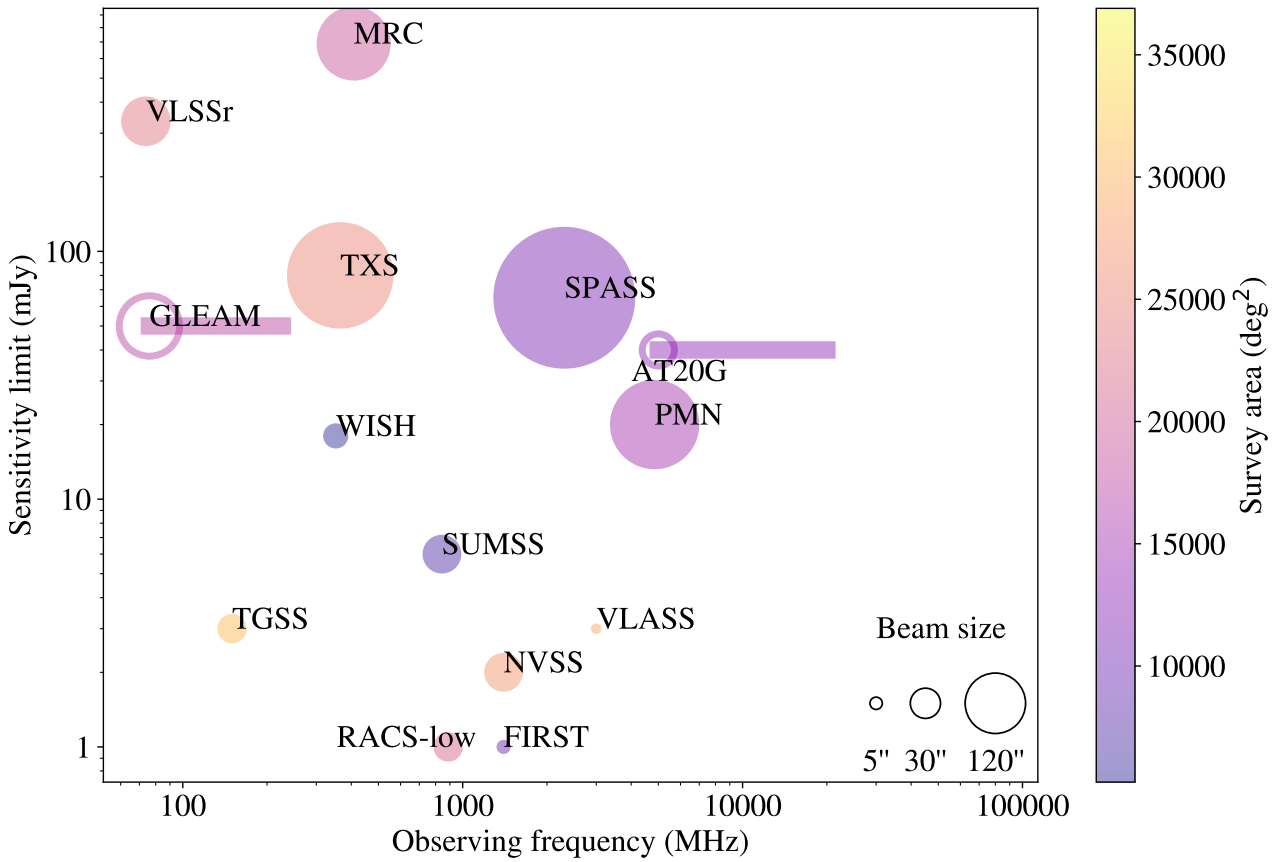


Figure 2.1: Frequency coverage of radio surveys to date which have observed more than $\sim 10^2$ square degrees of southern equatorial sky. Rectangular regions indicate the maximum frequency range of broadband surveys, which report flux density measurements at multiple frequencies in their respective catalogues. Surveys with measurements at a single frequency are indicated as circles, with the size of the circle indicative of beam size and the colour indicative of survey area.

2.3 SED construction

In order to identify interesting classes of radio AGN by their spectral shape, we must first construct their broadband SEDs in a reliable and reproducible manner. Ideally, these SEDs should cover a sufficiently large range of frequencies so as to identify spectral peaks which may manifest anywhere from a few hundred megahertz through to the gigahertz regime. Doing so will allow us in particular to capture a representative sample of the peaked spectrum population spanning a range of both ages and redshifts (recalling that the frequency of the spectral peak is inversely proportional to the linear size - and hence the age - of a source, but that this peak may be lowered in the observer's frame due to cosmological redshift; Snellen et al. 2000; Vries et al. 2009; Jeyakumar 2016). To date, the southern equatorial sky has been relatively well sampled in frequency space, as outlined in Figure 2.1, though the sensitivity and epoch of this coverage varies wildly. Since no single survey or instrument is yet capable of achieving comparatively broad coverage, we must combine data from multiple surveys, instruments and even epochs to achieve our stated goals.

Our approach to SED construction is outlined in the following sub-sections. In short, we rely on public datasets available via the VizieR catalogue access tool (Ochsenbein et al., 2000), combined with position-based cross-matching, to build up reliable SEDs containing flux density measurements of physically-associated sources observed at a range of frequencies. The minimal assumptions in our cross-matching algorithm are what sets this approach apart from the SPECFIND tool (Vollmer et al., 2005; Stein et al., 2021). Since SPECFIND assumes a spectral shape at the stage of cross-matching catalogues, the tool largely disregards flux density measurements which deviate from this expected trend as 'waste', though some work on identifying PS sources peaking at specific frequencies has been done recently (see Stein et al. (2021) Section 5). However our purely position-based matching algorithm keeps all flux density measurements within a certain distance threshold, allowing

us to better capture the more complex spectral curvature of interesting AGN sub-classes, such as the peaked spectrum sources. The caveat to using such minimally-informed crossmatching is that there is a higher chance of either matching non-physically-associated flux densities at different frequencies, or of making only a partial match between complex sources resolved in some surveys and not others. These downsides are mitigated by using the survey with the equal-highest resolution (aside from ATPMN and Sajina11 which are targeted surveys of point-like sources) as our reference survey, and by selecting a statistically-robust match radius determined individually for each survey pair, as outlined in Sections 2.3.1 and 2.3.2 respectively.

2.3.1 Catalogues

The first step in a uniform approach to broadband radio spectral modelling is to determine how to compile relevant flux density measurements and then, once compiled, to assign them to individual sources.

On the one hand, at the level of individual sources, we require flux density measurements across several decades in frequency space in order to construct reliable SEDs with which to identify spectral shape (including any peaks). On the other hand, at the sample level these flux density measurements should ideally be drawn from all-sky, or at least large area ($\geq 10^3$ deg^2) untargeted surveys in order to maximise both the reliability and completeness of the full sample. These two concerns, the reliability of individual SEDs, and the completeness of the full sample, must be balanced against each other in determining how to draw flux density measurements from the literature.

After a manual search of both the tables collated in Stein et al. (2021) and those elsewhere in the literature, we created a list of 25 candidate surveys outlined in Table 2.1. Of these, 17 were selected for final inclusion in our SED fitting, which can be found above the dividing line in said table.

Broadly speaking, we favoured radio surveys which fit the following criteria:

1. Large area, untargeted surveys ($\geq$ 5,000 deg^2) to minimise selection bias.
2. Recent observation (since 2000) to minimise the effects of blazar variability in multi-epoch SEDs and flux scale inconsistencies, especially since the Australia Telescope Compact Array (ATCA) flux scale was revised in 1994 to more closely align it with that of the VLA, and many Southern equatorial surveys have been conducted with the ATCA (Reynolds, 1994).
3. Observing frequency between 30 MHz-100 GHz to capture a broad range of spectral peaks, and hence a peaked spectrum sample covering a range of ages and redshifts.

Where large frequency bands lack surveys fulfilling all 3 criteria, we have occasionally chosen to include data from surveys fulfilling only one of either the area or epoch criteria, in order to maximise spectral coverage across our sample (e.g. the inclusion of the MRC and TXS surveys around 400 MHz). However, the modular nature of our fitting framework means that these surveys can be easily exchanged for more suitable ones as new data is made public.

From these surveys, a reference dataset must be selected upon which to base our SED construction. A closer consideration of all three criteria above led to the selection of RACS-low (Hale et al. 2021; emphasised in the table) for this purpose. Not only is it recent (observations date from 2020) and of an excellent sensitivity for our work (1 mJy at $\sim$ 800 MHz), but it also covers the entire southern, equatorial sky at a positional accuracy of 1-2 arcseconds (McConnell et al., 2020; Hale et al., 2021). Therefore, for this work all other included surveys were matched against the RACS-low catalogue in order to build the SEDs for individual objects, and the positions of sources are given by their coordinates in the RACS-low catalogue.

In some cases, targeted follow-up campaigns of large area surveys at different frequencies have been included (e.g. Ricci et al. 2006), because they provide additional sky coverage at certain frequencies not well covered by the untargeted surveys already selected. These are indicated by a "N/A" value in the survey area column of Table 2.1. We have also extended the frequency space explored up to the milimetre regime with the inclusion of the ALMA Calibrator Catalogue (ACC; Bonato et al. 2019), to allow us to capture the very youngest peaked spectrum sources which turn over at tens of gigahertz or higher.

Name	Frequency (GHz)	Beam FWHM (arcsec)	$n_{sources}$	Area (deg^2)	Instrument	Flux limit (mJy)	Flux limit Frequency (GHz)	Final Obs. Date	r_{match} 95% (")	Reference
VLSSr	0.0738	75	92964	30939	VLA	335	0.074	2007	25	Lane et al. (2014)
TGSS	0.15	25	623604	36900	GMRT	3	0.15	2012	18	Intema et al. (2017)
CCA	0.080-0.16	222	2213	N/A	CCA	1200	0.16	1984	43	Slee et al. (1995)
GLEAM	0.076-0.227	120	307455	24831	MWA	50	0.2	2014	25	Hurley-Walker et al. (2017)
TXS	0.365	360	66841	31538	Texas Int.	80	0.365	1983	16	Douglas et al. (1996)
MRC	0.408	172	12141	25770	Molongolo	690	0.408	1974	22	Large et al. (1981)
SUMSS	0.843	45	211050	8100	MOST	6	0.843	2003	14	Mauch et al. (2013)
RACS-low	**0.8875**	**25**	**2462639**	**28020**	**ASKAP**	**1**	**0.8875**	**2020**	**N/A**	**Hale et al. (2021)**
NVSS	1.4	45	1773484	33827	VLA	2	1.4	1997	11	Condon et al. (1998)
SPASS	2.307	645	23389	16000	Parkes	65	2.307	2010	48	Meyers et al. (2017)
PMN	4.850	252	50841	23265	Parkes	20	4.850	1990	43	Griffith et al. (1993)
Randall11	1.4-8.6	40	42	N/A	ATCA	1500	2.7	2008	129	Randall et al. (2012)
ATPMN	4.8-8.6	2.8	9040	7757	ATCA	7	4.85	1994	18	McConnell et al. (2012)
Sajina11	4.86-90	16	159	N/A	VLA	40	20	2010	19	Sajina et al. (2011)
AT20G	5-20	34	5867	20086	ATCA	40	20	2008	19	Murphy et al. (2010)
Ricci06	18.5-22	120	250	N/A	ATCA	1000	5	2002	103	Ricci et al. (2006)
ACC	85.8 - 373	115	3364	N/A	ALMA	1	95	2018	22	Bonato et al. (2019)
VLSS	0.074	80	92965	30939	VLA	0.2	0.074	2007	–	Cohen et al. (2007)
PKSCAT90	0.080-22	60	8264	5007	Parkes	50	1.4	1990	–	Wright & Otrupcek (1990)
WISH	0.352	18	90357	5252	WRST	18	0.352	1998	–	De Breuck et al. (2002)
FIRST	1.4	5	946432	10575	VLA	1	1.4	2011	–	Helfand et al. (2015)
VLASS	3	2.5	3381277	33880	VLA	3	3	2019	–	Gordon et al. (2021)
C-BASS	5	2700	–	41,253	C-BASS	–	–	–	–	King et al. (2010)
CRATES	8.4	–	11131	N/A	various	65	4.8	2006	–	Healey et al. (2007)
Kühr	0.012-89	–	518	N/A	various	–	–	1980	–	Kuehr et al. (1981)

Table 2.1: A summary of all of the wide-field surveys considered as part of our Bayesian framework. Those in the top half of the table have been included in the final flux density tables for sources of interest, those below the line were excluded from fitting for reasons further elaborated in-text. RACS-low (Hale et al., 2021) was used as the base catalogue to which all other catalogues were matched, and is emphasised in the table. Empty entries either indicate that the value could not be found in the literature, or varied substantially across the survey. Surveys with an area marked as "N/A" were targeted followup campaigns.

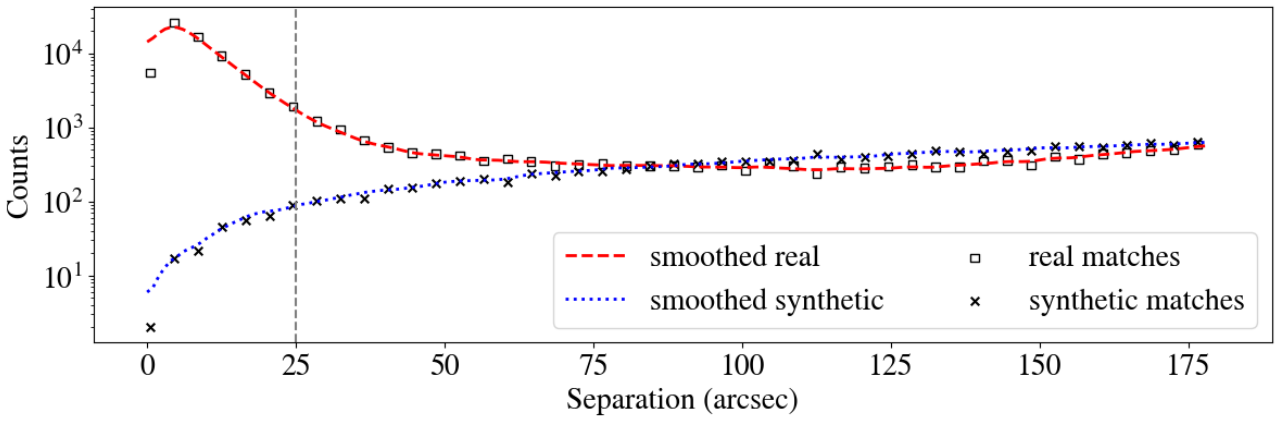


Figure 2.2: A representative histogram showing the separation between GLEAM sources (Hurley-Walker et al., 2017) and their counterparts in the real RACS-low reference catalogue (unfilled squares), and also with their counterparts in the synthetic reference catalogue (filled crosses). The red and blue dashed curves are the counts smoothed. The vertical line indicates the 25 arcsecond separation threshold at which the synthetic cross-matches reach 5% of the real cross-match count.

The surveys which were ultimately discarded were ruled out on the basis of flux scale errors (VLASS, VLSS; as documented in Gordon et al. (2021) and Cohen et al. (2007) respectively), a highly irregular beam (WISH) or poor *uv*-coverage leading to missing flux on scales comparable to that of our reference catalogue (FIRST), or incomplete documentation of the observing campaigns from which flux densities were derived (PKSCAT90). We additionally ruled out the use of the CRATES data owing to a lack of information about resolution across the survey area, and discrepancies with the AT20G flux scale (see Healey et al. 2007 section 3.3), and we likewise rule out the Kühr sample owing to the age of the catalogue and the assortment of observing parameters. Finally, we discard the C-BASS survey as its southern equatorial portion is not yet released, and its focus is on diffuse galactic emission.

2.3.2 Catalogue Cross-matching

Simple, position-based crossmatching against the reference catalogue (RACS-low) was performed for each catalogue in the top region of Table 2.1 using the CDS XMatch service (Pineau et al., 2020) accessed via Astroquery (Ginsburg et al., 2019). Matches were accepted as physically associated, and not a chance coincidence, if the separation between catalogued coordinates fell within the radius given in the r_{match} column in arcseconds. This value varies across surveys since each catalogue used has a different source density dependent on its frequency, sensitivity and beam size. In each case, the radius was chosen using a simple statistical method based on the approach of Best et al. (2005); Sadler et al. (2007); Ching et al. (2017) and others, who match radio sources with both real and randomly positioned optical counterparts to ensure the completeness and reliability of their real samples.

First, a 'synthetic' catalogue was created as a counterpart to the reference catalogue by shifting the coordinates of every source by 1 degree at a random position angle. The magnitude of the shift was chosen to be large enough to remove the physical clustering produced by extended, multi-component sources, which Blake et al. (2003) demonstrated is significant out to only 0.3 degrees in both NVSS and FIRST. This resulting synthetic catalogue preserves the source density of the original, and can therefore be meaningfully compared with it. We then performed cross-matching against both the synthetic and real RACS-low catalogues for every other survey of interest, and examined the number of matches as a function of radius, as shown in Figure 2.2. We determine the maximum acceptable match radius to be the point at which matches with the synthetic catalogue comprise no more than 5% of the number of matches with the real reference catalogue when the counts of both are smoothed across bins using a Gaussian filter. All crossmatches below this threshold are statistically very likely to represent flux density measurements of real, physical sources observed across the various surveys, and are hence included as reliable measurements in the SEDs of individual sources.

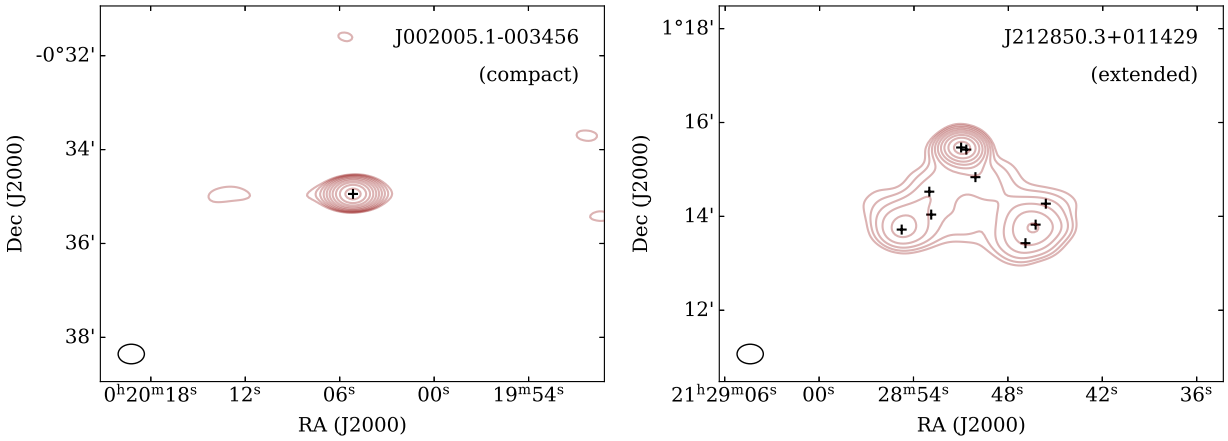


Figure 2.3: Examples of sources which appear compact (left) and extended (right) in the RACS-low observations. Red contours are logarithmically spaced between $5 - 200$ mJy. Black crosses indicate the positions of the Gaussian components belonging to the island source displayed by the radio contours. The RACS-low beam size is given in the bottom left of each plot.

2.3.3 Source Compactness and Complexity

Since all of our chosen surveys were observed at different resolutions, and many of them with interferometers, the morphology of a source has the potential to affect its SED shape. In general, radio sources can be broadly grouped into two morphological categories: compact, and extended. This classification depends on the resolution of the observations (since at some level, all radio sources must be extended), but it is nevertheless a useful distinction in the context of SED fitting, particularly for the identification of peaked spectrum sources, which are known to be extremely compact, with a linear diameter of $\lesssim 1$ kpc at all observable radio frequencies (O'Dea & Saikia, 2021). Sources of each kind as seen in RACS-low are displayed in Figure 2.3. In terms of the relevance of source morphology for RadioSED, sources with extended structure may suffer from under-reported flux densities due to emission on scales larger than $\lambda/2B_{min}$ being resolved out (by the Nyquist-Shannon sampling theorem), where B_{min} is the shortest baseline length. While we do not attempt to correct for this in our SED fitting, we do check for the compactness of each source across 3 orders of magnitude in frequency space, so that extended sources which may suffer from missing flux can be flagged for later interpretation if desired. We note also that across the RACS-low catalogue, $\sim 27\%$ of sources brighter than 10 mJy are extended on scales which lead to multi-component Gaussian fits. Therefore, while the majority of sources fit by RadioSED will likely be compact, some significant fraction will be extended at gigahertz frequencies, requiring a more careful interpretation of their multifrequency SED.

As stated, we obtained a measure of source compactness at three representative frequencies from across our selected surveys by utilising three surveys with the widest coverage: GLEAM ($\sim 10^{-1}$ GHz), RACS-low ($\sim 10^{0}$ GHz) and AT20G ($\sim 10^{1}$ GHz; Murphy et al. 2010).

At the lowest frequencies, compactness on arcminute scales was obtained by taking the ratio of peak to integrated flux across the wide GLEAM band (170-231 MHz), using the `Fpwide` and `Fintwide` catalogued values (Hurley-Walker et al., 2017).

This process was repeated at just below 1 GHz for RACS-low using the `Fpk` and `Ftot` catalogue entries which give an indication of compactness on arcsecond scales (Hale et al., 2021). For this survey, the number of Gaussian components into which each source was divided (reported as N_g in the catalogue) was also recorded as an extra measure of source extent, since sources with complex, resolved morphology will be fit by multiple Gaussian components within the one island.

At high frequencies, flux-based measures of compactness in the image plane could not be used, since the AT20G survey only recorded peak fluxes and made use of only the short baselines in the ATCA's hybrid arrays (Murphy et al., 2010). Instead, we made use of the additional AT20G High Angular Resolution Catalogue (Chhetri et al., 2013) which measured compactness on sub-arcsecond scales, reported as a visibility ratio, by utilising the raw 6-km visibilities from the original AT20G observing runs.

In all three cases, compactness is represented numerically on a linear scale from 0 - 1, with 1 being most compact. In the case of RACS-Low, we obtained our additional measure of extension from the component breakdown, where this time a value of 1 indicates the most compact sources (fit by 1 Gaussian component in the image plane), and anything higher indicates some extension (fit by multiple components). Since these measures of compactness cover three orders of magnitude in frequency space, they are a useful complement to the classifications returned from SED fitting alone. For example, if a source with an SED that turns over is compact in all three surveys, this is a good indication we have identified a true compact, peaked spectrum source. On the other hand, if a source has a compactness measure that is inversely proportional to frequency, that may be an indication of a compact core embedded in larger, extended lobes, which tend to manifest most clearly at low frequencies due to their steep spectral index (Miley & De Breuck, 2008; Blandford et al., 2019). Thus these compactness measures provide a useful, if simplistic, indication of source morphology as it pertains especially to identifying peaked spectrum sources, and they do so without the need for high resolution imaging.

In future work, additional measures of source compactness may be easily incorporated into this framework. In particular, recent and ongoing observations of interplanetary scintillation made with both the MWA (Morgan et al., 2022) and ASKAP (Chhetri et al., 2023) will provide an invaluable measure of source compactness down to sub-arcsecond scales at frequencies at and below ~ 1 GHz.

2.3.4 Source Variability

The variability of radio AGN is a complex topic and no doubt our understanding is not yet complete, so we limit our discussion here to a consideration of the known variability mechanisms that might impact our SED fitting.

AGN variability may be intrinsic, caused by the clumpiness of the circumnuclear material (e.g. Risaliti et al., 2002, on the clumpy torus), a changing rate of accretion onto the supermassive black hole (Czerny et al., 2009), or by variations and instabilities within the radio jets themselves which may produce propagating shocks (e.g. Marscher & Gear, 1985; Kovalev & Kovalev, 2006; Hovatta et al., 2008), and all of these may be beamed, increasing the degree of observed variability and producing a blazar (Blandford et al., 2019).

Variability may also be extrinsic, caused by changing diffraction and refraction of light from the source as it passes through the turbulent interstellar medium of the Galaxy (Interstellar Scintillation - "ISS"; Rickett 1968; Rickett et al. 1984) or the solar wind (Interplanetary Scintillation - "IPS"; Clarke 1964 see also Hewish et al. 1964). All of these mechanisms may operate at once, producing complex variability across the radio spectrum on timescales from milliseconds to years, and in magnitude from a few percent, to several times the minimum recorded flux density (e.g. Torniainen et al., 2005; Ross et al., 2021). For this work, our main interest is in how this variability could affect the SED shape of our sources when combining multi-epoch data, as well as in how transient spectral peaks could contaminate any eventual peaked spectrum source sample. We therefore care about the cadence at which variability occurs, the rate of variability within the radio AGN population, and the degree to which it occurs, as our sample is most likely to suffer from contamination or incompleteness when all three are extreme.

First of all, we must consider the temporal cadence at which we expect radio AGN to vary. Since IPS typically occurs on timescales of a few seconds in arcsecond-scale sources (Hewish et al., 1964), its effects will be averaged out by the integration time of the surveys used here; the ATCA has a standard integration cycle of 10 seconds[3], and of the VLA surveys, VLSSr uses the shortest integration time at 10 seconds (Cohen et al., 2007). Other surveys here use comparable or longer integrations. By contrast, ISS fluctuations affecting the most compact sources happen on timescales of hours to days and also exhibit an annual cycle (Rickett, 2001; Bignall et al., 2003). Likewise, intrinsic variability has been observed over periods ranging from months to years at gigahertz frequencies (Hovatta et al., 2007; Koay et al., 2018). Therefore, while IPS will not have a meaningful impact on the flux densities used for our SEDs, both ISS and intrinsic variability may impact our propsosed SED analysis.

Fortunately, the rate and degree of radio variability within the compact, radio AGN population both appear to be relatively low, with quasar-type AGN exhibiting more variability than their galaxy-type counterparts (Tinti et al., 2005). Sadler et al. (2007) studied the variability of compact sources at 20 GHz in an untargeted survey,

[3] ATCA User Guide

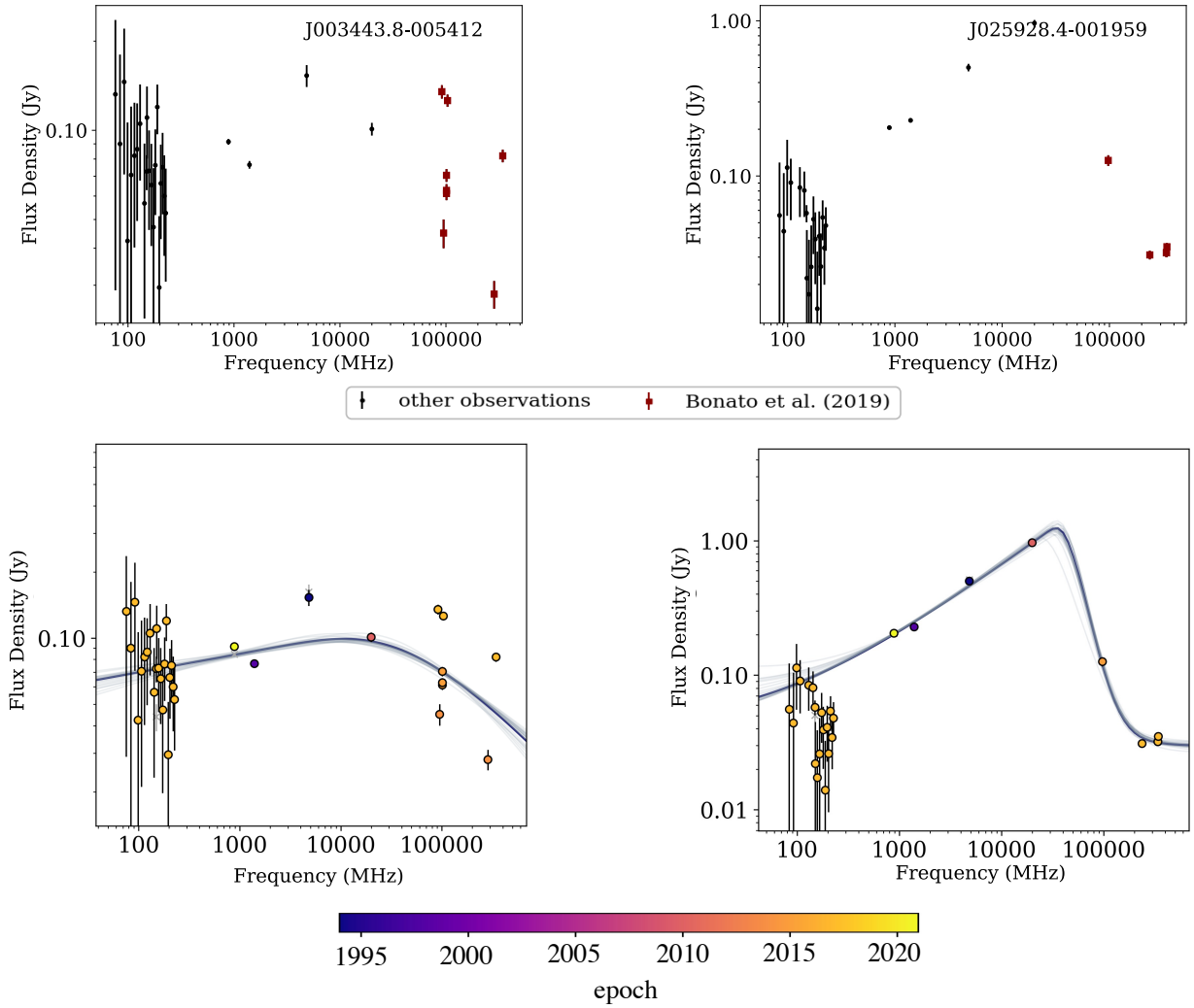


Figure 2.4: An example of a source exhibiting significant ($> 10\sigma$) variability in the ALMA band (left), and one with a relatively stable flux density (right). In the top panel flux density measurements from the ALMA Calibrator Catalogue (Bonato et al., 2019) are indicated by dark red squares, black points indicate the other flux density measurements from the surveys in Table 2.1. In the bottom panel, observations are coloured by observing epoch as drawn from the publication date of the respective survey in Table 2.1, with the model from RadioSED shown in grey. From a manual inspection of this representation, J003443.8-005412 (left) appears to exhibit more variability between sampled epochs than does J025928.4-001959 (right).

finding a median variability of 6.9% with only 5% of sources exhibiting variability at the 30% level or higher. Likewise at lower frequencies, Ross et al. (2021) examined variability in the GLEAM band below ~ 300 MHz and found it to be low, with only $\sim 1.5\%$ of their full sample classed as variable, and of this variable subset only 16% (51 out of 21,558 total sources) were found to exhibit extreme variability leading to changes in their spectral shape over the 3 year monitoring period. They did however note that peaked spectrum sources were over-represented in the variable subset, comprising about 30-40% of this population, as were sources with high-frequency components. This is at odds with the definition of a peaked spectrum source as having a quiescent spectrum with variability below the 10% level (O'Dea, 1998). However, in other followup campaigns of broadband instantaneous observations, a large degree of variability from apparent peaked spectrum sources has been shown to be the result of contamination of peaked spectrum samples by flaring blazars with a self-absorbed component, which produces a transient spectral peak in an otherwise flat radio spectrum (Orienti et al., 2007; Torniainen et al., 2007; Mingaliev et al., 2012).

To better address the question of blazar contamination in our fitting, as well as to provide additional frequency coverage, we have chosen to include the ALMA Calibrator Catalogue (Bonato et al., 2019) in our SEDs. This contains all calibrator observations made between 2011 and 2018 and at 90-370 GHz, and many (though not all) of these sources are known blazars. In sources exhibiting variability in the ALMACAL catalogue (ACC) at the 10σ level, we do not use the flux density measurements when performing SED fitting, but we do flag the source as variable. We also return the maximum ALMA Variability Index (VI) as defined in Akritas & Bershady (1996) and Barvainis et al. (2005) for all sources with more than 2 measurements in any given ALMA band. Individual sources exhibiting ALMA variability and a quiescent spectrum are shown in Figure 2.4, though we stress that this perceived variability (or lack thereof) is closely related to the number of available ALMA observations, and should not be used as the sole discriminator between variable and quiescent sources. In addition to returning variability statistics from ALMACAL data, after fitting, RadioSED produces plots like those shown in the bottom panels of Figure 2.4, where the best fitting SED model (as described below) is shown alongside data points coloured by observing epoch. The observing epoch is derived from the publication date of each survey in Table 2.1, except in the case of ACC data, where the observing epoch is explicitly provided as part of the catalogues. This allows the user to perform a simple, visual check as to whether the sources' apparent SED shape is at all influenced by the multi-epoch nature of the observations. However, for a more complete understanding of variability and in conjunction with the results from RadioSED, users may also find it useful to crossmatch against the Roma-BZCAT catalogue (Massaro et al., 2015) to determine whether a source is a known blazar, as well as against the SPECFIND catalogue (Stein et al., 2021) to view flux densities spanning a broader time and frequency coverage. Furthermore, since γ-ray emitting AGN are almost always classified as blazars (Bose et al., 2022), an additional cross-match with Fermi sources from, for example, the catalogue of D'Abrusco et al. (2019) may prove useful in identifying further blazar-like sources beyond those identified using radio SED variability alone. However, since these high-energy observations cannot be used in a radio SED, this functionality is not inbuilt in RadioSED, and our own study of the Fermi sources in our sample is left to future work.

Therefore, we do not expect variability to have a large effect on the SEDs constructed using the compiled surveys in Table 2.1 owing to the low rate and degree of variability amongst the radio AGN population. There may be some distortion of individual SEDs due to variability, but the largest impact is likely to be due to blazar contamination in any given peaked spectrum sample. However, since these flares appear transient over a period of months to years, our choice to compile flux densities from several surveys from different epochs in fact lowers the likelihood of such contamination, as does our use of the ALMA Calibrator Catalogue. In fact, Torniainen et al. (2007) note the advantage of compiling multi-epoch flux densities to isolate true peaked spectrum sources from larger samples. The number of sources we expect to be contaminating blazars or variable AGN is further discussed in Section 2.5.2.

2.4 Bayesian Modelling

With our SEDs constructed from the surveys listed in Table 2.1, it remains now to fit them with suitable models in order to identify the peaked spectrum sources from amongst the larger population. To do this we employ a Bayesian approach, which allows us to not only infer the model parameters and their posterior probability

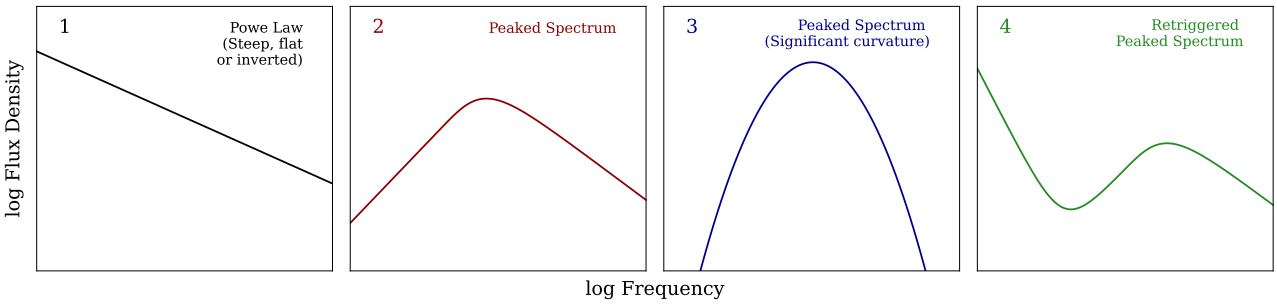


Figure 2.5: Indicative shapes of each of the 4 model types implemented by default as part of RadioSED.

densities for each model type, but also to compare different models and make an informed choice about which one best explains the data in question. Since our models can capture various degrees of curvature (as discussed below), we require a minimum of 4 flux density measurements at unique frequencies to perform our model inference. As we have been careful to construct our SEDs from large area surveys, many of which have flux limits on the order of a few milliJanskys, sources not meeting this minimum threshold are typically relatively faint in RACS-low ($S_{0.8\,\mathrm{GHz}} \leq 15$ mJy), though some may also be variables or transients.

2.4.1 Models

A set of four analytical models are applied to each SED using Bayesian inference. These are displayed in Figure 2.5, and each capture a physically distinct category of radio SED by modelling flux S_ν (or its logarithm) as a function of frequency ν as follows:

1) A steep, flat, or inverted spectrum source

Such sources are described by the simple functional form of a power law in Equation 2.1. These three SED shapes (steep, flat, and inverted) vary functionally only by the sign of the spectral index α, and hence are the simplest SEDs to fit. Physically, the flat spectrum sources are typically found to be quasars or non-flaring blazars, often with an elevated incidence of gamma-ray emission (e.g. Healey et al., 2007; Kimball & Ivezić, 2008; Mahony et al., 2010), or the compact cores of extend radio sources which can be distinguished clearly on VLBI scales (e.g. Pushkarev & Kovalev, 2012). The steep and inverted spectrum sources, by contrast, are thought to be peaked spectrum sources whose peak falls outside the observed frequency range, either due to being extremely compact (and hence exhibiting a high frequency peak), or extended out to ~ 10 kpc so that the peak frequency has shifted down into the megahertz regime (O'Dea & Saikia, 2021). Their SED shape can be characterised by:

$$S(\nu) = S_0 \nu^\alpha \tag{2.1}$$

where S_0 is the amplitude of the synchrotron spectrum in Janskys. By convention, the division between steep and flat-spectrum sources is often placed at a spectral index $\alpha = -0.5$ (e.g Urry & Padovani, 1995) which we adopt here for consistency. Less well-defined is the distinction between flat and inverted spectrum sources, but we here adopt $\alpha = 0.5$ so that the range of flat spectrum indices is symmetric about zero.

2) A peaked spectrum source

The simplest peaked spectrum source model is given in Equation 2.2. This was first used by Snellen et al. (1998) to characterise peaked spectrum sources, and it is now one of the more popular functional forms for describing these AGN, as it is simple to read off useful parameters like the peak frequency, peak flux and spectral indices (e.g. Callingham et al., 2017; Shao et al., 2020; Wołowska et al., 2021; Su et al., 2022).

$$S_\nu = \frac{S_\mathrm{p}}{1 - e^{-1}} \left[1 - e^{-(\nu/\nu_\mathrm{p})^{\alpha_\mathrm{thin} - \alpha_\mathrm{thick}}}\right] \left(\frac{\nu}{\nu_\mathrm{p}}\right)^{\alpha_\mathrm{thick}} \tag{2.2}$$

In the above equation, S_p is the flux density at peak frequency ν_p and α_thin, α_thick are the spectral indices in the optically thin (above ν_p) and thick (below ν_p) regions. This model reduces to the case of a homogeneous, synchrotron self-absorbed source when $\alpha_\mathrm{thick} = 2.5$, but otherwise represents some inhomogeneous absorber which may be affected by either free-free absorption (FFA) or synchrotron self-absorption (SSA), such as a clumpy, circumnuclear medium (Risaliti et al., 2002), or transient, radiation-driven outflows of ionised gas (Wada et al., 2023). Even when this function was first used to fit peaked spectrum sources, Snellen et al. (1998) noted that it sometimes failed to accurately capture SED structure far away from the peak, which brings us to our next two functional forms.

3) A peaked spectrum source with significant curvature

Some peaked spectrum sources are not well fit by Equation 2.2. Instead, they can be described by a parabolic functional form in logarithmic space, or a log-parabola, as in Equation 2.3.

$$\log S_\nu = a + \log \nu \, [b + c \log \nu] \tag{2.3}$$

This form was originally used by Orienti et al. (2007) and Orienti et al. (2010) to characterise peaked spectrum sources because it was a good fit to their data, even though the parameters a, b and c were not thought to hold any physical meaning. However, Duffy & Blundell (2011) outline how the curvature parameter (c here, q in that work) can be used alongside the peak frequency ν_p and peak emissivity (L_p, found from the peak flux density S_p, distance, and size of the source) to derive the magnetic field strength and properties of the electron distribution in synchrotron emitting plasma lobes. Ross et al. (2021) used a similar log-parabola form to fit sources peaking at low frquencies in the GLEAM band, and the same function identified here is also used by Callingham et al. (2017) and Nyland et al. (2020) at radio frequencies. In the radio to sub-mm regime, a similar function is used to fit broadband, synchrotron self-absorbed emissions of blazars (e.g. Abdo et al., 2010; Chen et al., 2023), where there is some evidence that the degree of curvature may correlated with peak frequency and is thus able to be related back to either statistical or stochastic particle acceleration under certain conditions (Tramacere et al., 2011; Chen, 2014).

4) A re-triggered peaked spectrum source

In some cases, peaked spectrum sources exhibit an upturn again at the lowest frequencies, producing an SED well fit by the functional form in Equation 2.4.

$$S_\nu = \frac{S_\mathrm{p}}{1 - e^{-1}} \left[1 - e^{(\nu/\nu_\mathrm{p})^{\alpha_\mathrm{thin} - \alpha_\mathrm{thick}}}\right] \left(\frac{\nu}{\nu_\mathrm{p}}\right)^{\alpha_\mathrm{thick}} + S_0 \nu^\alpha \tag{2.4}$$

This is a linear combination of Equations 2.1 and 2.2, with all of the same parameters from each of those. Physically, these sources are interpreted as AGN undergoing multiple cycles of activity on short timescales, or exhibiting distinct knots of emission along their jets, with the older, more diffuse emission producing the upturn at low frequencies (Baum et al., 1990; Edwards & Tingay, 2004; Hancock et al., 2010). They are often, though not always, observed at the centre of cluster environments (Hogan et al., 2015; Callingham et al., 2017).

2.4.2 Bayesian Inference

For each model, we test the hypothesis ($\mathcal{M}$) that it describes the true form of an SED given model parameters ($\boldsymbol{\theta}$), and SED data ($\boldsymbol{d}$), using Bayes' theorem:

$$\Pr(\boldsymbol{\theta}|\boldsymbol{d}, \mathcal{M}) = \frac{\Pr(\boldsymbol{d}|\boldsymbol{\theta}, \mathcal{M}) \Pr(\boldsymbol{\theta}|\mathcal{M})}{\Pr(\boldsymbol{d}|\mathcal{M})}. \tag{2.5}$$

The $\Pr(\boldsymbol{\theta}|\mathcal{M})$ term in the numerator of Equation 2.5 is the prior probability, which encodes our belief about the parameters $\boldsymbol{\theta}$ for model $\mathcal{M}$ *a priori*. Since in our case we want to fit a large sample of sources about which we know very little, we want uninformative priors which make no assumption about the spectral shape of

each source beyond constraining our parameter space to physically meaningful values. In general, this means we have used prior distributions uniform in logarithmic space (also known as a Jeffreys prior; Jeffreys 1946) for scale parameters (i.e. peak frequency and flux), and distributions uniform in linear space for location parameters (i.e. spectral indices). The limits placed on these distributions are to ensure physically sensible parameters, meaning we restricted spectral indices to values $-4 < \alpha < 4$, peak flux values to $0.03 < S_{\rm peak} < 15$ Jy and peak frequency values to $0.05 < \nu_{\rm peak} < 100$ GHz across all models where these parameters are applicable. The only exception to this is the curvature parameter c for the curved peaked spectrum model (Equation 2.3) for which we assume a truncated Gaussian prior which approximates the distribution of the curvature found for the sample described in Callingham et al. (2017) (there curvature is given as q). A more detailed breakdown of the priors for each model is provided in Table 2.2.

Table 2.2: Priors used in RadioSED for the four models outlined in Section 2.4.1, plus priors for the Gaussian process used for SEDs with covariant GLEAM data as discussed in Section 2.4.6. The `log_const` factor in the Gaussian process represents the vertical scale of the kernel, and the M_{00} factor represents the horizontal scale over which the kernel operates in frequency space.

Model	Parameter	Prior Type	Prior range
Power Law	S_0	log-uniform	0.1 – 10×10^7 mJy
(Equation 2.1)	α	linear-uniform	-4 – 4
Simple peaked spectrum	$S_{\rm p}$	log-uniform	30 – 15×10^3 mJy
(Equation 2.2)	$\nu_{\rm p}$	log-uniform	50 – 10×10^4 MHz
	$\alpha_{\rm thick}$	linear-uniform	0 – 4
	$\alpha_{\rm thin}$	linear-uniform	-4 – 0
Curved peaked spectrum	a	linear-uniform	-100 – -1
	b	linear-uniform	1 – 50
(Equation 2.3)	c	truncated Gaussian	$\mu = -0.5, \sigma = 5$, [-1×10^4, 0]
	$S_{\rm p}$	log-uniform	30 – 15×10^3 mJy
	$\nu_{\rm p}$	log-uniform	50 – 10×10^4 MHz
Re-triggered	$\alpha_{\rm thick}$	linear-uniform	0 – 4
(Equation 2.4)	$\alpha_{\rm thin}$	linear-uniform	-4 – 0
	S_0	log-uniform	0.1 – 10×10^7 mJy
	α	linear-uniform	-4 – 4
Gaussian process component	`log_const`	log-uniform	3×10^{-7} – 2
	M_{00}	log-uniform	5 – 30 MHz

The $\Pr(\boldsymbol{d}|\boldsymbol{\theta}, \mathcal{M})$ term in Equation 2.5 is the likelihood function $\mathcal{L}$, representing the probability of the SED data ($\boldsymbol{d}$), given the current model $\mathcal{M}$ with parameters $\boldsymbol{\theta}$. This is the key term that guides our model inference, as it allows us to update our belief about the model parameters using the available data. Since the data are assumed to be Gaussian distributed about the mean of model $\mathcal{M}$ with standard deviation given by the assumed noise of the signal(s), the probability density function in this case is a product of Gaussian PDFs over the vector of data. This is calculated simply using a multivariate Gaussian:

$$\mathcal{L} \equiv \frac{1}{\sqrt{(2\pi)^N \det \mathbf{C}}} \exp\left[-\frac{(\mathbf{d}-\mathbf{m})^{\rm T}\mathbf{C}^{-1}(\mathbf{d}-\mathbf{m})}{2}\right] \tag{2.6}$$

Here, N is the length of the data vector $\mathbf{d}$, $\mathbf{C}$ is the covariance matrix for these observations, and $\mathbf{m}$ is the vector of modelled flux densities. As the GLEAM sub-band fluxes are known to exhibit some covariance (Hurley-Walker et al., 2017, Section 5.4), this full matrix formulation of the likelihood must be used for fitting sources with GLEAM data (and indeed, could be used for any sources with additional covariant data points introduced in the future). However, in cases where all data points are independent and uncorrelated, Equation 2.6 reduces to:

$$\mathcal{L} = \frac{1}{\sqrt{(2\pi)^N \prod_i \sigma_i^2}} \exp\left[-\sum_i \frac{(d_i - m_i)^2}{2\sigma_i^2}\right] \tag{2.7}$$

where σ_i is the uncertainty reported on flux density d_i. The posterior distribution of parameters $\boldsymbol{\theta}$ for each individual model is inferred by traversing the prior parameter space until a stopping condition is reached where the model $\mathcal{M}(\bar{\theta})$ well describes the data with best-fitting parameter set $\bar{\theta}$. The exact implementation of this it outlined in Section 2.4.6 below.

Censored data

Since RadioSED is designed around several untargeted, large-area surveys, we can also make use of non-detections in a systematic way to better constrain sparse SEDs. For a source t that falls within a survey area but does not have a reported flux density, we can assume an upper limit of $\overline{x}$ Jy using the detection threshold given in the relevant survey paper. Therefore, a source may have flux density measurements $[d_1, ...d_i]$ and upper limits $[\bar{d_{i+1}}, .., \bar{d_k}]$. In such cases, we separate the likelihood into different components for observations and censored data. Following, for example, Klein & Moeschberger (2003), we define the likelihood for the censored data using the cumulative distribution function of the likelihood for uncensored observations. In this case with a Gaussian likelihood for observations, for the censored data we will use the the error function (erf()) to calculate $Pr(d < \bar{d_k})$. In this way the likelihood function becomes:

$$\mathcal{L} = \prod_{x=1}^{i} \frac{1}{\sqrt{2\pi\sigma_x}} \exp\frac{(d_x - m_x)^2}{2\sigma_x^2} \prod_{y=i+1}^{k} \left[1 + \mathrm{erf}\left(\frac{d_y - m_y}{\sigma_y}\right)\right] \tag{2.8}$$

where m is the flux density given by the model, d is the observed/censored flux density, and σ is the measurement uncertainty, as in Equations 2.6 and 2.7. All further steps proceed as with an SED comprising observed data only.

2.4.3 Parameter Estimation

Once the stopping condition is reached, we infer the most likely values for model parameters $\boldsymbol{\theta}$ which can be further used for a more detailed and physically-motivated analysis the radio source in question. Using a random sample of draws from the posterior $\Pr(\boldsymbol{\theta}|\mathbf{d}, \mathcal{M})$ to create several realisations of model $\mathcal{M}$, we derive a median value and 68% credible interval on each parameter θ. In the case of models 2.1 and 2.2, we can directly obtain measures of spectral indices, and in Model 2.2 we can also obtain the peak frequency and flux. For models 2.3 and 2.4 inferring useful measurements like spectral indices is more complex as these are not directly encoded in the functional parameters. Instead for Equation 2.3, the peak frequency is derived as the axis of symmetry for a log-parabola: $\log(\nu_{\mathrm{peak}}) = \frac{-b}{2c}$, with the corresponding peak flux S_ν obtained by evaluating the function at this frequency ν_{peak}. Spectral indices α_{low} and α_{high} are derived by fitting a power law as in Equation 2.1 either side of the peak, following (Torniainen et al., 2005) and (Orienti et al., 2007). In the case of Equation 2.4, intervals for the peak frequency, flux and optically-thin spectral index, α_{thin} can be read from the function and returned, but we choose to derive these empirically from the data by identifying turning points in the curve, in addition to returning intervals for a pair of parameters we term the *trough frequency* and associated *trough flux*, where the SED reaches a local minimum before increasing again at lower frequencies. These aid in our identification of steep spectrum sources with mild curvature, which are sometimes well-described by this model and falsely classified as re-triggered peaked spectrum sources (more on this below).

2.4.4 Model Comparison & Selection

Once all four models given in Section 2.4.2 have been applied to a single SED, our Bayesian inference method also allows us to robustly select which one of these is the preferred class of model thanks to the remaining term in the denominator of Equation 2.5.

This $\Pr(\mathbf{d}|\mathcal{M})$ term is the Bayesian evidence $\mathcal{Z}$, which serves to normalise the posterior probability, and can be calculated numerically by marginalising the likelihood and prior of a model over that model's parameters

$\boldsymbol{\theta}$. This marginalisation means the evidence captures information about how well the overall model class $\mathcal{M}$ fits the data, bearing in mind all possible parameter values $\boldsymbol{\theta}$ are explored from the prior distribution, not just the best fitting parameter set $\overline{\theta}$. In this way, a direct comparison of the evidence for different models enables us to determine which model $\mathcal{M}$ best fits the data without over-fitting and accounting for varying model complexity (e.g Sivia & Skilling, 2006, Section 4). This comparison term is usually called the Bayes factor $B_{1,2}$ and is found by taking the ratio:

$$B_{1,2} = \frac{\mathcal{Z}_1}{\mathcal{Z}_2} \tag{2.9}$$

so that a value of 2 means model $\mathcal{M}_1$ is twice as likely to be the true model as $\mathcal{M}_2$. In theory, the model with the highest Bayes factor from a pairwise comparison of all four models is will be the best fitting model without over-fitting the data.

In practice, there is always an uncertainty associated with the numerically-derived model evidence $\mathcal{Z}$ (e.g. Speagle, 2020, Appendix A6), and our choice of prior distributions (especially their limits) may have a mild effect on the calculated evidence, and hence on the derived Bayes factors. Accordingly, we follow the evidence thresholds of Kass & Raftery (1995), and consider any Bayes factors $0 < B_{1,2} < 0.5$ to represent only a marginal preference for $\mathcal{M}_1$ over $\mathcal{M}_2$. In instances where this is the case, and $\mathcal{M}_1$ is more complex than $\mathcal{M}_2$ (i.e. the number of parameters $N(\boldsymbol{\theta_1}) > N(\boldsymbol{\theta_2})$), we reject $\mathcal{M}_1$ as the most likely model to represent the data in favour of $\mathcal{M}_2$, following the principle of Occam's Razor.

2.4.5 Source Classification

Once the most likely model type has been determined from a Bayesian analysis, the classification of the source is taken from this model following the definitions in Section 2.4.1. However, there are some subtleties in the use of these models and how they relate to the canonical definition of a peaked spectrum source, which should be taken into account when creating a robust sample of peaked spectrum AGN. We outline here three refinements to classifications that are performed as part of RadioSED, although we note that further refinement may be useful to the user after close inspection of their sources.

Firstly here, we note that there have historically been two main methods for identifying peaked spectrum sources. In small samples, often involving targeted followup observations, a peaked spectrum source has been defined as having an optically thin spectral index, $\alpha_{\rm thin} < -0.5$ (where $\alpha < 0$ indicates a flux density which is decreasing as frequency increases), mimicking the canonical boundary between steep and flat spectrum sources (Urry & Padovani, 1995; O'Dea, 1998), and a spectral curvature ($\alpha_{\rm thick} + \alpha_{\rm thin} > 0.6$) (de Vries et al., 1997; Edwards & Tingay, 2004).

Across larger samples of up to a few thousand sources but with sparse flux density measurements, the identification has been done with as few as three independent flux densities which are used to create 2, two-point spectral indices at high and low frequencies (what constitutes 'high' and 'low' is not fixed but dependent on the availability of the data in any given study). These spectral indices are plotted against each other in radio colour-colour space, where one quadrant of the resultant plot with $\alpha_1 < 0$ and $\alpha_2 > 0$ indicates the presence of a spectral peak in some range of frequencies spanned by the flux density measurements used in calculation. This is the approach taken by Murphy et al. (2010), Callingham et al. (2017) and others (see e.g. Figure 15 in Murphy et al. (2010) for a good example of this method), but it it does not provide a direct correspondence to the first method of identification. Indeed, Callingham et al. (2017) recognised the $\alpha_{\rm thin} < -0.5$ requirement in the first method as an arbitrary threshold in a continuous spectrum of possible spectral indices, and instead used the radio colour-colour diagram to classify any sources with $\alpha_{\rm low} > 0.1$ and $\alpha_{\rm high} < 0$ as peaked spectrum for their work (note again that they use spectral indices $\alpha_{\rm low}$ and $\alpha_{\rm high}$ between fixed frequencies, rather than $\alpha_{\rm thin}$ and $\alpha_{\rm thick}$ determined by the location of the spectral peak). We strike a balance between these two approaches since our prior space for the peaked spectrum models extends beyond the $\alpha_{\rm thin} < -0.5$ threshold (see Table 2.2), and we classify any source best fit by Models 2.2, 2.3 & 2.4 but with $0.1 > \alpha_{\rm thin} > -0.5$ or $\alpha_{\rm thick} + \alpha_{\rm thin} < 0.6$ a '*soft peaked spectrum*' source. This allows for easy comparison with older, literature-derived samples, whilst also recognising the continuous distribution of spectral indices in physical sources.

Secondly, as mentioned in Section 2.4.3, some sources for which the most likely model is a re-triggered

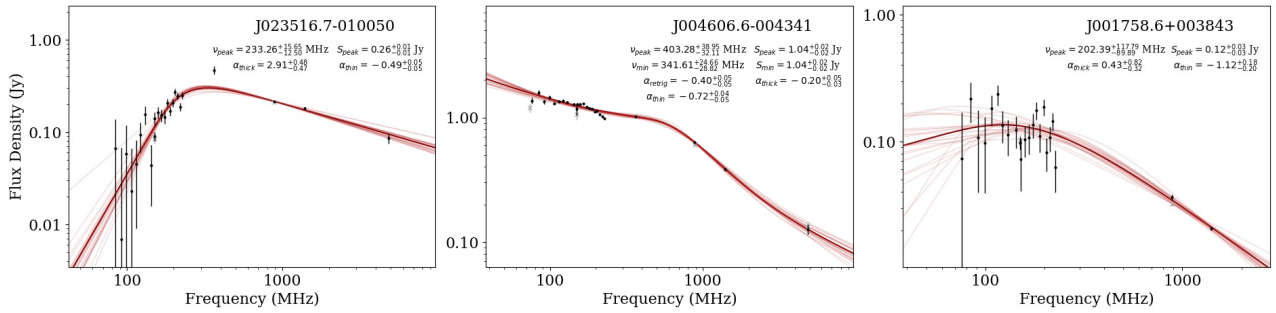


Figure 2.6: An example of each type of source which is best fit by a peaked spectrum-like model, but which requires some simple analysis of the fit parameters to fully understand the subtleties of the SED. Left is a source that would canonically be deemed flat spectrum with $\alpha_{\rm thin} > -0.5$, but is classified here as a "soft peaked spectrum" source. Middle is a source which is best fit by the re-triggered peaked spectrum model, and yet all three spectral indices have $\alpha < 0$, indicating it is in fact a steep spectrum source with some degree of spectral curvature but no turnover. Right is a source best fit by Model 2.2, and yet the fit parameters are poorly constrained with $\Delta S_{\rm peak} \geq 0.4 \times S_{\rm peak}$, so the source is re-classified as complex. The parameter estimates with derived uncertainties for the best fitting model are shown in the top right of each sub-plot for reference. The dark red line is the best fit from the posterior of the selected model type, light red lines are random draws from the posterior to represent graphically the range of best-fitting parameters.

model (Model 2.4), may in fact only exhibit slight spectral curvature and no discernible peak or turnover. In this case, the peaked classification is rejected in favour of a 'complex' or steep/flat/inverted spectrum classification (dependent on the spectral indices inferred for the model).

At present, one final post-processing step is performed to identify poorly-constrained peaked spectrum classifications based on the parameter estimation described in Section 2.4.3. If a source is classified as peaked spectrum, but the uncertainty on its peak frequency and/or peak flux density is so large that $\Delta{\rm val} \geq 0.4 \times {\rm val}$, the peaked spectrum classification is marked as uncertain, but we do not modify the choice of most likely model. This step was included after a manual inspection of the synthetic data in Section 2.5.2 and the test sources discussed more in Section 2.5.3, and is found to identify a number of uncertain peaked spectrum classifications, which may occur due either intrinsic variability distorting the SED, or low S/N flux density measurements with large uncertainties.

Sources requiring each type of additional analysis for robust classification are shown in Figure 2.6. These are real sources drawn from our pilot field which will be discussed in more detail in the second paper in this series.

2.4.6 Implementation

Our model inference is implemented in Python using the Bilby Bayesian Inference package (Ashton et al., 2019) as an interface. Dynesty (Speagle, 2020) performs the actual model inference using nested sampling, as this allows for robust evidence estimates which are necessary for model comparison (Skilling, 2004, 2006). Within Dynesty, the dynamic nested sampling algorithm is used to improve performance by allocating the number of live points in a run dynamically, adding more points in a series of groups or *batches* as the sampler moves towards successively higher areas of likelihood (Higson et al., 2019a). This allows the sampler to more accurately capture a complex posterior. The default behaviour for RadioSED is to ensure 1,000 effective samples in the final posterior or to progress through 10 batches, whichever comes first. For sources with GLEAM data, we also make use of the George Gaussian Process package (Ambikasaran et al., 2016) to jointly fit for the data covariance and the model mean SED, since the exact form of this covariance is unknown (Hurley-Walker et al., 2017, Section 5.4).

As of publication date, RadioSED makes use of Dynesty version 2.1.2 (Koposov et al., 2023), Bilby version 2.1.1, and George version 0.4.0. An overview of the steps involved in running RadioSED on a given source is shown in Figure 2.7.

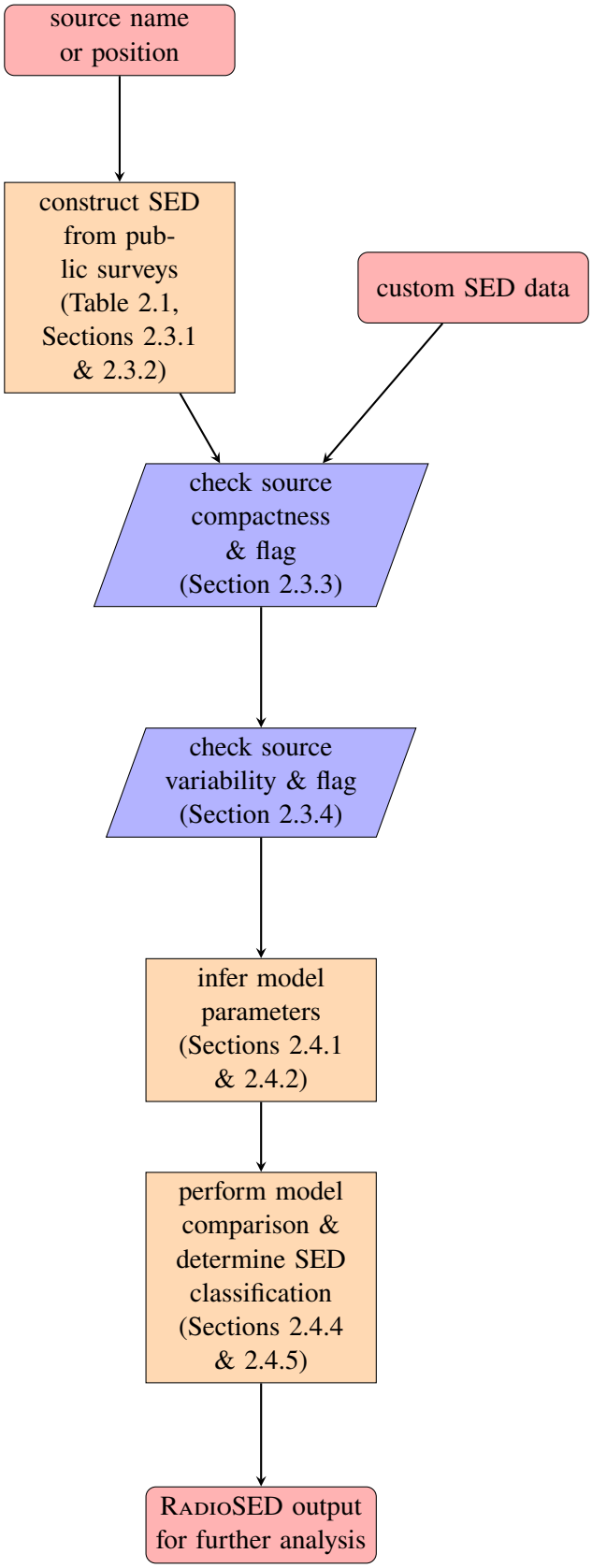


Figure 2.7: An overview of the RadioSED inference procedure for an individual source. Users can either input a source name/position and allow RadioSED to construct the SED using the method outlined in this paper, or input their own radio SED data to be fit.

2.5 RadioSED Verification & First Results

We outline here a first application of RadioSED to both synthetic and real-world data, demonstrating its accuracy and reliability as a tool for identifying young, obscured AGN from radio continuum data alone. We focus first on a discussion of individual models and potential sources of uncertainty in the modelling procedure, before moving on to consider whether our inhomogeneous SEDs could lead to contamination of any eventual peaked spectrum sample in the context of AGN variability. Finally, we examine the performance of RadioSED in recovering the SEDs of radio sources with well-characterised spectral shapes from the literature.

2.5.1 Parameter Recovery & Sampling Errors: Nestcheck

Here we examine the reproducibility of the results output by RadioSED using synthetic broadband flux density measurements. We validate output for each of the 4 models outlined in Section 2.4.1 separately as detailed below.

Since nested sampling by construction never reaches a steady state, but proceeds always towards regions of the posterior with a higher likelihood, it can be difficult to validate nested sampling-based approaches. In particular, quantifying implementation-specific uncertainties is tricky as we cannot rely on autocorrelation-based heuristics which are typically used in MCM approaches to determine whether successive runs will produce reproducible output (i.e. whether the runs converge and the posterior samples are suitably uncorrelated). Therefore, to ensure that the fits produced by RadioSED are accurate and reproducible, we make use of the Nestcheck package (Higson et al., 2019b), which is designed explicitly to address this difficulty. The package provides three main useful features; first it characterises the uncertainty of a nested-sampling algorithm between successive runs, secondly it quantifies implementation-specific sources of error, and finally it provides useful diagnostic plots to summarise these effects. This third feature, the diagnostic plots, also means the output of Nestcheck can be easily adapted to test the ability of RadioSED to recover injected parameters as well.

We consider first potential inter-run variations of RadioSED. The nested sampling algorithm is stochastic in nature, and so some low-level variation of the posterior and evidence output between runs is to be expected. However, if this variation is too high, estimates of model parameters and Bayesian evidence become unreliable. We do not expect a large inter-run variation here as RadioSED makes use of a standard and robust nested sampling package which incorporates stopping criteria based on minimising uncertainty in both the posterior and evidence (Speagle, 2020, Appendix A). Indeed this is confirmed by Nestcheck's diagnostic plots, shown in Figures 2.8 - 2.11. These plots present the uncertainty of the posterior of two separate nested sampling runs for a synthetic spectrum representing each model type outlined in Section 2.4.2. The synthetic spectra were generated

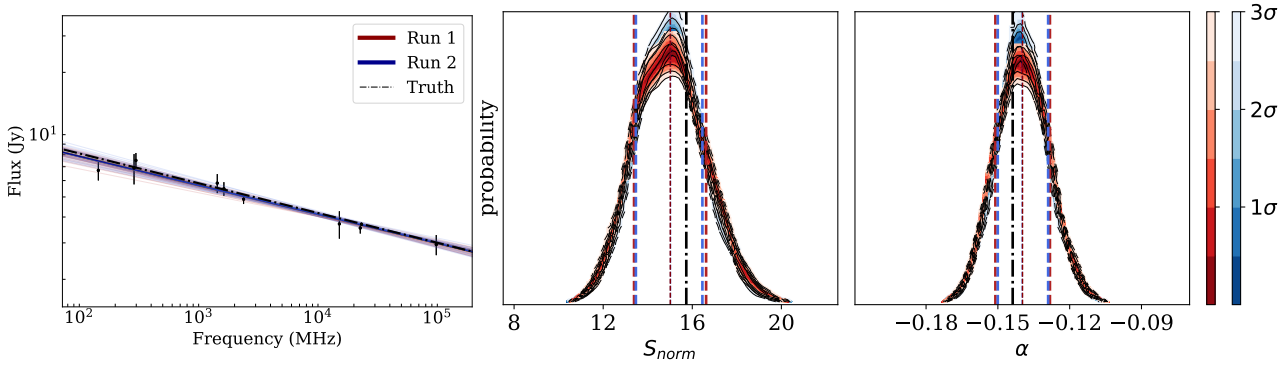


Figure 2.8: Left: output fits from RadioSED applied twice to a synthetic source obeying a power law of the kind in Equation 2.1. Dark lines show the best fit from the model posterior for each run (red and blue for runs 1 and 2, respectively), and lighter lines show random draws from the posterior. Dot-dashed line in black is the true SED of the source, from which observations are drawn with a 10% Gaussian scatter.Right: diagnostic plots showing the pdf as a function of model parameters and produced by Nestcheck for repeated fitting of this same source. Dot-dashed black lines indicate injected parameter values, dotted coloured lines indicate output parameter means and dashed lines are the associated 1-σ errors.

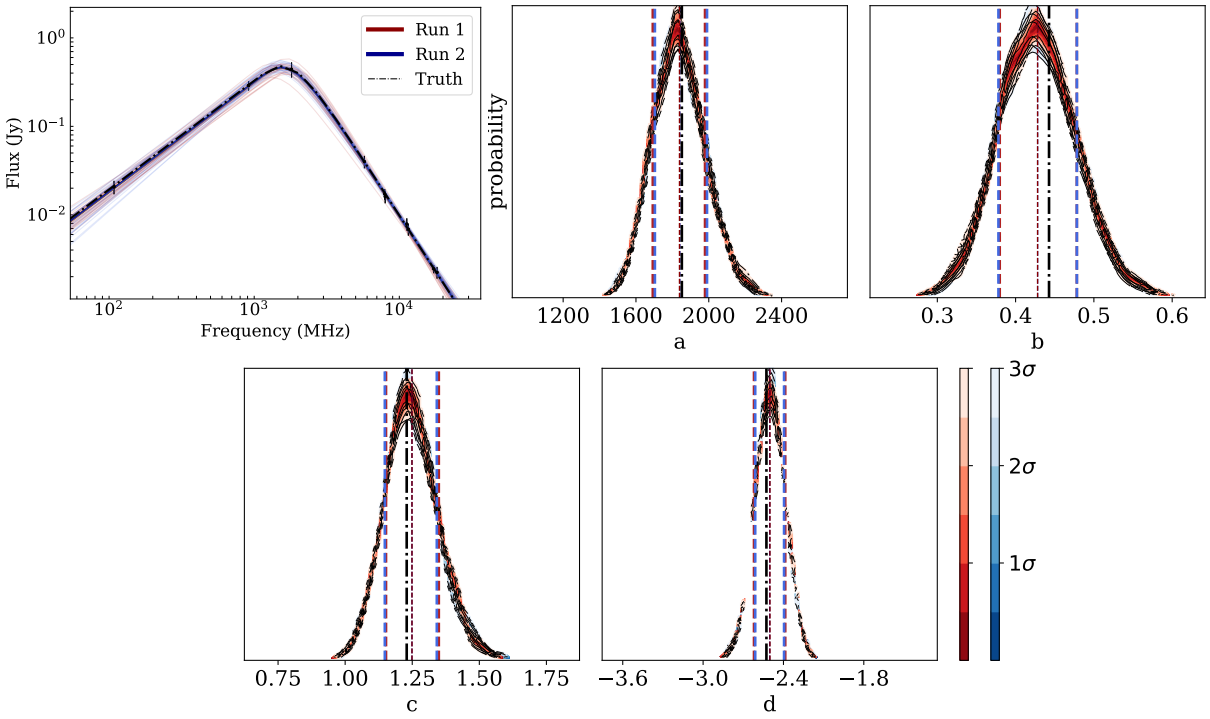


Figure 2.9: The same as Figure 2.8 but for a source obeying Model 2.2.

by sampling from the prior for each model type to create an SED shape, and then sampling from this SED at 10 random frequencies to create flux density measurements. To emulate real data for a non-varying source, these flux densities were given a 5% Gaussian scatter, with uncertainties set at 10% of this flux value, plus an additional component of Gaussian noise, again at the 5% level. After running RadioSED twice, uncertainties on each posterior were obtained by bootstrap re-sampling the individual runs (Higson et al., 2018). The fact that both posterior probability distributions, indicated by the different colours, occupy similar regions of probability space demonstrates that the stochasticity of nested sampling does not have a substantial effect on the estimates output by RadioSED.

Secondly, and perhaps more importantly, we consider implementation-specific effects. These may arise due to everything from subtle inaccuracies in the construction of the prior and likelihood distributions, to the structure of the nested sampling software itself (Higson et al., 2019b). It is therefore possible that such effects may arise in RadioSED due to elements of the code which build upon established packages. Nestcheck derives the uncertainty in the Bayesian evidence ($\log \mathcal{Z}$) and parameter estimates due to implementation-specific effects by calculating the uncertainty over several runs and comparing this to the uncertainty expected from stochasticity alone (calculated via bootstrapping a single run, as discussed above). These calculations were performed using 10 runs for each model type, and the results are summarised in Figure 2.12.

In Figure 2.12, the bars at the bottom indicate the fraction of the total standard deviation σ_{var} which is due to implementation-specific effects. If this value exceeds $\frac{1}{\sqrt{2}}$ as indicated by the dashed horizontal line, it means that the uncertainty in a parameter estimate is due mainly to implementation effects, if it is less than this threshold then the uncertainty is largely caused by the inherent stochasticity of nested sampling. This is due to the expected form of the variance, as outlined in Higson et al. (2019b), where $\sigma^2_{\text{val}} = \sigma^2_{\text{imp}} + \sigma^2_{\text{bootstrap}}$. For both the power law (Equation 2.1) and simple peaked spectrum (Equation 2.2) models, the stochasticity of nested sampling dominates the uncertainties, while for both the peaked spectrum with curvature (Equation 2.3) and re-triggered models (Equation 2.4), implementation effects dominate the uncertainty of all except the Bayesian evidence ($\log \mathcal{Z}$). However, this is only meaningful when considered in the context of the overall fractional uncertainty on these values, which is shown along the top of Figure 2.12. As we can see, the fractional error is below 5% for all parameters and evidence values (the dot-dashed black line), and below 1% (dot-dashed grey line) for all except the Bayesian evidence for models 2.3 and 2.4, and the normalisation factor for model 2.4. Since the Bayesian evidence is used in model comparison and selection, the fact that the uncertainty on

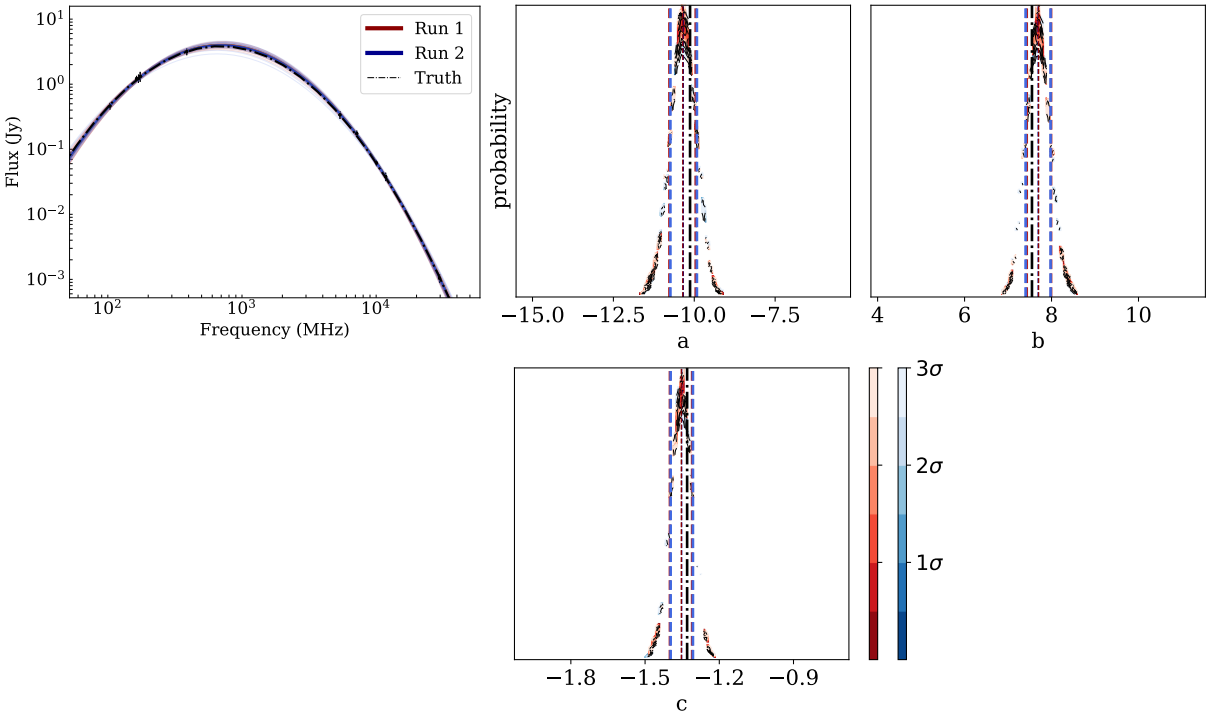


Figure 2.10: The same as Figure 2.8 but for a source obeying Model 2.3.

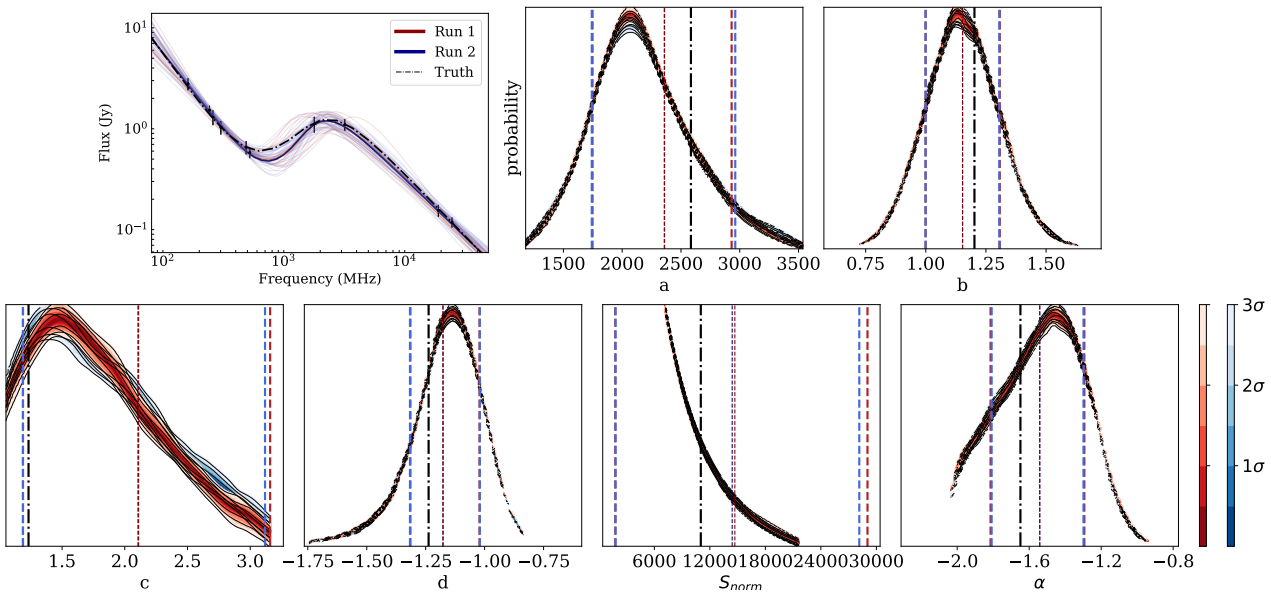


Figure 2.11: The same as Figure 2.8 but for a source obeying Model 2.4.

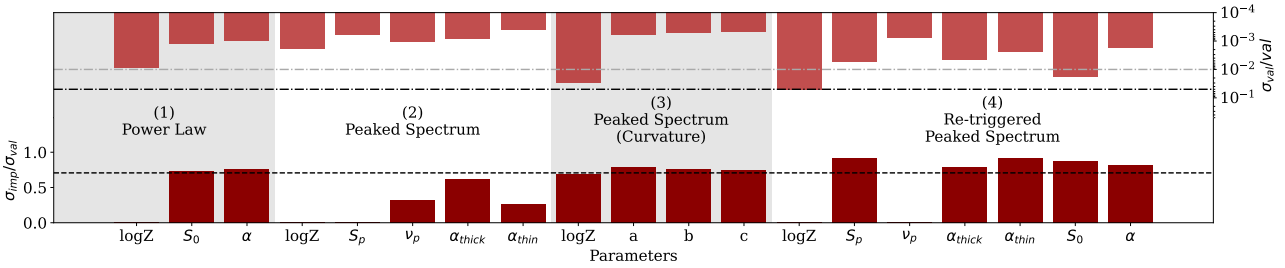


Figure 2.12: Summary plot showing the small fractional uncertainties in model parameters derived using RadioSED (top), as well as breaking down the uncertainty into contributions from implementation effects, and from the inherent stochasticity of nested sampling algorithms (bottom). All uncertainties sit close to or below the black dot-dashed (top) and black dotted (bottom) lines, indicating that RadioSED will be sufficiently accurate for deriving SED parameters with which to perform meaningful, physics-driven analysis. Uncertainties were derived using the Nestcheck package. The black dot-dashed line indicates a fractional error of 5%, and the grey a fractional error of 1%.

these values for models 2.3 and 2.4 is at the 5% level reinforces our choice to implement a correction to the model selection procedure, as outlined in Section 2.4.4. Overall though, while implementation-specific effects contribute equally or more than stochastic effects in models 2.1, 2.3 and 2.4, the fractional uncertainties remain sufficiently low for all parameters of interest.

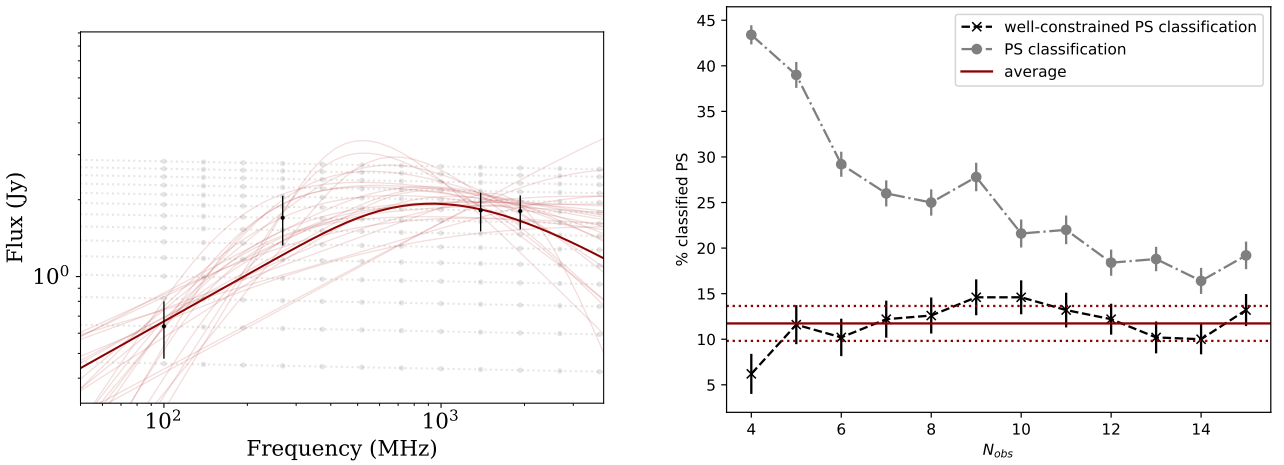
Because the diagnostic plots produced by Nestcheck show the uncertainty on the posterior distribution with reference to the median value of the model parameters returned by a sampling run, it is also possible to use these results to validate how well RadioSED can recover the true parameters for various SED shapes. In Figures 2.8 - 2.11, the dotted red and blue lines indicate the mean values for each parameter derived from each nested sampling run, while the dashed lines represent 68% confidence intervals on these values as derived from the posterior distributions. To make these plots useful for verifying how well RadioSED can recover the true parameters of an SED, we have modified Nestcheck's diagnostic diagrams, adding the true injected parameter value for each spectrum as the dot-dashed line in black. We have also added an additional sub-plot (the leftmost sub-plot in each Figure) showing the synthetic SED along with the best output from each run of RadioSED, 50 draws from the posterior to give an indication of parameter range, and the true SED shape. In all cases RadioSED has recovered the injected parameter values to within the 1-σ uncertainties shown in the figures, demonstrating it is suitably accurate for use on real-world observations.

2.5.2 Variability & Sample Contamination

As mentioned in Section 2.3.4, it is conceivable that the radio variability of AGN could lead to false detections of peaked spectrum sources when combining multi-epoch flux density measurements. While the use of the ALMA Calibrator Catalogue provides one measure of variability at sub-mm frequencies, variability across much of the frequency space used in our modelling is difficult to quantify from the surveys incorporated into RadioSED. Therefore, to better understand the rate of false peaked spectrum detections due to variability, we have undertaken some simple, Monte Carlo-based simulations.

We began by creating a mock-variable flat spectrum source from which to draw 'multi-epoch' observations. This synthetic source, shown in Figure 2.13a, was created by using Equation 2.1 with a spectral index $\alpha = -0.02$ and an amplitude varying between $0.5 \leq S_0 \leq 3.2$, to mimic broadband spectral variability. We chose a flat spectrum source since McConnell et al. (2012) found sources with $\alpha > -0.5$ exhibited most variability between 4.8-8.65 GHz. Overlaid on this variable source is a mock set of five multi-epoch observations, made by selecting a single flux density measurement at 5 unique frequencies. By repeatedly sampling this synthetic source using between 4-15 unique frequency 'measurements', we can construct a suite of mock SEDs to be run through RadioSED, using a Monte Carlo method to explore false detection rates as a function of $N_{\rm obs}$, the number of observations at unique frequencies.

After creating 500 mock SEDs per unique $N_{\rm obs}$ and running these through RadioSED, the false detection rate of peaked spectrum sources was found to be approximately 11.8±1.9% from amongst variable, flat spectrum



(a) Mock variable flat spectrum source (dotted grey lines), with an overlaid "multi-epoch" SED (black points) and fit from RadioSED (dark red). Uncertainties are 10% of the flux value plus an additional 10% Gaussian scatter.

(b) Fraction of mock variable SEDs misclassified as peaked spectrum due to multi-epoch sampling. Note especially that the rate appears much higher if we consider all sources misclassified as peaked (grey dot-dashed), and do not discount the unreliable or poorly constrained ones (black dashed) as defined in Section 2.5.2. This shows the importance of our reliability criteria. Uncertainties on all measurements are bootstrapped. The dark red line shows the average misclassification rate across all $N_{\rm obs}$, and the red dotted lines are the 1σ uncertainties on this value.

Figure 2.13: Mock spectral tests of contamination by variable sources when using radio catalogues from multiple epochs.

sources. This was after discounting sources unreliably classed as peaked spectrum, as defined in Section 2.5.2 and shown in Figure 2.13b. We also found that in 96% of mis-classified sources, the false peak was three measurements or fewer from the edge of the observing band, which may be a helpful heuristic for identifying mis-classified sources in the future, though further work will need to be done in comparing this with real data for secure peaked spectrum sources.

Considering this in the broader context of peaked spectrum samples, we expect the overall contamination of any peaked spectrum sample produced using RadioSED to be low. If we assume conservatively that 5% of radio AGN are significantly variable in some region of observed frequency space, based on the discussion in Section 2.3.4, and we expect 12% of these to be mis-classified as peaked spectrum from our simulations, that is only 0.6% of the radio population mis-classified as peaked spectrum. Or, put another way, if 15% of an unbiased radio population is classified as peaked spectrum, we expect about 4% of these classifications to be incorrect. This is much lower than the mis-classification fractions seen in for example Torniainen et al. (2007) and Edwards & Tingay (2004), who re-observed known peaked spectrum sources identified in single epochs, and revealed variability in a large fraction of them by utilising a multi-epoch analysis. Since RadioSED begins from a multi-epoch approach, such contamination is in fact likely to be much lower in the peaked spectrum samples produced in the method outlined in this work. A deeper consideration of radio variability and its impact on multi-epoch radio SEDs will be considered as part of our second paper in the context of a larger sample of radio sources with real world, observational data.

2.5.3 Test Sources

With the characteristics of RadioSED better understood we now apply it to a selection of radio sources with various published SED shapes. The purpose of this exercise is to determine how well our framework can recover expected spectral parameters for sources known to fall into each of the four categories given by equations 2.1 - 2.4.

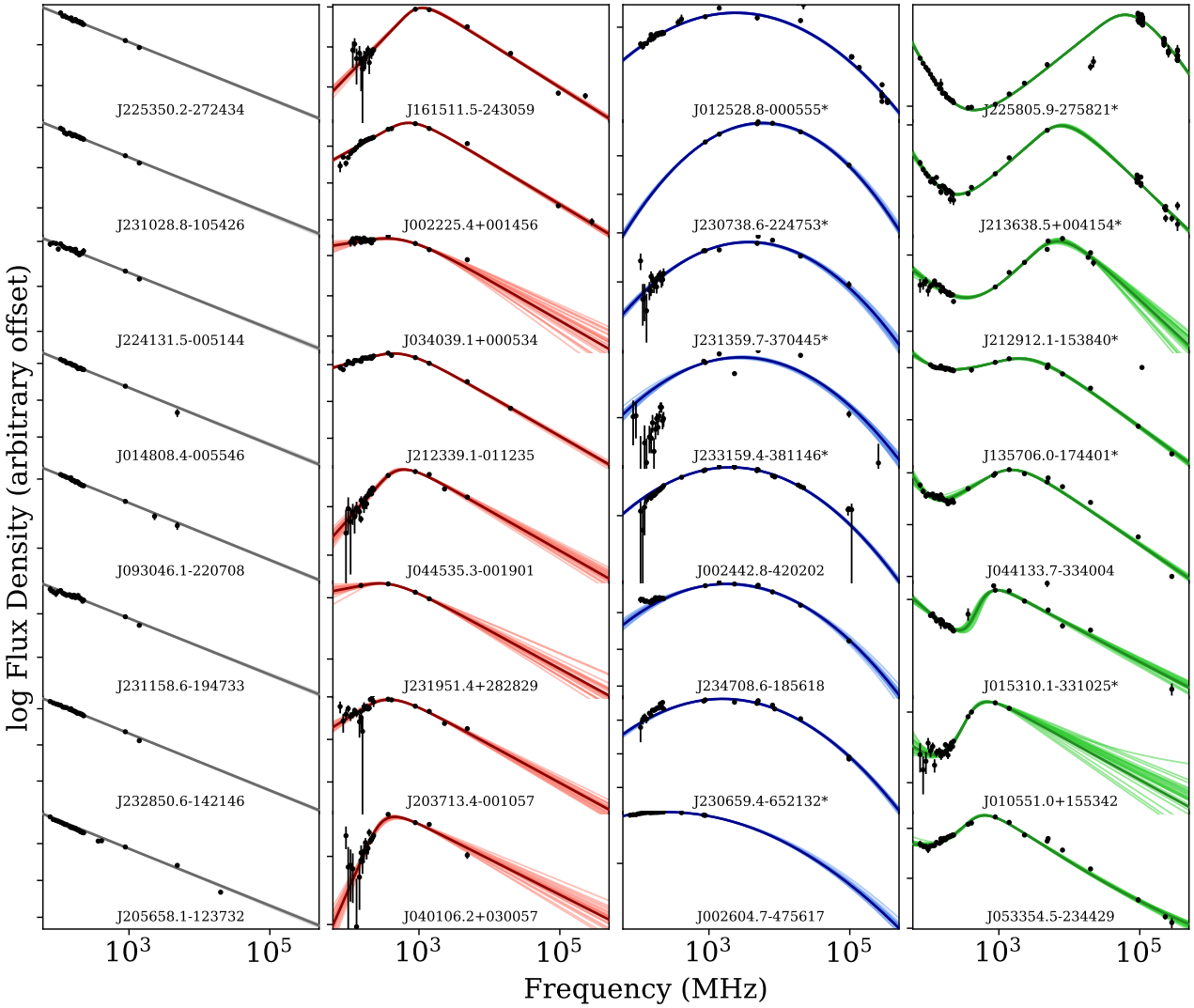


Figure 2.14: A representative sample of sources from the literature with well-characterised SED shapes used to verify that RadioSED is able to recover expected spectral parameters. Columns map onto the different SED shapes from Figure 2.5, so these sources have previously been identified as having a power law SED corresponding to Model 2.1 (left column), a peaked spectrum following Model 2.2 (second left), a peaked spectrum with significant curvature after Model 2.3 (right middle), or an upturn at lower frequencies (right), fit by Model 2.4 as described in Section 2.4.1, and shown in Figure 2.5. The y-axis for each individual source has been dynamically scaled to show the full extent of the model between 100 MHz-500 GHz, which causes the slopes of all the Power Law sources (left column) to appear almost identical. For each source, the darkest line shows the best fit of the most probable model from RadioSED, while the array of lighter lines represents 25 draws from the posterior of this model to give an indication of the spread in parameter values.

To obtain a selection of power law (Model 2.1) and peaked spectrum (Model 2.2) sources, we consulted the SPECFIND V3.0 catalogues (Stein et al., 2021). We constrained our search to spectra with at least 8 flux density measurements in the final SPECFIND SED, and a declination $-30^\circ < \delta < 30^\circ$ to obtain sources within the area where the RACS-low and NVSS surveys overlap. In addition to SPECFIND, we consulted Callingham et al. (2017) to obtain a selection of sources with a low-frequency upturn (Model 2.4), as these sources were explicitly discussed in Section 6 of that work, and a manual inspection of the SEDs for sources in the AT20G high angular resolution catalogue (Chhetri et al., 2013) revealed a number of peaked spectrum sources with significant curvature based on their changing spectral indices as a function of flux density (Model 2.3). Some additional sources from the literature known to the authors were further added to this sample, to allow us to present an equal number of sources within each model category. The fits of these sources obtained using RadioSED are given in Figure 2.14, and the parameters of their fits are presented in Table C.1.

The parameters given by RadioSED are broadly consistent with those found in the literature for this sample of sources, bearing in mind the constraints under which many literature parameters were originally derived. For example, peaked spectrum sources identified in Stein et al. (2021) assume a fixed peak frequency of either 325 MHz or 1.4 GHz, and the spectral indices are typically derived as α_{low} and α_{high} between fixed frequencies, rather than $\alpha_{\rm thick}$ and $\alpha_{\rm thin}$ either side of the peak, as is done in RadioSED. More importantly, it is apparent from Figure 2.14 that our framework can appropriately model SEDs spanning this diverse range of shapes and spectral parameters. The successful fitting of these sources is therefore a promising, early indication of the of the efficacy of RadioSED in large, statistical studies of radio spectral shapes.

However, as noted in Sections 2.3.4 and 2.5.2, there remains the possibility of blazar contamination amongst any given peaked spectrum sample. Indeed from amongst these sources, a search of the SIMBAD database (Wenger et al., 2000) revealed 10 sources with either a primary or secondary blazar classification. The names of these sources are marked with an asterisk in Figure 2.14 and Table C.1. Of these, 9 sources have blazar classifications derived from the Massaro et al. (2015) Roma-BZCAT catalogue where they are classed as 'quasar-like' blazars, meaning they were identified on the basis of a flat radio spectrum between 0.8-5 GHz and broad, optical emission lines. The remaining source, J230738.6-224753, is identified as a blazar candidate in the CGRaBS survey (Healey et al., 2008). However, the broadband SED modelling performed here reveals significant curvature in each of the blazar-identified 'flat spectrum' sources, highlighting the need for broadband coverage when searching for beamed radio emissions. Indeed in the original Roma-BZCAT paper, Massaro et al. (2009) noted that GPS sources and blazars are sometimes confused in radio-selected samples due to poor spectral coverage "both in frequency and time". As a result of this difficulty, any known GPS sources otherwise meeting the Roma-BZCAT criteria for a blazar were included in the first edition of that catalogue. Of course in the case of a secure blazar classification, an object cannot be a true Peaked Spectrum source in the canonical sense of young radio AGN with minimal variability, but it is clear that we must be careful to check any such classification against the full set of mutliwavelength data available for a source. In the case of the blazar-identified sources shown here, many do exhibit high X-ray luminosities and γ-ray emissions consistent with true blazars, but we have still chosen to include their radio SED fits for demonstrative purposes. Not only do they reinforce the ability of RadioSED to recover peaks previously identified in the literature, but they also emphasise the need for additional tools, beyond SED analysis, to identify true Peaked Spectrum sources. We therefore recommend to any users of RadioSED that a crossmatch with up to date blazar catalogues be used in conjunction with the outputs from our framework, and indeed our forthcoming work on our chosen pilot field will make use of just such ancillary data.

2.6 Conclusions & Future Work

In this paper we have presented RadioSED, a new SED fitting framework tailored to broadband, radio frequency observations and capable in particular of identifying peaked spectrum sources across a wide range of ages and redshifts. RadioSED incorporates pre-matched, large area radio catalogues spanning 0.072 – 370 GHz to provide the user with meaningful, consistent SEDs with which to do science (Section 2.3), and these SEDs are reliably fit using a rigorous Bayesian approach (Section 2.4). We have validated RadioSED's ability to recover injected parameters by making use of the Nestcheck package (Section 2.5.1), and our own simple Monte Carlo simulations have shown that the misclassification rate of variable sources in any given sample is likely to be

low, sitting at 11.8 ± 1.9% (Section 2.5.2). Not only that, but by applying this framework to a representative sample of sources with well-known SED shapes from the literature, we have shown that RadioSED can reliably recover the expected spectral properties of real radio sources using only the radio surveys outlined in Section 2.3.1, and the models from Section 2.4.1.

As to future work, we are currently completing an investigation into the radio AGN within the Stripe 82 field (Kerrison et al. in prep), for which a full, multiwavelength analysis is likely to reveal important details about their host galaxies and local environments (e.g. Hogan et al., 2015). A full investigation of this test field in comparison to synthetic spectra like those discussed here in Section 2.5.2 will also help to further clarify the effects of variability on multi-epoch SED fitting, all of which will pave the way for larger, more reliable, and more complete samples of radio AGN in various classes. Alongside this, the modular nature of RadioSED means that it can be continually improved as the next generation of large area surveys collect in public repositories, and our understanding of the nuances behind radio SED modelling grows. In particular, new and forthcoming IPS measurements at both megahertz and gigahertz frequencies could aid greatly in the identification of compact sources (Morgan et al., 2022; Chhetri et al., 2023), while the GLEAM-X (72-231 MHz; Hurley-Walker et al. (2022a)) RACS-mid (1.36 GHz; Duchesne et al. (2023)) and RACS-high (1.65 GHz) surveys will provide additional continuum data to better constrain individual SEDs. Ultimately, RadioSED promises to help us better understand the formation, triggering and local environments of radio AGN by working alongside existing IR-Optical-UV SED fitting codes to provide a more holistic, multiwavelength picture of AGN of all ages and evolutionary states.

Acknowledgements

The authors wish to thank the anonymous referee for their comments and suggestions which led to many improvements in the final version of this paper.

This research was supported by an Australian Government Research Training Program (RTP) Scholarship. This research was also supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013.

This research made use of the VizieR catalogue access tool, cross-match service XMatch and SIMBAD database provided by CDS, Strasbourg (Ochsenbein et al., 2000; Wenger et al., 2000). We also made use of Astroquery (Ginsburg et al., 2019) and Astropy:[4] a community-developed core Python package and an ecosystem of tools and resources for astronomy (Astropy Collaboration et al., 2013, 2018, 2022).

This scientific work uses data obtained from Inyarrimanha Ilgari Bundara / the Murchison Radio-astronomy Observatory. We acknowledge the Wajarri Yamaji People as the Traditional Owners and native title holders of the Observatory site. CSIRO's ASKAP radio telescope is part of the Australia Telescope National Facility[5]. Operation of ASKAP is funded by the Australian Government with support from the National Collaborative Research Infrastructure Strategy. ASKAP uses the resources of the Pawsey Supercomputing Research Centre. Establishment of ASKAP, Inyarrimanha Ilgari Bundara, the CSIRO Murchison Radio-astronomy Observatory and the Pawsey Supercomputing Research Centre are initiatives of the Australian Government, with support from the Government of Western Australia and the Science and Industry Endowment Fund.

Data Availability

All flux density measurements used to construct the SEDs in this work are drawn from publicly available catalogues accessible via the CDS VizieR catalogue service. The code for RadioSED, including crossmatch details for the radio surveys discussed, is accessible on Github and available for use: `https://github.com/ekerrison/RadioSED/` (DOI: 10.5281/zenodo.8336846). Data on the individual fits presented in this paper is available upon request to the authors.

[4] http://www.astropy.org

[5] https://ror.org/05qajvd42

3

A new sample of PS sources

Emily F. Kerrison[1,2,3], Elaine M. Sadler[1,2,3], Vanessa A. Moss[3,1], Elizabeth K. Mahony[3], Laura Driessen[1], Kathryn Ross[4,5], Kovi Rose[1,3], Dougal Dobie[1,6], and Tara Murphy[1,6]

[1]Sydney Institute for Astronomy, School of Physics A28, University of Sydney, NSW, 2006, Australia
[2]ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia
[3]ATNF, CSIRO Space and Astronomy, PO Box 76, Epping, NSW 1719, Australia
[4]ICRAR, International Centre for Radio Astronomy Research, Curtin University, Bentley, WA 6102, Australia
[5]AusSRC, Australian SKA Regional Centre, Curtin University, Bentley, WA 6102, Australia
[6]ARC Centre of Excellence for Gravitational WAve Discovery (OzGrav), Hawthorn, Victoria 3122, Australia

This chapter is published as *RadioSED - II. Discovering the peaked spectrum radio sources in Stripe 82.* Kerrison E. F., Sadler E. M., Moss V. A., Mahony E. K., Driessen L., Ross K., Rose K., Dobie D., Murphy T., 2025. *Monthly Notices of the Royal Astronomical Society* 543, 3895–3914. doi:10.1093/mnras/staf1643

There is nothing more useless than another unidentified radio source!

Condon (2002), on use and purpose of radio surveys

This chapter in context

With a new framework for identifying PS sources in place, it is time to put it to the test. In this chapter, we apply RadioSED to the sources in a well-defined, $\sim 300\,\mathrm{deg}^2$ region of sky (corresponding to approximately 1 per cent of the full survey area of ASKAP-FLASH). From within this area, we construct a new, uniform sample of 359 PS sources, which is more than an order of magnitude larger than the set of previously known PS sources in the field. Our new sample is also complete for observed peak frequencies ν_{peak} spanning 70 MHz − 20 GHz at flux densities $S_{\mathrm{peak}} \geq 0.85\,\mathrm{Jy}$, an unprecedented bandwidth made possible by DR1 of GLEAM-X (Ross et al., 2024), which provides critical, low frequency sensitivity improvements over its predecessor, GLEAM. Furthermore, as this work is only using pre-existing data from large-area radio surveys, this approximate areal density of 1 source/sq. deg. can be readily scaled up to a predicted $\sim 30,000$ PS sources across the southern sky visible to these surveys, and to ASKAP-FLASH. Samples of this size will be readily achievable on short timescales provided the imminent, all-sky data release of GLEAM-X (Hurley-Walker et al., 2022a).

The gains to be made with samples this large are great, but J. Condon is right that even the largest sample of PS sources has limited utility without multi-wavelength ancillary data. At a bare minimum, spectroscopic redshifts are required to constrain the luminosity of a source and its restframe turnover frequency, obtaining from the second an estimate of its linear size (e.g O'Dea & Baum, 1997; Jeyakumar, 2016). Beyond that, the redshift of a PS sample allows us to constrain population-scale evolution (e.g Slob et al., 2022; Zhai et al., 2025), while a more detailed study of spectral lines can reveal more about the local environment, including multiphase outflows and potential jet-gas interactions as discussed in Section 1.2.3. It is for this reason that the remainder of this chapter turns to a multi-wavelength analysis of our new PS sample, relying again solely on the data from large, public surveys. What results is a richly-described population of PS AGN which are distributed across redshift space out to at least $z \sim 3$, a natural sample to exploit for future analysis of the cosmological evolution of these young active galaxies.

Statement of contribution

This chapter faithfully reproduces Kerrison et al. (2025), which was accepted for publication in the *Monthly Notices of the Royal Astronomical Society*, Vol. 543, Issue 4, pp. 3895 − 3914 (September 2025), with minor stylistic adaptations to fit the thesis style. I am the first author, with Elaine Sadler, Vanessa Moss, Elizabeth Mahony, Laura Driessen, Kathryn Ross, Kovi Rose, Dougal Dobie and Tara Murphy as co-authors. The research problem was conceived of by me as an extension to my work in Chapter 2. Some co-authors provided access to additional radio data. Specifically, K. Ross provided early access to GLEAM-X, (Ross et al., 2024), and D. Dobie and T. Murphy provided early access to VAST Pilot Survey data, (Murphy et al., 2013, 2021). L. Driessen and K. Rose assisted with analysing and interpreting radio lightcurves from VAST, and E. Sadler, V. Moss and E. Mahony provided guidance on the multiwavelength analysis of the sample. I wrote the paper and conducted the analysis, with all co-authors reviewing the manuscript before submission.

3.1 RadioSED applied to Stripe 82: Abstract

This paper is the second in a series presenting RadioSED, a Bayesian inference framework for constructing, modelling and classifying radio spectral energy distributions from publicly-available surveys. We focus here on the application of our framework to SDSS Stripe 82. Not only do we recover all eleven previously-published peaked spectrum sources from the literature within this region, but we increase the number of known peaked spectrum sources here by more than an order of magnitude. We investigate the variability properties of our peaked spectrum sample, and find that overall they exhibit a low degree of variability, consistent with previous samples of peaked spectrum active galactic nuclei. The multiwavelength properties of these sources reveal that we have selected a population comprising largely distant ($z \geq 1$), powerful active galaxies. We find that the most compact jets are located preferentially in quasar-type hosts, with galaxy-type hosts home to slightly more extended radio structures. We discuss these findings in the context of current and forthcoming radio surveys.

3.2 Introduction

In Kerrison et al. (2024c) (henceforth Paper I) we presented a new Bayesian inference framework to construct and model radio spectral energy distributions called RadioSED. Here, we put this framework to the test by using it to identify a new sample of peaked spectrum radio active galactic nuclei ('AGNs'), the properties of which are examined using available multiwavelength data.

Young radio AGNs within a few thousand years of triggering are thought to possess radio jets which have not yet expanded out beyond the gas-rich environment of their host, with the resulting strong magnetic fields and dense ambient medium leading to absorption at low radio frequencies. It is this absorption which leads to a peak in their broadband radio spectral energy distribution (SED), and from which they commonly take the name 'Peaked Spectrum' (PS) sources (O'Dea & Saikia, 2021). However, this is not the only reason for a radio AGN to exhibit a peaked spectrum, indeed some PS sources are thought to be not necessarily young, but impeded in their growth by the dense, ambient medium of their host (e.g. Callingham et al., 2015; Keim et al., 2019). Until recently, the largest, uniform sample of these PS was produced by Callingham et al. (2017). This sample contained sources peaking between 72–700 MHz, and provided unprecedented spectral coverage below the spectral turnover, yet owing to the relatively high flux density limit, the sample had a low density of only 0.06 sources/deg^2. Table 3.1 compares that work to other recent samples from the literature, as well as to our new sample presented here. We note that of these samples, the first to follow Callingham et al. (2017) was the Ross et al. (2021) study of low-frequency variability which has a lower source density again owing to the requirement for multiple GLEAM epochs. This was followed closely by the work of Slob et al. (2022) using the LOw Frequency ARray (LOFAR: van Haarlem et al. 2013) to identify a comparative sample of PS sources in the northern equatorial sky. This was recently superseded by the work of Ballieux et al. (2024), who identified several thousand PS sources in the northern sky by combining survey data from LOFAR, and the Very Large Array (VLA) spanning 60–160 MHz (LOFAR – the LOFAR Two-metre Sky Survey) and 1.4/3 GHz (VLA – the VLA Sky Survey and NRAO VLA Sky Survey) in separate subsamples. However, a comprehensive sample of PS sources, with peaks spanning a broader frequency range and extending to higher frequencies, is crucial for any statistical study of radio AGN evolution, since the linear size of a PS source (a proxy for its age) is expected to be inversely proportional to the restframe frequency of its spectral peak (Snellen et al., 2000; Vries et al., 2009; Jeyakumar, 2016).

Table 3.1: Parameters for several of the most recent PS samples from the literature, compared to those presented in this work.

Selection Frequency (MHz)	Sample Size	Survey Area (deg^2)	Source Density (deg^{-2})	Reference
72 – 230	1,483	24,831	0.06	Callingham et al. (2017)
72 – 230	123	8,000	0.015	Ross et al. (2021)
42 – 1400	373	740	0.50	Slob et al. (2022)
120 – 3000	8,032	5,635	1.43	Ballieux et al. (2024)
42 – 1400	506	740	0.68	Ballieux et al. (2024)
888	359	300	1.20	This work

This paper presents the first such sample of PS sources selected initially using a single, narrow frequency band, yet with spectral peaks spanning 70 MHz – 20 GHz in the observers' frame. This use of a single selection frequency, along with the application of our Bayesian modelling framework RadioSED, and our focus on a well-defined region of sky, means that our sample completeness can be very well constrained. In fact, this sample of 359 sources is at least 90 per cent complete down to 200 mJy across the full range of observed peak frequencies captured. We focus our search on the 300 deg^2 Sloan Digital Sky Survey (SDSS) Stripe 82 field (Abazajian et al., 2009), a region rich with multiwavelength coverage from radio frequencies through to TeV γ-rays. Despite this abundance of data, only eleven PS sources have been identified in the field to date, as shown in Figure 3.1. It is therefore a useful test of RadioSED to apply it to this well-studied field, and the resulting sources in turn form a clean sample with which to probe the properties of PS radio galaxies. Section 3.3 provides further details as to the composition of our sample, Section 3.4 presents the radio properties of the sample, including measures of source structure and variability, in Section 3.5 we consider the optical and IR properties of the hosts of our PS sources, and finally in Section 3.6 we return again to the radio properties of our sources in the context of galaxy evolution. Conclusions and a summary of our findings are presented in Section 3.7.

Throughout this work, we adopt a flat, Λ cold dark matter (ΛCDM) cosmology in line with values from Planck Collaboration XIII (2016); $\Omega_\mathrm{m} = 0.308$, $\Omega_\Lambda = 1 - 0.308$, and $H_0 = 67.8 \mathrm{km\,s^{-1}Mpc^{-1}}$.

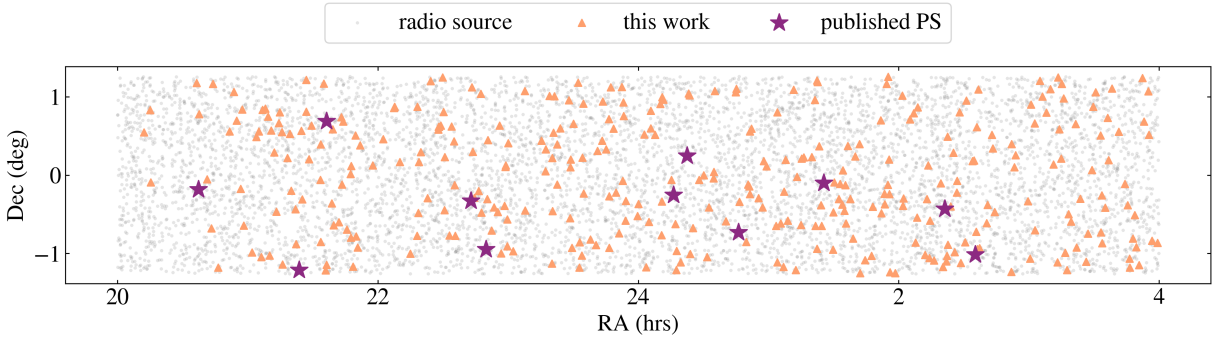


Figure 3.1: A sky plot of the Stripe 82 field, showing the distribution of our sample of sources with peaked SEDs (orange triangles) compared to all PS sources previously published in the literature (purple stars). All RACS-low sources in the field are overplotted as light grey points for reference.

3.3 The sample: peaked spectrum radio sources in stripe 82

The initial construction of our sample is intentionally simple; we make cuts in position and flux density on a parent sample of radio sources, and then perform radio spectral fitting. Using the 888 MHz Rapid ASKAP Continuum Survey (RACS-low; Hale et al., 2021) catalogue as our reference, we select out only those sources within Stripe 82 ($300^\circ < \alpha < 60^\circ$, $-1.26^\circ < \delta < 1.26^\circ$), and which have an integrated flux density $S_{888\,\mathrm{MHz}} > 10$ mJy. This flux density cut ensures our parent sample is at least 95 per cent complete (see Section 6 of Hale et al. 2021), whilst simultaneously removing the many star-forming galaxies which dominate the microJansky radio sky, and which might otherwise contaminate our final AGN sample (e.g. Padovani, 2016). This leaves us with 7 251 radio sources in Stripe 82, all of which were run through RadioSED to construct and then fit their broadband radio spectral energy distributions. The distribution of peaked sources from amongst this sample is shown Figure 3.1, where we have overplotted the previously published PS sources as purple stars, to illustrate the expansion in sample size achieved with this work.

3.3.1 Radio spectral fitting

To identify the peaked spectrum sources from amongst this parent catalogue, we first fit them using RadioSED with the default selection of radio surveys as outlined in Paper I, including the GaLactic and Extragalactic All-sky Murchison Widefield Array (GLEAM) survey at the lowest frequencies (Hurley-Walker et al., 2017; Tingay et al., 2013; Wayth et al., 2018). However, since the publication of Paper I, the GLEAM eXtended

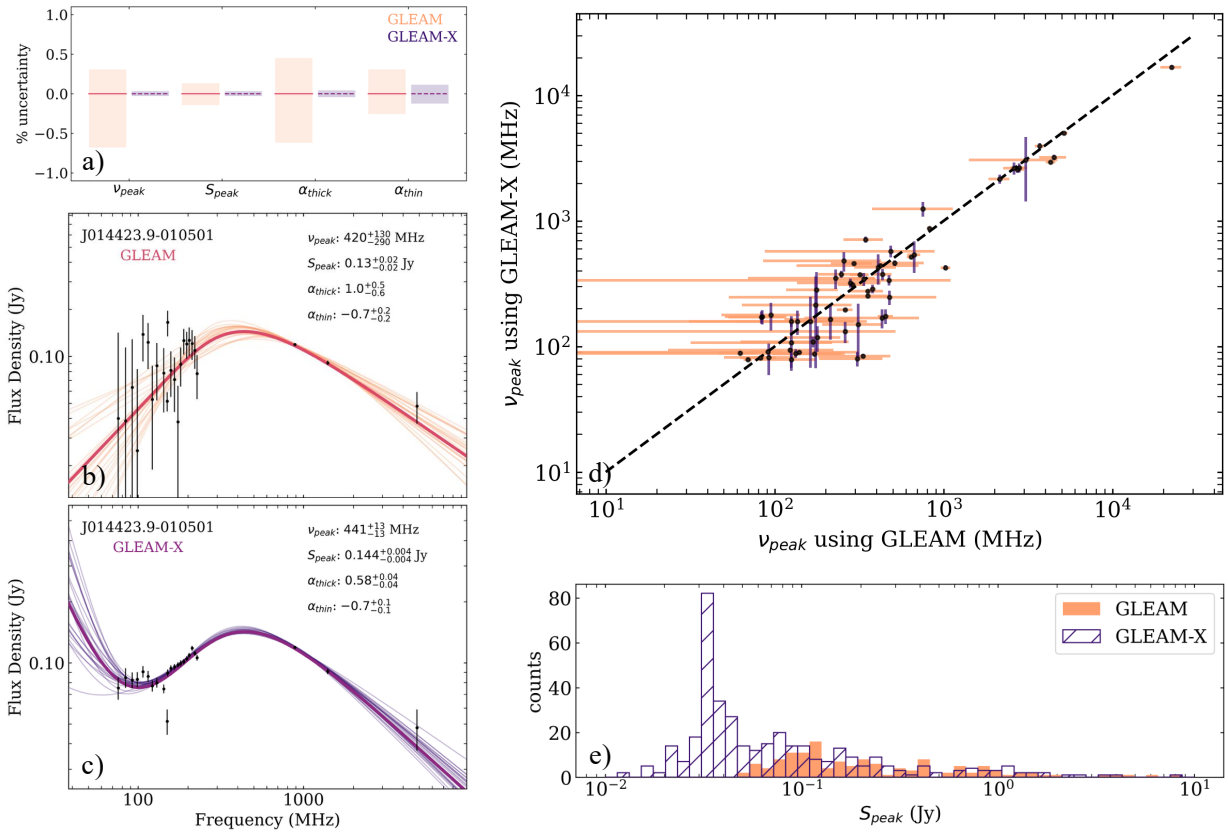


Figure 3.2: A summary of the accuracy of RadioSED fits when using GLEAM data (orange) as opposed to data from the second release of GLEAM-X (purple). Panel a) shows the magnitude of the errors on best fit parameters for a typical peaked spectrum source in our sample. The solid, orange lines are set to zero and represent the mean parameter values derived using GLEAM data, the orange shaded regions indicating the fractional uncertainty on each of these. The dotted purple lines indicate the mean parameter values derived using GLEAM-X data relative to the GLEAM-derived parameters, with purple shading indicating the (significantly smaller) fractional uncertainties on these. Panels b) and c) show the best fit model for this same source when using GLEAM data (panel b), orange) as opposed to GLEAM-X (panel c), purple). The values of the fit parameters are inset in these panels as text. Panel d) shows the peak frequency and associated uncertainty derived from the best fitting RadioSED model. This is for all sources in Stripe 82 classified as peaked spectrum using both GLEAM and GLEAM-X data, with the black dotted line indicating a 1-1 relationship. Panel e) shows the number of PS sources identified using each catalogue as a function of their peak flux density ($S_{\rm peak}$)

.

survey (GLEAM-X) DR II has been released which covers Stripe 82 eastwards of $\alpha > 310°$, and pushes down the 95 per cent completeness level to 16 mJy, over an order of magnitude fainter than the original GLEAM catalogue (Hurley-Walker et al., 2022a; Ross et al., 2024). Accordingly, we re-ran our SED fitting using this new, low frequency catalogue with a statistically-selected match radius of 12 arcseconds, determined using the method outlined in Paper I, to investigate any potential enhancements to the sample. As expected, this led to a significant improvement in both the size of the peaked spectrum sample, and in the precision of individual, peaked spectrum fits. A summary of these improvements is shown graphically in Figure 3.2.

More specifically, incorporating GLEAM-X data reduced the uncertainty in fit parameters from 34 per cent to 13 per cent on average across all peaked spectrum sources common to both fitting runs. In many sources for which the optically thin slope is solely constrained by either GLEAM or GLEAM-X data, this reduction is even more significant, as in the case of J014423.9–010501 shown in panels a)–c) of Figure 3.2, where the more sensitive GLEAM-X data reveals a low-frequency flattening, possibly indicative of an older epoch of activity. The reduction in uncertainties across the broader sample is shown for model-derived peak frequencies ($\nu_{\rm peak}$) in panel d), where it is shown to be consistent across the two runs to within the larger uncertainties from the

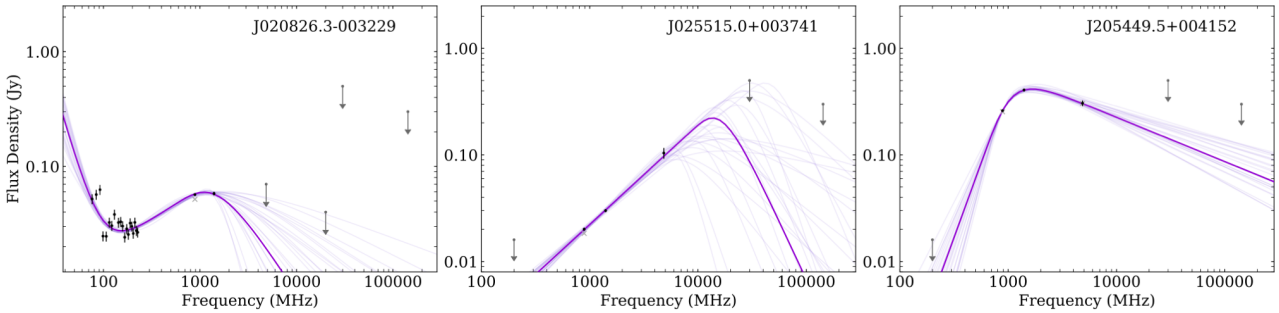


Figure 3.3: Examples of sources for which the peak can be identified only with the addition of censored datapoints from the surveys outlined in Table 3.2. Upper limits are indicated by arrows. The best fit model is shown as a purple, solid line, with lighter lines indicating draws from the posterior.

GLEAM-derived models. Our ability to constrain the peak frequency is of particular importance, as this has been shown to correlate directly with the linear size, and hence, the age of a source (O'Dea, 1998). In addition to the precision of individual fits, the size of our peaked spectrum sample increased fourfold, from 92 to 346 with the use of GLEAM-X data, extending down to sources with a peak flux 11.7 ± 1 mJy, as shown in panel e). This is an exciting hint of what is to come with the larger, peaked spectrum sample of Ross et al. (in preparation) covering the full, GLEAM-X DRII region. And it is for these reasons that we proceed with the sample derived using the GLEAM-X data for the remainder of this paper.

Since we are focused on a well-defined region of sky, we were also able to incorporate survey limits into our fitting as censored data points within individual SEDs. These additional constraints are outlined in Table 3.2, where the limiting flux densities are taken as the 95 per cent completeness limits drawn from the respective papers. The implementation of censored fitting itself was described in Paper I. These upper limits led to the inclusion of thirteen additional PS sources in our final sample, and improved constraints on the spectral properties of several more sources in which a peak had already been identified. Examples of sources which benefit from the use of censored data are shown in Figure 3.3. Overall, our sample therefore comprises 359 PS sources within the Stripe 82 region. The fit parameters of these PS sources are given in Appendix D, and the full catalogue is available in the online version of this paper.

Table 3.2: The survey limits incorporated into RadioSED for the identification of the PS sample in this paper, in addition to those already described in Paper I. The chosen flux density is the 95 per cent completeness limit of the relevant survey, and if it applies only to a sub-section of Stripe 82, this constraint is given in the table.

Survey	Frequency (GHz)	Limit (Jy)	Region of Stripe 82	Reference
GLEAM-X	0.2	0.016	$\alpha > 310°$	Ross et al. (2024)
PMN	4.85	0.07	full	Griffith et al. (1993)
AT20G	20	0.04	$\alpha < 200°$, $\delta < 0°$	Massardi et al. (2011)
PCCS	30	0.5	full	Ade et al. (2014)
PCCS	143	0.3	full	Ade et al. (2014)

For a source to be classified as PS, it must have observations either side of the predicted peak, and it must have fractional error $\frac{\Delta\alpha_{\rm thick}}{\alpha_{\rm thick}} < 1$ to avoid ambiguity in spectral shape (where $\alpha_{\rm thick}$ is the spectral index below the broadband peak). Furthermore, as discussed in Paper I, we have separated out those sources which have spectral indices shallower than required for the canonical definition of a PS source (i.e. $-0.5 < \alpha_{\rm thin} < 0$ or $0 < \alpha_{\rm thick} < 0.5$) into a "soft peaked spectrum" (SPS) category within the catalogue. This is for purely historical reasons, to allow for easy comparison with older PS samples from the literature. For our analysis here though, the PS and SPS sources will be collectively analysed as, and assumed to be, peaked spectrum sources.

Overall, from the 7 251 radio sources within our selected field, 4 458 could not be fit due to an insufficient number of flux density measurements or meaningful constraints provided by upper limits. This population of unfitted sources has a median flux density of 18.04 mJy, and 90 per cent have a flux below 70 mJy, so they

comprise mainly faint sources in the field.

From amongst the sources which could be reliably fit, PS sources make up approximately 13 per cent, which falls comfortably within the range of population fractions calculated by Ballieux et al. (2024) at one end (3.9 per cent for sources peaking at MHz frequencies) and O'Dea (1998) at the other (20 per cent typically peaking at GHz frequencies). The SED plots for the full sample of 359 PS sources can be found in the supplementary material online.

3.3.2 Completeness: sample limits

Although our initial selection criterion was to impose a single flux density cut (10 mJy at 888 MHz), the completeness of our sample will be a function of both peak frequency and peak flux density, since a source must be detected in at least three unique frequency bands to be fit. Within the region of GLEAM-X DR II (eastwards of $\alpha = 310°$), we estimate the completeness of our PS sample using a simple Monte Carlo test. We generate just over 23 000 synthetic spectra with spectral indices uniformly sampling the interval $0.5 < |\alpha| < 2.1$, peak frequencies logarithmically sampling the interval between 50 MHz –30 GHz, and peak flux densities spanning 5 mJy – 2.8 Jy, also logarithmically spaced. For each combination of spectral parameters, we generate flux density measurements at frequencies corresponding to the surveys used in fitting, and keep only those above the 95 per cent completeness limit in each survey, otherwise replacing them with upper limits where these were used in constructing our sample (we refer the reader to Paper I, Table 1 for further information on these surveys). The resulting synthetic spectra are fit with RadioSED, and the criteria from Section 3.3.1 are applied to determine which would fall into the 'PS' category as defined here. The results of this test are summarised in Figure 3.4, where the colour gradient indicates the expected completeness from these simulations.

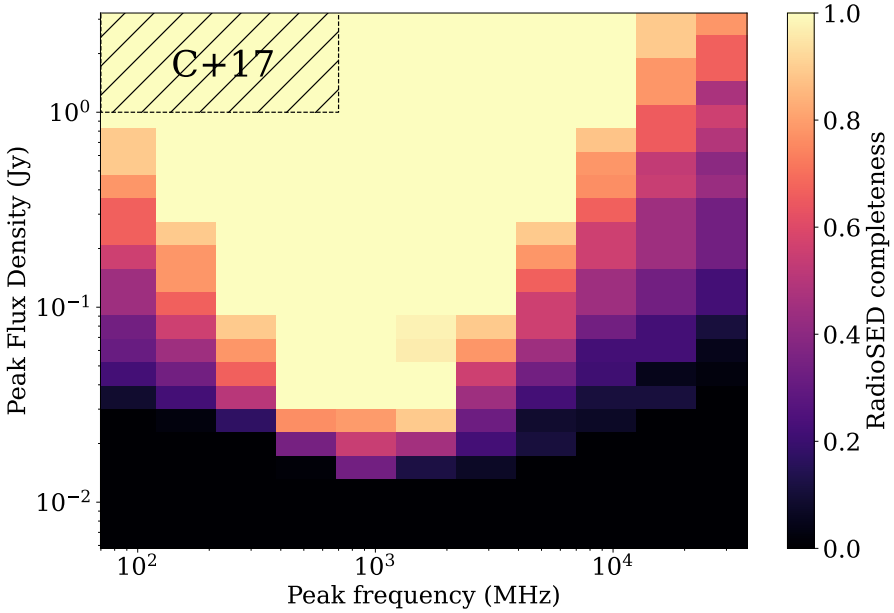


Figure 3.4: The completeness of our PS sample as a function of peak flux density and peak frequency, divided into logarithmically-spaced bins. Completeness calculations were performed by feeding synthetic spectra to RadioSED, and using the set of parameters specified in Section 3.3.1. The hatched region indicates the area of parameter space within which the all-sky sample of Callingham et al. (2017) was estimated to be complete.

Evidently, our sample completeness varies as a function of both peak frequency and peak flux density, and even at our selection frequency of 888 MHz we do not reach 100 per cent completeness at 10 mJy. This is undoubtedly due to the higher flux density limits of the other surveys required for fitting, which cannot adequately sample spectra which peak at or below a few tens of mJy. Nevertheless, we are most complete to sources peaking at or around our 888 MHz selection frequency. These simulations also suggest that our sample is complete across a large fraction of the peak parameter space. For comparison, we overplot the hatched region in Figure 3.4 to indicate the completeness of the Callingham et al. (2017) sample. To our knowledge, this is

the only other PS sample to date for which completeness statistics have been calculated, and it is clear that the present sample promises a vast improvement in the diversity of PS sources we can detect. The inclusion of additional, high-frequency surveys, along with the more sensitive GLEAM-X data, and censored fitting with upper limits all allow us to identify PS sources that are fainter and peak at higher frequencies than the Callingham et al. (2017) sample. The former of these improvements may lead to the identification of PS sources that are either intrinsically less powerful or more distant, while the latter suggests we are sensitive to a population in which the jets are even more compact. This sample may be therefore thought of as a narrow, deep companion to the wide-area and shallow sample in that work, and a precursor to the large-area samples that will be possible with future GLEAM-X releases, and eventually, with data from SKA–low and mid.

3.3.3 Completeness: recovering published sources

Another measure of the completeness of our sample is our ability to recover known peaked spectrum sources. Guided by the references in Callingham et al. (2017) and Ballieux et al. (2024), we searched a number of papers for peaked spectrum sources within the Stripe 82 field boundaries. These include: Stanghellini et al. (1997); O'Dea (1998); Dallacasa et al. (2000); Edwards & Tingay (2004); Tinti et al. (2005); Torniainen et al. (2007); Labiano et al. (2007); Randall et al. (2011); Callingham et al. (2017) and Slob et al. (2022). After removing duplicates and those sources classified as "Compact Steep Spectrum" (i.e. without any observed spectral turnover), we were left with the eleven sources outlined in Table 3.3.

We recovered peaked SEDs for all eleven previously known PS sources within our sample. The fit parameters output by RadioSED are given in Table 3.3, to aid comparing our results to those from the literature. The SEDs for these sources are shown in Figure 3.5.

Of these eleven sources, J023516.7–010050 would canonically be classified as flat spectrum since it has $\alpha_{\text{thin}} > -0.5$, while J001611.0–001510 is on the borderline of a flat spectrum classification. However, as can be seen from Figure 3.5 these are well fit by PS models, and our spectral indices agree broadly with those derived by Callingham et al. (2017). Accordingly J023516.7–010050 is classified as "soft peaked spectrum" for compatibility with previous definitions. The other nine sources show clear peaked spectrum SEDs, although J001611.0–001510, J012528.8–000555, and 213638.5+004154 appear in the 5th edition of the Roma-BZCAT blazar catalogue as "Flat Spectrum Radio Quasars" (Massaro et al., 2015), calling into question whether they are true Peaked Spectrum sources. We consider this question further in Section 3.3.5, but ultimately we keep these sources in our PS sample, as they do not possess a flat spectrum, a key component of their classification in Roma-BZCAT.

While the spectral parameters derived by RadioSED are not always in close agreement with those from the literature, they appear to appropriately characterise the SED shapes in Figure 3.5. Some discrepancy between parameters from the literature and RadioSED is evidently due to the broader frequency coverage of observations used in this work compared to previous approaches, but a deeper analysis beyond this is challenging, as many of the fits from older works use data that was not published in tabular form. However, we demonstrated in Paper I that RadioSED can accurately recover injected parameters in synthetic fits, so any discrepancy is real, and must be rooted in the subtleties either of the data used, the fitting methods, or both.

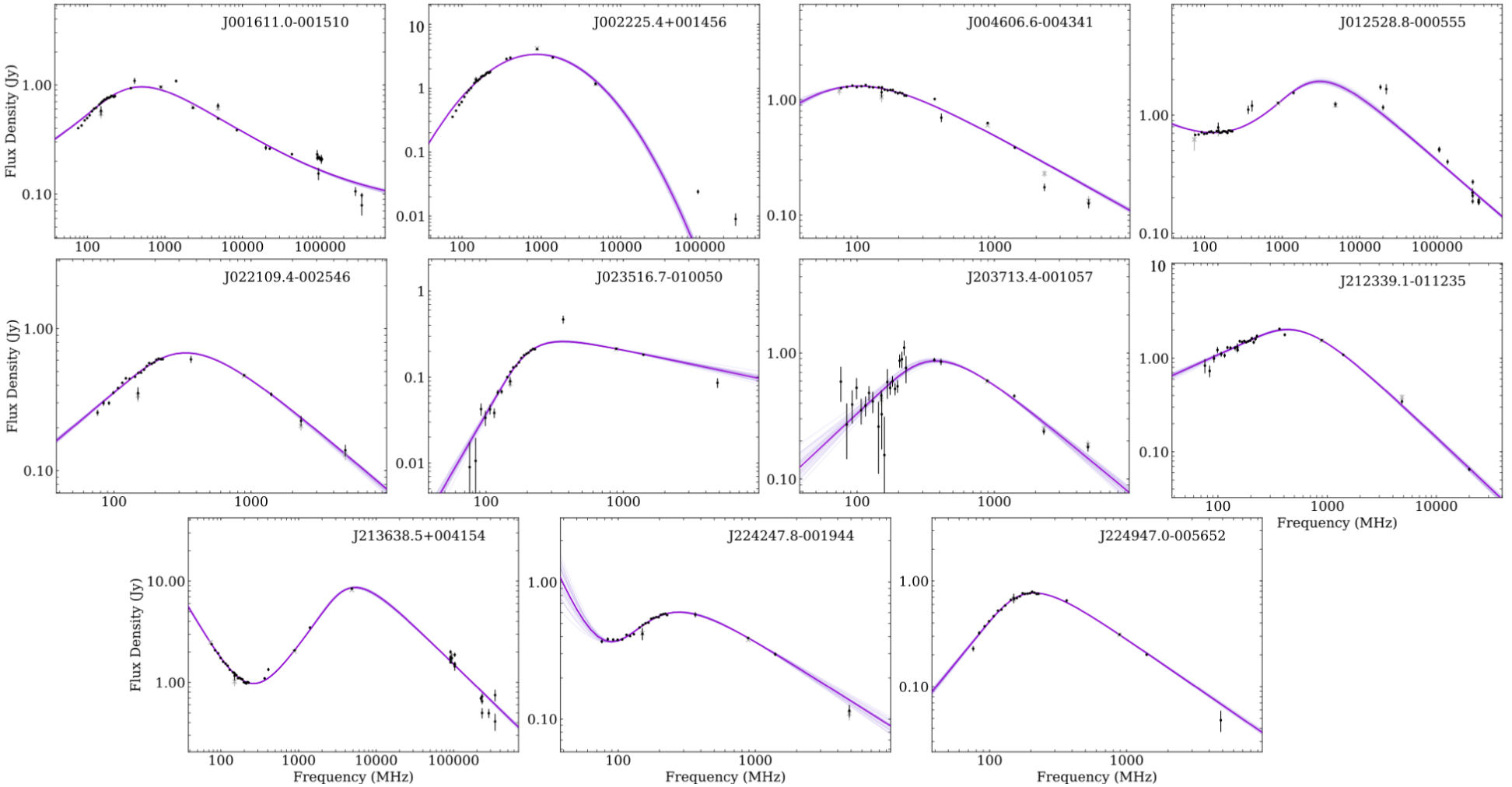


Figure 3.5: SEDs of all eleven known peaked spectrum sources from the Stripe 82 field obtained using RadioSED. The dark line in each plot shows the best fit of the most probable model from RadioSED, while the array of lighter lines represents 25 draws from the posterior of this model.

Table 3.3: Peaked Spectrum sources within the Stripe 82 field that were previously identified in the literature.* Source names are from the RACS-low catalogue, and literature values for model parameters (peak frequency: $\nu_{\rm p,\,lit}$, peak flux: $S_{\rm p,\,lit}$, and optically thick and thin spectral indices: $\alpha_{\rm thick,\,lit}$, $\alpha_{\rm thin,\,lit}$) are drawn from the citations listed, with uncertainties quoted where available. All other model parameters are derived from RadioSED. Citations are not exhaustive for each source, but represent those works in which the source was first identified as peaked spectrum, and from which the literature parameters were derived. The citations in Column 11 are as follows: S85: Spoelstra et al. (1985), O98: O'Dea (1998), E04: Edwards & Tingay (2004), T07: Torniainen et al. (2007), and C17: Callingham et al. (2017). Where a model parameter is unknown or not applicable to a particular source, that column is marked with a dash.

Name	$\nu_{\rm p,\,lit}$ (GHz)	$S_{\rm p,\,lit}$ (Jy)	$\alpha_{\rm thick,\,lit}$	$\alpha_{\rm thin,\,lit}$	$\nu_{\rm p}$ (GHz)	$S_{\rm p}$ (Jy)	$\alpha_{\rm thick}$	$\alpha_{\rm thin}$	$\alpha_{\rm retrig}$	Refs.
J001611.0–001510	>0.843	–	0.42 ± 0.08	–	$0.526^{+0.015}_{-0.015}$	$0.885^{+0.005}_{-0.005}$	$0.67^{+0.01}_{-0.01}$	$-0.52^{+0.01}_{-0.01}$	$-0.002^{+0.006}_{-0.006}$	C17
J002225.4+001456	0.76	–	0.46	–	$0.873^{+0.012}_{-0.012}$	$3.39^{+0.03}_{-0.03}$	$1.59^{+0.006}_{-0.007}$	$-3.77^{+0.05}_{-0.05}$	–	O98; T07
J004606.6–004341	0.09 ± 0.04	1.41 ± 0.19	–	-0.59 ± 0.09	$0.117^{+0.006}_{-0.005}$	$1.312^{+0.006}_{-0.009}$	$0.69^{+0.07}_{-0.06}$	$-0.66^{+0.01}_{-0.01}$	–	C17
J012528.8–000555	>0.843	–	0.46 ± 0.09	–	$3.31^{+0.17}_{-0.17}$	$1.79^{+0.04}_{-0.05}$	$0.77^{+0.03}_{-0.03}$	$-0.59^{+0.01}_{-0.01}$	$-0.38^{+0.03}_{-0.03}$	C17
J022109.4–002546	0.35 ± 0.07	0.72 ± 0.08	0.90 ± 0.21	-0.91 ± 0.32	$0.385^{+0.012}_{-0.011}$	$0.667^{+0.003}_{-0.003}$	$0.81^{+0.02}_{-0.02}$	$-0.82^{+0.02}_{-0.02}$	–	C17
J023516.7–010050	0.31 ± 0.08	0.27 ± 0.11	3.62 ± 1.85	-0.34 ± 0.34	$0.223^{+0.007}_{-0.006}$	$0.215^{+0.007}_{-0.007}$	$2.76^{+0.13}_{-0.11}$	$-0.33^{+0.03}_{-0.03}$	–	C17
J203713.4–001057	0.40 ± 0.08	0.86 ± 0.05	0.83 ± 0.26	-0.96 ± 0.24	$0.404^{+0.038}_{-0.033}$	$0.863^{+0.025}_{-0.027}$	$1.02^{+0.14}_{-0.13}$	$-0.89^{+0.04}_{-0.04}$	–	C17
J212339.1–011235	0.34 ± 0.07	1.75 ± 0.05	0.71 ± 0.14	-0.67 ± 0.14	$0.648^{+0.026}_{-0.025}$	$1.85^{+0.04}_{-0.04}$	$0.53^{+0.03}_{-0.03}$	$-1.1^{+0.01}_{-0.02}$	–	S85; T07; C17
J213638.5+004154	5.9	–	1.6	-0.88 ± 0.08	$5.28^{+0.18}_{-0.18}$	$8.65^{+0.11}_{-0.1}$	$1.08^{+0.02}_{-0.02}$	$-0.75^{+0.01}_{-0.01}$	$-1.31^{+0.02}_{-0.02}$	O98; E04; C17
J224247.8–001944	0.44 ± 0.15	0.77 ± 0.18	0.52 ± 0.18	-1.56 ± 0.90	$0.242^{+0.014}_{-0.012}$	$0.583^{+0.012}_{-0.015}$	$1.4^{+0.14}_{-0.13}$	$-0.63^{+0.03}_{-0.04}$	$-2.3^{+0.4}_{-0.4}$	C17
J224947.0–005652	0.26 ± 0.09	0.79 ± 0.05	0.92 ± 0.29	-1.18 ± 0.25	$0.196^{+0.003}_{-0.003}$	$0.77^{+0.002}_{-0.002}$	$1.59^{+0.04}_{-0.04}$	$-0.89^{+0.01}_{-0.01}$	–	C17

*During the preparation of this manuscript, Sun et al. (2025) was published presenting a new analysis of PS sources, primarily using GLEAM data. These PS sources are not included in this analysis.

3.3.4 Reliability

The reliability of a new sample of PS radio AGNs is difficult to determine a priori. In Section 4.2 of Paper I, we reasoned that we might expect a misclassification rate of approximately 4 per cent in any sample of PS sources found using RadioSED, based on simulations. However, this assumed a certain degree of variability amongst flat spectrum radio sources, and that they comprise a certain fraction of the radio population — neither of which is particularly well constrained in the literature. We might therefore take this 4 per cent as a fiducial contamination rate in our sample, which would suggest some 14 of our sources are not true PS AGNs. However, to say more requires a deeper, multiwavelength consideration of the sources in question, which we will return to again in the context of variability in Section 3.4.3.

3.3.5 Blazar contamination

At the highest level, a blazar is an AGN in which the jet axis is aligned along our line of sight such that its emissions are relativistically beamed. This produces apparent superluminal motion, and causes the source flux density to vary rapidly across the full electromagnetic spectrum. In a gigahertz radio image with arcsecond resolution, a blazar will therefore appear as a compact source. Although a PS radio AGN could of course have its jets aligned in this way, the likelihood of observing such an object is probably very low, and in any case its emissions would be dominated by the blazar-like characteristics of its aligned jet. For this reason, those working on PS sources tend to consider blazars a separate class of object (e.g. Orienti et al., 2007; O'Dea & Saikia, 2021). By contrast, the original Roma-BZCAT catalogue of blazars included PS sources peaking at or above 1 GHz (Massaro et al., 2009, Sec. 3), and these have propagated through into the latest version of that work. More recently, Behiri et al. (2025) showed that a number of Fermi blazars in fact have broadband radio spectra with complex shapes including peaked and low-frequency upturns indicative of episodic activity. Furthermore, many works identifying PS sources (including this one) have necessarily had to combine observations from different epochs in order to obtain the broadband spectral coverage necessary for classification. The dynamic variability of blazars combined with this heterogeneous data makes any such sample somewhat susceptible to blazar contamination, though such contamination should be naturally reduced in the coming years with the publication of more contemporaneous, widefield, and broadband surveys from the SKA and its precursors and pathfinders. Nevertheless, at present this makes separating out blazars from candidate PS radio AGNs a difficult task using blazar catalogues alone.

We have already seen in Section 3.3.3 that three of the eleven previously-published PS sources in our sample are found within the Roma-BZCAT catalogue of blazars. However, in each case they are classed as "Flat Spectrum Radio Quasars" (FSRQs), a classification which is typically based on only two flux density measurements, plus an optical spectrum indicative of a quasar. A crossmatch between the 5th edition of the Roma-BZCAT and the PS sources in Stripe 82 reveals a further 13 matches, but again 10 of these are classified as FSRQs, with only the final three exhibiting featureless optical spectra of a BL Lac type, clearly indicative of a blazar (Massaro et al., 2015). These three are excluded from further analysis as PS sources. Since it is known that PS sources can be hosted by both galaxies and quasars, we do not discount the 10 FSRQ sources from our sample on this basis alone, though we do note that a visual inspection of the SEDs of J001611.0–001510, J012528.8–000555, and 213638.5+004154 in Figure 3.5 (the previously-published PS sources also in the Roma-BZCAT) reveals some scatter around the best fit model in each case, which may be indicative of radio variability consistent with episodic blazar activity.

Finally, since the Roma-BZCAT relies on both optical spectra and radio flux density for classification, it will not include those sources in our sample lacking optical spectra. This may mean there are additional blazar candidates contaminating our sample which cannot be identified by this method alone. We leave further consideration of potential blazar contamination to Section 3.4.3, where we consider the other key metric by which blazars are typically identified; variability.

3.4 Radio properties

PS sources are expected to be compact, with linear sizes $\lesssim 1$ kpc, corresponding to angular scales $\theta \sim 1.2$ arcsec at $z = 1$, and to have a radio spectrum with a low degree of variability. Here, we study the sub-arcsecond

compactness of our sample using relatively novel compactness measures at 162 MHz and 20 GHz. We also consider the variability of our sample at $\nu \sim 800$ MHz and consider the possibility of both intrinsic variation, and flux modulation by an external scattering screen. The radio frequency data used in the following subsections can be found in Appendix E, with the full version available online.

3.4.1 Source Structure from continuum surveys

To obtain first-order constraints on the source structure we measure compactness based on two of the surveys used in SED fitting. Following the procedure used in survey description papers such as Bondi et al. (2008); Shimwell et al. (2019), and Hale et al. (2021), we determine which of our sources are unresolved up to the resolution limit of both RACS-low (15 arcsec at 888 MHz) and GLEAM-X ($\sim$ 45 arcsec at 200 MHz), two of the surveys with most complete coverage for our sample. We construct an envelope within which unresolved sources are expected to reside when the ratio of their total to peak fluxes ($F_{\rm tot}/F_{\rm pk}$) is plotted against their signal to noise (SNR). In the case of RACS-low, this envelope was already derived in Hale et al. (2021) Section 5.2.1, and we use those same parameters here. In the case of GLEAM-X DRII, we follow the method from that paper to derive an envelope of the form:

$$\frac{F_{\rm tot}}{F_{\rm pk}}_{\pm} = 0.995 \pm 0.281 \times {\rm SNR}^{-0.313} \tag{3.1}$$

These envelopes are shown in the two panels of Figure 3.6. Although it may seem at first glance that several of our sample fall outside of this envelope, its shape only takes into account the global scatter about an expected ratio, $\frac{F_{\rm tot}}{F_{\rm pk}}$ for an unresolved source, and not the uncertainty on the individual flux density measurements. In reality, the uncertainty on both $F_{\rm tot}$ and $F_{\rm pk}$ in our sources means that, to first order, all could reasonably be considered unresolved in RACS-low, and all bar two are unresolved in GLEAM-X. The two which may show extended structure at MHz frequencies (J020137.2+010057 and J023036.5–005122) are indicated in Figure 3.6 by large black crosses, though a visual inspection of their GLEAM-X images does not reveal any obvious structure.

However, if we consider the length scales which are probed at the resolution of these surveys assuming a flat, ΛCDM cosmology, we are only sensitive to structures $D \geq 30$ kpc at a redshift $z = 0.1$ (with RACS-low), increasing to structures as large as $D \geq 40$ kpc at $z = 5.0$ (with GLEAM-X). This far exceeds the typical size of a PS source, which is thought to be between a few hundred parsecs up to 1-2 kpc, where they begin to transition to compact steep spectrum sources, in which the peak falls below the range of frequencies currently observable (O'Dea & Saikia, 2021). Thus these surveys alone are not enough to properly constrain the structure of our PS sample.

3.4.2 Source structure: sub-kiloparsec scales

To perform further analysis on continuum source structure, Very Long Baseline Interferometry (VLBI) is often used to image milliarcsecond structure thanks to both the high frequency of observations ($\geq$ 5 GHz) and the intercontinental baselines involved ($\geq 10^3$ km). However in this case, VLBI will resolve out much of the structure we are interested in, as even a 50 mas resolution image probes length scales 0.01–0.3 kpc between redshifts $z = 0.1$–5. Furthermore, VLBI catalogues are highly incomplete due to the stringent selection criteria typically applied to obtain telescope time, complicating sample-level inference. Ideally, we would like to be able to probe structures between these two spatial regimes of continuum surveys and VLBI, and fortunately there are still tools that can achieve this for us.

One such tool is interplanetary scintillation (IPS), a phenomenon where background radio sources with sub-arcsecond compact components (like Quasars or PS sources) exhibit variations in flux density over seconds-long timescales due to changes in the intervening solar wind (Clarke, 1964). Jeyakumar et al. (2000) established IPS as a useful analysis technique for PS sources, using measurements at a range of solar elongations and position angles relative to the solar wind vector to measure source structure in conjunction with VLBI on different angular scales. A novel technique for widefield IPS detections was developed by Morgan et al. (2018) for use with the MWA, the same instrument used for the GLEAM-X survey. Although this widefield survey technique does not provide measurements over a range of solar elongations and position angles like the targeted observations in

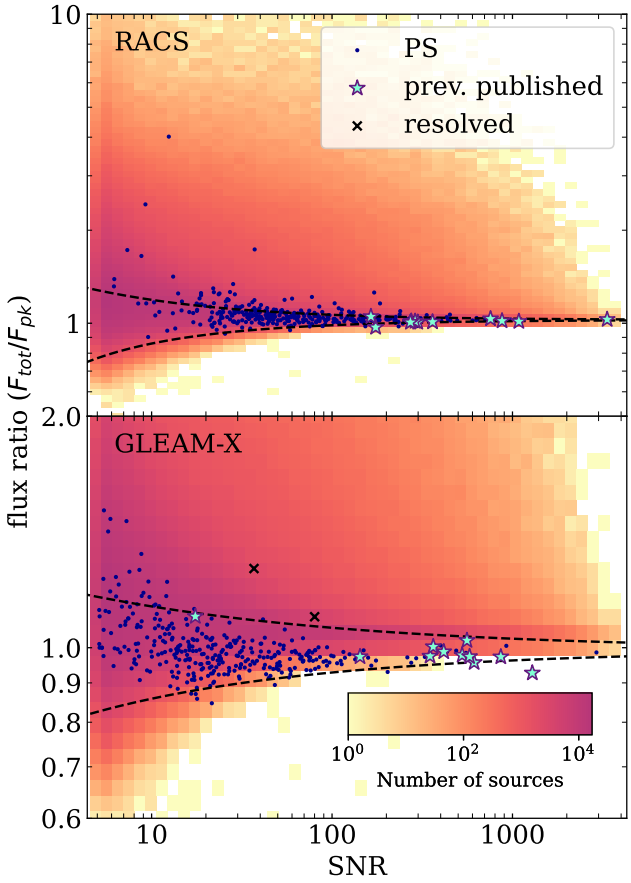


Figure 3.6: Heatmap of the flux ratio $\frac{F_{\rm tot}}{F_{\rm pk}}$ for all sources in RACS-low (top) and GLEAM-X (bottom), overplotted with our PS sources (navy dots) and previously-published PS sources in Stripe 82 (cyan stars). The dotted lines inscribe the envelope of unresolved sources as defined in Hale et al. (2021) Section 5.2.1.

Jeyakumar et al. (2000), it does offer complete coverage of our sample above the sensitivity limit of the survey. Indeed, using data obtained at 162 MHz, the MWA-IPS technique is sensitive to compact components with a ~ 0.3 arcsec angular diameter, and the strength of their scintillation is expressed as a 'Normalised-Scintillation Index' ('NSI'; Chhetri et al., 2018), where a value of 1 indicates all of the flux is contained in the ≤ 0.3 arcsec compact core, and values less than this indicate the fraction of the total MWA flux scintillating. Chhetri et al. (2018) used this technique to identify 632 sources exhibiting some level of IPS in a 5-minute observation spanning 900 deg^2 of sky. They note that a large fraction of their moderate-to-strongly scintillating (NSI > 0.7) sources are PS, and that all 21 sources from the Callingham et al. (2017) sample in their field have an NSI > 0.7. As follow-up to that work, Jaiswal et al. (2022) combined IPS analysis with VLBI measurements in a manner similar to the earlier Jeyakumar et al. (2000) study, while Sadler et al. (2019) identified the hosts of these scintillating sources as distant radio galaxies and quasars, with a median redshift $z \sim 1.5$, and approximately 30 per cent residing at redshift $z > 2$.

How does this compare to our sample? Of the 359 sources we have identified, 74 are included in the latest MWA-IPS catalogue (i.e. they have a 5σ detection in the continuum image (Morgan et al., 2022, and catalogue in prep.). Of these, 40 have secure IPS detections (at least one 5σ detection in the variability image), a further 24 have a marginal IPS detection ($< 5\sigma$ in the variability image), and the final eleven have upper limits on their NSI. A ~ 20 per cent detection rate (in the sense that a "detection" appears in the IPS catalogue - whether as measurement or upper limit) is not surprising given our original selection frequency for this sample was much higher than the 162 MHz MWA observations. Indeed, of the 289 sources not detected with IPS, 286 have flux densities $S_{200\,\rm MHz} \leq 100$ mJy, and two of the three remaining sources (J014423.9–010501 and J204315.0–004037) have "re-triggered" type SEDs with a low frequency upturn indicative of older, extended emission. It is therefore unsurprising that their flux in the MWA band is resolved out at the angular resolution required for IPS. Nevertheless, it is useful to consider how this population compares to the one studied in Chhetri et al. (2018), which was selected purely based on IPS statistics.

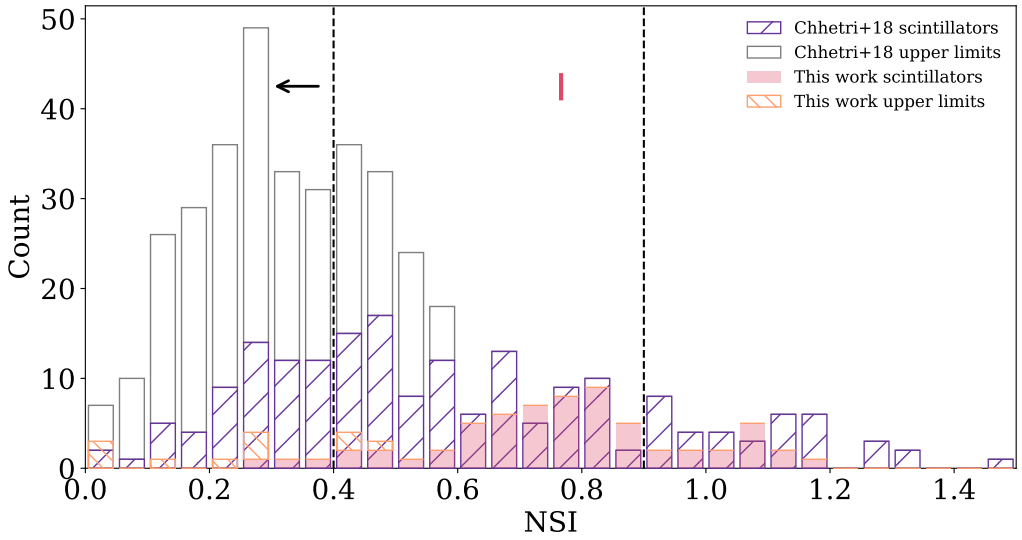


Figure 3.7: Distribution of NSIs for our sample (red bars) compared to the sample of Chhetri et al. (2018) (right-hatched purple bars) including upper limits (orange-hatched for our sample, black outline for the Chhetri et al. (2018) sample). The vertical dashed lines indicate the thresholds in Sadler et al. (2019) for 'moderately scintillating' (0.4 < NSI < 0.9 and 'strongly scintillating' (NSI > 0.9) sources. The red bar indicates the median NSI of our sample, while the black arrow indicates an upper limit on the median of the Chhetri et al. (2018) sample for reference.

In Figure 3.7, we show the distribution of NSI for our sample of PS sources, compared to the sample of Chhetri et al. (2018), including upper limits on NSI in both cases. In our sample the distribution of NSIs is skewed towards 1 with a median of 0.77, while the Chhetri et al. (2018) reference sample contains many more censored datapoints, so that we can only estimate an upper limit of 0.38 on that sample median (we note values above 1 here are due to scatter). This is encouraging, as it suggests our sample tends to have more flux contained in sub-arcsecond structure than the general population of megahertz radio sources. Indeed, almost all of our sources are at least 'moderately scintillating' by the Sadler et al. (2019) criterion (NSI > 0.4), indicating at least 40 per cent of their flux at 162 MHz is contained in a compact component. However, where all 21 PS sources contained in the Chhetri et al. (2018) sample had NSI > 0.7, only two thirds of our sample meet this threshold. This is in part due to our sample containing sources peaking at gigahertz frequencies and higher, which are typically only detected in the 162 MHz IPS measurements if they have a low-frequency upturn indicative of older, extended emission. These findings are summarised in Table 3.4.

At higher frequencies, there is another tool used by Chhetri et al. (2013) to probe comparable sub-arcsecond structure called the visibility ratio. In that work, the ratio of complex visibilities from long and short baselines was used to measure the relative flux contained within 0.5 arcsecond structure at 20 GHz, by making use of the original visibility data from the Australia Telescope Compact Array (ATCA) 20 GHz survey (AT20G; Murphy et al., 2010). Of our sample, 21 sources are contained within that catalogue, and all of these have visibility ratios > 0.9, indicating more than 90 per cent of their flux is contained within a sub-arcsecond component. Since both this visibility ratio and the NSI from IPS probe comparable angular scales with metrics spanning the same range, we have combined their values in Figure 3.8, where the central fill-colour of points indicates their NSI, and the annulus colour is the visibility ratio for the same source.

Table 3.4: The distribution of NSIs for our sample, and the Chhetri et al. (2018) sample for comparison.

IPS	NSI	Number (per cent)	
		This work	Chhetri et al. (2018)
Low	< 0.4	9 (12%)	59 (31%)
Moderate	0.4 - 0.7	21 (28%)	70 (36%)
High	> 0.7	45 (60%)	64 (33%)

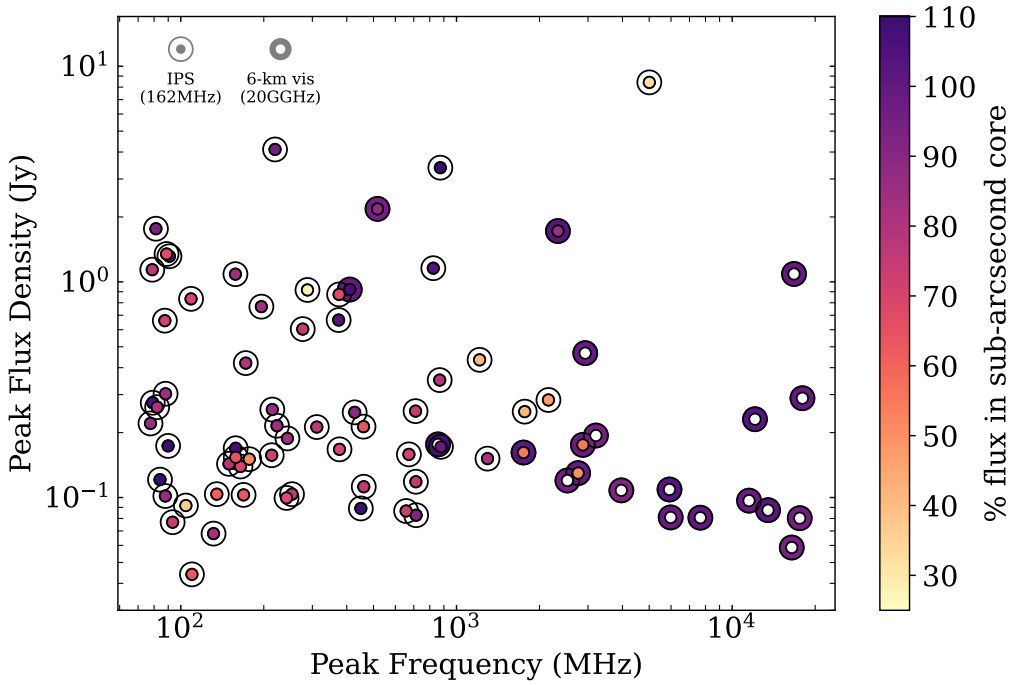


Figure 3.8: Sources in our full PS sample which have a sub-arcsecond compactness measure from either IPS at 162 MHz (central filled circle) or an AT20G visibility ratio at 20 GHz (filled annulus). Sources detected in the 20 GHz survey tend to have their flux entirely localised to a ≤ 0.5 arcsec region at that frequency, while sources detected in IPS have anywhere from 30–100 per cent of their flux contained within a single, compact component. If a source was detected in both the MWA-IPS survey and the AT20G survey, this is indicated by a datapoint with both a shaded inner circle and annulus.

Evidently, there are selection effects at play here. Even a strong, Jy-level source peaking at 20 GHz in the observers' frame will almost certainly fall well below the sensitivity threshold of the IPS survey at 162 MHz, and vice versa. Indeed, there are only 3 sources which are detected and measurably compact at both frequencies. Therefore by combining data from each we can more uniformly study the structure of our full PS sample. Even so, the 162 MHz IPS measurement is more likely to probe the absorbed component below the spectral turnover in our sample, and the visibility ratio will probe the converse, since the subset of our sample detected in these surveys peaks between $\sim$ 100 MHz–20 GHz. Any extension in a PS source should be most visible below the spectral peak, where the absorbed emission traces the very edges of the jet structure. It is therefore not surprising that the NSI of many objects peaking $>$ 200 MHz is somewhat lower than 1. Although a large fraction of their flux still comes from a compact core, some non-negligible amount may be traced to more extended structure. By contrast, the visibility ratio is probing the optically thin emission above the peak, which, based on the data here, is coming almost entirely from a sub-arcsecond component in those sources detected at 20 GHz.

To put these values back into the context of source size, in our ΛCDM cosmology, both the NSI and visibility ratio probe scales spanning 0.3–2.5 kpc at redshifts $0.1 < z < 5$, which neatly encapsulates the range of sizes PS sources are thought to span. What Figure 3.8 demonstrates then, is that where our sources are detected with the 20 GHz visibility ratio, they do indeed appear to have a linear extent less than $\sim$ 2 kpc. Those detected by the 162 MHz IPS technique may have some extended structure at lower frequencies, possibly indicative of older jet activity, but nevertheless a significant fraction of their flux still comes from a compact, parsec to kiloparsec-scale component. Further expansion upon the IPS work of Morgan et al. (2018); Chhetri et al. (2018) and the visibility ratios of Chhetri et al. (2013), increasing the sensitivity and sky coverage of both samples would offer an invaluable contribution to the study of sub-arcsecond source structure, allowing us to probe structures too small for typical continuum imaging, and yet too large for usual VLBI baselines.

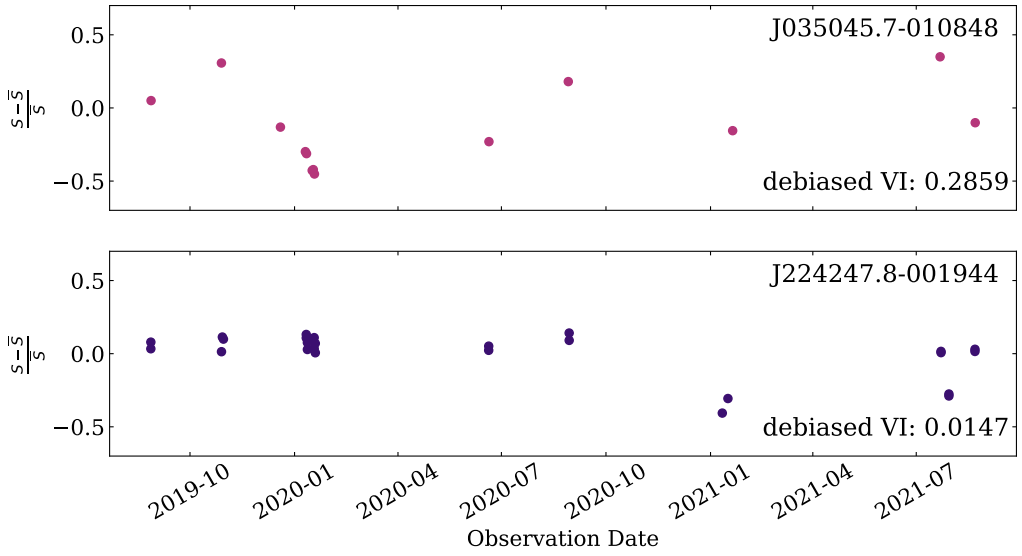


Figure 3.9: Two example radio lightcurves drawn from amongst our PS sample, one with a debiased variability index indicative of variability (top panel), and one that is essentially non-varying to within the sensitivity of the VAST lightcurve (bottom panel).

3.4.3 Radio variability

Alongside its linear extent, another key feature of a PS source is its non-varying radio spectrum. The exact threshold for "low variability" in a broadband SED is not well defined in the literature, but it is accepted that, at the very least, the overall spectral shape of a PS AGN should not appreciably change on timescales less than a few years (Edwards & Tingay, 2004; Hancock et al., 2010). By contrast, Ross et al. (2021) showed that the variability of sources peaking at or below 250 MHz is greater than in their higher-peaked counterparts, but that many of these sources were identified with known blazars.

We here consider the variability of our sample at our selection frequency (888 MHz) by making use of another ASKAP survey, the Variables and Slow Transients Survey (VAST; Murphy et al. 2013). The VAST Pilot surveys, which we use here, comprise two phases, the first of which is made up of 162 × 12 minute observations made between August 2019 and August 2020, with an observing cadence of between 1 day and 8 months (Murphy et al., 2021). Phase II data extends these observations until November 2021. Between both Pilot Surveys, the entirety of the Stripe 82 field is covered by at least nine unique epochs; the radio lightcurves of our sample will therefore be sufficiently well sampled for simple variability calculations. Since the VAST Pilot observations span two years, studying PS variability with them is a valuable test to see whether our sample behaves more like the targeted samples of Edwards & Tingay (2004), Jauncey et al. (2003) and others which are largely non-varying but typically monitored at ≥ 1.4 GHz, or like the low-frequency sample of Ross et al. (2021). Two example lightcurves, one showing significant variability, and one non-varying, are shown in Figure 3.9.

To measure the radio variability of our sources, we follow Barvainis et al. (2005) and Sadler et al. (2006) in using the debiased Variability Index (VI):

$$VI = \frac{100}{\langle S \rangle} \sqrt{\frac{\sum [S_i - \langle S \rangle]^2 - \sum \sigma_i^2}{N}} \tag{3.2}$$

which is calculated on a source-by-source basis and essentially measures a per cent variation in the source flux density. Here, S_i is an individual flux-density measurement, σ_i is its associated uncertainty, N is the number of data points in the lightcurve, and $\langle S \rangle$ is the mean flux density across the lightcurve. Since this statistic takes into account the uncertainty on individual datapoints and it does not specify a time-cadence for observations, it can be used to meaningfully compare variability measures across our entire sample even though our sources fall into several different VAST pointings, and thus were observed at slightly different times and cadences. It will also be minimally impacted by any variations in field calibration.

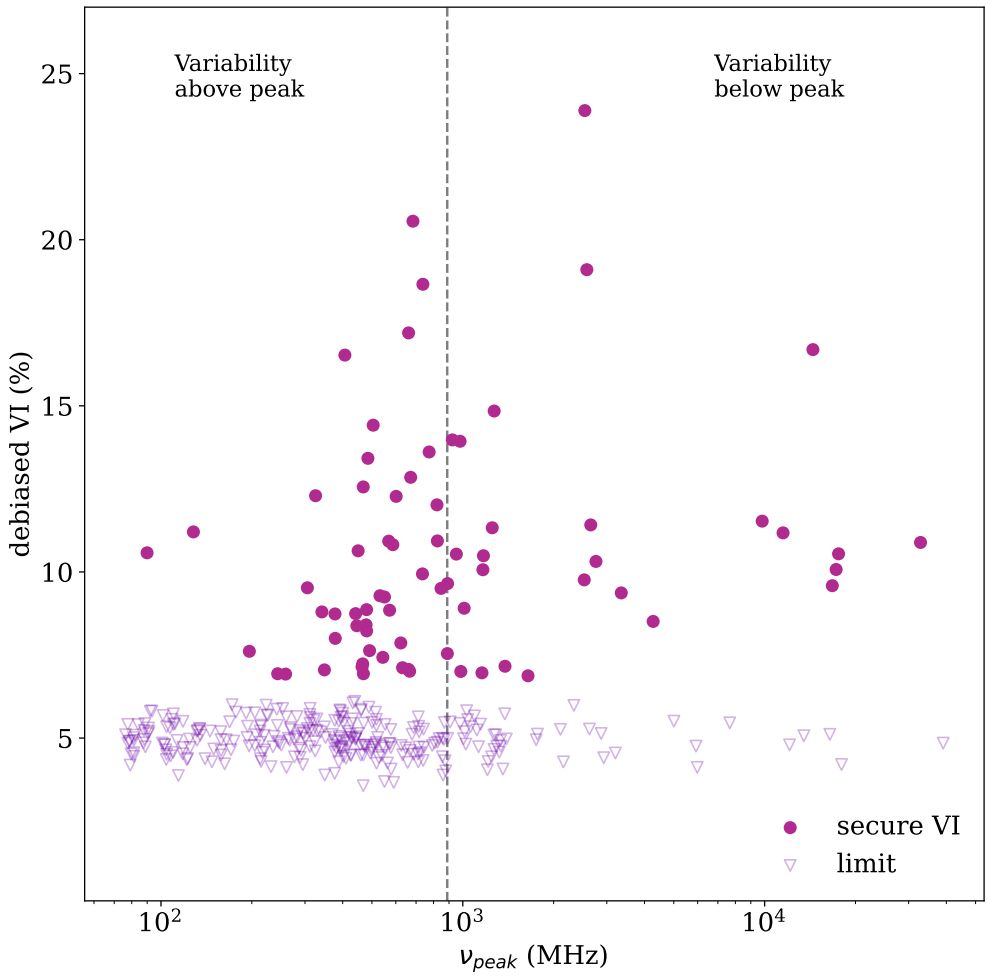


Figure 3.10: The debiased Variability Index of our sources as a function of peak frequency $\nu_{\rm peak}$. Sources for which the measured VI falls below the sensitivity of the VAST Pilot lightcurves are indicated by open arrows, filled circles are those with a secure VI measurement. The vertical dashed line marks the observing frequency of VAST, which divides the plot into sources where we have effectively measured variability above the peak (left) and below the peak (right). Sources for which we have only upper limits on VI have been scattered about a mean value of 5 per cent for ease of viewing.

It is possible for the numerator within the square root of VI to return a negative value if the uncertainties are large enough (an issue particularly for low signal-to-noise sources). In such cases, we follow Sadler et al. (2006) in first setting the VI for these sources to be negative, and then using maximally-negative VI from across our sample to define the sensitivity limit of our data. In this case, the value was found to be 6.8 per cent, and so we consider all VIs below this value as upper limits, indicating a variability < 6.8 per cent on the timescales of the VAST pilot survey. This value was calculated using the peak flux density of sources in each epoch (rather than the integrated flux density) to minimise the effect of the source fitting algorithm.

The variability of our PS sample is summarised in Figure 3.10, where the VI is plotted against the peak frequency of the sources, as obtained from RadioSED. It is clear that the vast majority (73 per cent or 259 sources) have only an upper limit on their VI, consistent with a very low degree of variability. Of the remaining 96 sources, none have a VI above 25 per-cent. We note that 4 sources from amongst our 359 set do not have a measured VI from the VAST Pilot data. This is because in source finding, they were divided into multiple components in at least some of the VAST images (often due to strong sidelobes confusing the sourcefinding algorithm), so it would not be meaningful to compare their measured flux densities across different epochs.

However, we must now return to the question of what constitutes "a low degree of variability", as 25 per cent seems on first inspection relatively high. Indeed, at 4.8 GHz Edwards & Tingay (2004) found 18/20 of

their GPS sources had a variability index less than 8 per cent (the median of their larger, parent sample), and at 20 GHz Hancock et al. (2010) found an average variability of 10–12 per cent for their sample of 21 candidate GPS sources. Our sample median is in line with these previous results, coming to 8 per cent for just those sources above the sensitivity limit. Yet the fact remains that there is a significant tail to the distribution of VI in our sources, which warrants further consideration.

It has been previously found through VLBI observations that some PS sources exhibit extrinsic variability due to Interstellar Scintillation (ISS; Jauncey et al., 2003). To test whether this may be at play here we adopt the model presented in Hancock et al. (2019), which characterises how Refractive Interstellar Scintillation (RISS) can manifest as radio frequency variability in compact sources – including extragalactic sources – on timescales of months to years. The model assumes a scattering disk (source size) of approximately $\theta_{\rm src} = 1$ mas at the galactic latitude of the Stripe 82 field, which is used to produce a modulation index (m_p) at the location of every input source. However, it is unlikely that the majority of flux in our PS sources is concentrated in an area this small. Accordingly, we scale the output m_p from the model by a factor:

$$m_e = \left(\frac{\theta_{\rm scat}}{\theta_{\rm src}}\right)^{\frac{7}{6}} m_p \tag{3.3}$$

following equations 12 and 17 from Hancock et al. (2019). The question now reduces to a question of the angular size of our sources. Conservatively, we assume an angular size of $\theta_{\rm src} = 0.5$ arcsec, which is the approximate scale to which the IPS measurements were sensitive in Section 3.4.2. For an additional constraint, we crossmatch our sample with the 'Astrogeo' VLBI catalogue of Petrov & Kovalev (2025), which is an extensive sample of VLBI sources observed at 2.2 GHz and up. This provides us with VLBI measurements for 74 of our sources, of which 19 have VIs above the sensitivity threshold of our data. The shortest baselines used in the Astrogeo catalogue are somewhere below 1 000 km, if we take this and the lowest observing frequency (2.2 GHz) as reference values, a source detected in that catalogue will have some significant flux on scales $\theta \sim 1.5 \times 10^{-7}$ deg. Again, making the assumption that flux is still contained on this angular scale at the VAST frequency (888 MHz), we can use this as a new value for $\theta_{\rm src}$ in those sources appearing in that catalogue. This is what is shown in Figure 3.11, where the filled circles in the top panel correspond directly to the filled circles in Figure 3.10 as sources with a well-constrained VI. Sources that appear as unfilled circles in all three panels are those below our sensitivity limit in VI, so no useful comparison to the models can be made.

Clearly, the majority of sources with well-constrained VI have variability which is above the level of RISS with the source sizes assumed, which would suggest an intrinsic source of variability. Crucially though, over half (12/19) of the sources which appear in Astrogeo have a modelled VI that is larger than that measured here, suggesting that the radio variability detected in VAST could be consistent with external effects caused by a scattering screen. This is not the first time such a phenomenon has been observed towards PS sources, as Jauncey et al. (2003) identified at least one of their GPS candidates which exhibited variability directly attributed to ISS. Since the Astrogeo catalogue is not complete, in the sense that it does not contain VLBI measurements for every known source in any region of sky, it is entirely possible that more of our PS sample have a significant fraction of their flux contained within small angular scales, and that some of the variability seen here may indeed be due to RISS.

To say more on the variability of these sources is beyond the scope of this paper, so we end this section simply by reiterating that our PS sample on the whole does show a low degree of variability, with a median of just 8 per cent amongst those sources above the sensitivity limit in VI. This is in line with the higher-frequency findings of Jauncey et al. (2003); Edwards & Tingay (2004) and Hancock et al. (2010), and none of our sources show extremely high variability indicative of blazar-like activity. Though some of our sources do exhibit variability as high as 25 per cent, some portion of this may be due to RISS, especially given that a number of sources detected in the VLBI Astrogeo catalogue have significant flux on scales compact enough to induce a large degree of scattering. However some of this variability may of course be intrinsic to the source and caused by events such as transient shocks or secular source evolution.

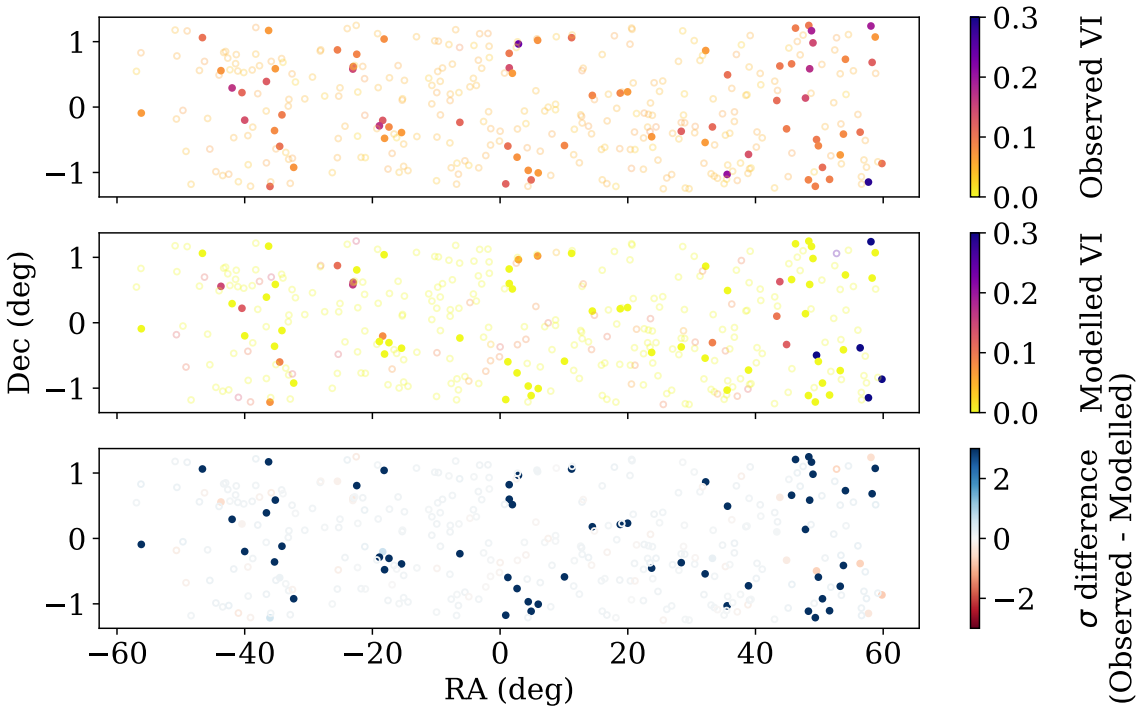


Figure 3.11: Observed debiased VI (top panel) compared to that predicted by the Hancock et al. (2019) model (middle) assuming a source size of either $\theta_{\rm src} = 0.5$ arcsec or $\theta_{\rm src} = 1.5 \times 10^{-7}$ deg if a source is included in Petrov & Kovalev (2025). The bottom panel shows the difference between the observed and modelled values. Unfilled circles are those sources where the observed debiased VI is below our sensitivity limit, and so comparison with the model is not meaningful.

3.5 The hosts of Peaked Spectrum sources

In this section we outline the properties of the host galaxies for our PS sample, where these sources have been detected at optical and/or infrared wavelengths. Since the main focus of this work is on the nature of the radio sources themselves, we restrict this discussion mainly to a consideration of the cosmological distances of our sources, as well as the classification of their hosts in the broadest terms as galaxies or quasars (QSOs). We show that 97 of our sources have spectroscopically identified redshifts, and a further 69 have some constraints on their distance from photometric redshifts. The spectra and photometric colours show that our sources are fairly evenly distributed amongst galaxy-type and QSO-type hosts, while combining this data with WISE photometry reveals that many of our sources are likely distant ($z \geq 1$).

3.5.1 Spectroscopy

We obtain spectroscopic redshifts for our sources by crossmatching primarily against the LARGESS (Ching et al., 2017), SDSS-DR16 (Ahumada et al., 2020), and the Dark Energy Spectroscopic Instrument Data Release 1 (DESI-DR1) (DESI Collaboration et al., 2025) catalogues. Additionally, we search the NASA/IPAC Extragalactic Database (NED) for supplemental spectroscopic redshifts, of which we obtained three. Since all of our sources are compact as outlined in Section 3.4.2, we simply search for optical matches within a 5 arcsecond radius across the catalogues which do not already provide a radio-optical association (SDSS-DR16, DESI-DR1 and NED).

In total, we obtain spectroscopic redshifts for 132 sources in our sample, of which 69 have QSO-like spectra, and 63 are identified as galaxies. From amongst these sources, there were 6 where DESI and SDSS classifications differed (one survey pipeline preferred a Galaxy-type spectral template, the other a QSO-type template), perhaps due to the differing fibre sizes capturing different ratios of light from the AGN and host galaxy. Manual inspection of these sources revealed unresolved point sources in the optical images, with visible components from both galaxy and AGN-driven emissions in the spectra, and in a more complex

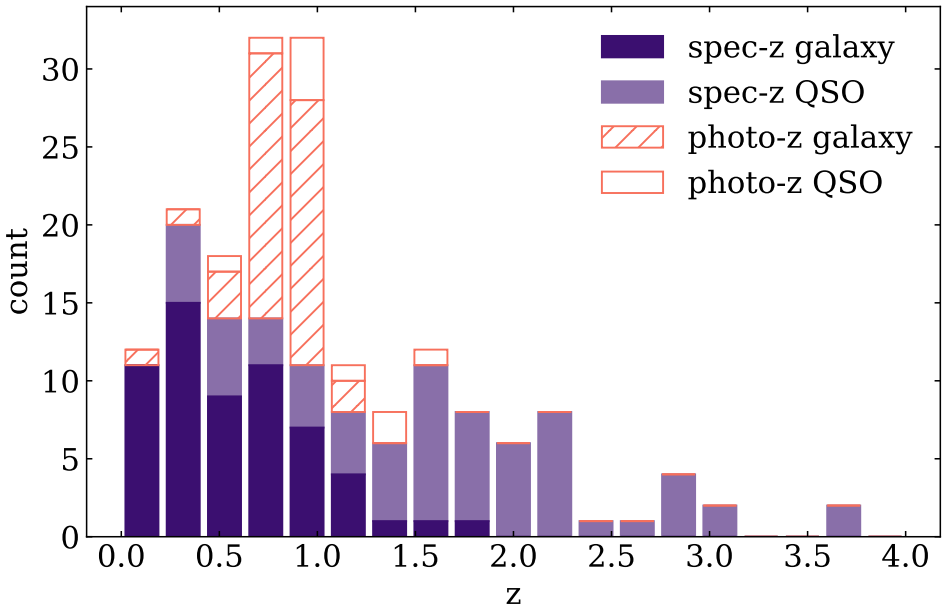


Figure 3.12: The distribution of spectroscopic (filled) and photometric (unfilled) redshifts for our sample, divided into galaxies (dark filled/hatched) and QSOs (light filled/unfilled). The apparent sharp declines in photometric redshifts after $z \sim 1$ and spectroscopic redshifts after $z \sim 2$ are discussed in text.

classification scheme these might be called 'Seyfert 1' or 'Broadline AGN' depending on their radio loudness. For simplicity here, we have included these in the 'QSO' class. The spectroscopic redshifts of these sources span $0.04 < z < 3.7$, though above $z \sim 1.7$ all of our spectroscopically-identified sources are quasars, a natural consequence of the magnitude limits on optical surveys. We engage in a fuller analysis of these redshifts below, where they can be considered alongside photometrically-identified redshifts.

3.5.2 Photometry

We supplement our spectroscopic redshifts with photometric redshift estimates drawn primarily from two large-area surveys which cover Stripe 82 in its entirety: the Quaia catalogue (Storey-Fisher et al., 2024), and the WISE-PS1 catalogue (Beck et al., 2022). Both of these combine photometry from the Wide-field Infrared Survey Explorer (WISE; Wright et al. 2010) with an optical survey, though each targets a different class of extragalactic sources. Specifically, Quaia combines unWISE and Gaia photometry to target quasar photometric redshifts, while WISE-PS1 makes use of Pan-STARRS1 photometry to target galaxies. These two datasets are therefore complementary, as they cover independent subsets of our sample. Furthermore, a comparison of the photometric and spectroscopic redshifts for those sources in our sample possessing both shows good agreement across both the Quaia and WISE-PS1 selected subsets. To these we add three photometric redshifts from data release 9 of the Legacy Survey (Dey et al., 2019; Duncan, 2022), and a further five photometric redshifts that were returned as part of our NED search above; in each case these help characterise sources which were otherwise lacking redshift information. Overall, we find 129 photometric redshifts for our sample, of which 47 (8 QSO/39 galaxy) provide constraints on sources without spectroscopic identifications.

A histogram of the combined spectroscopic and photometric redshift estimates is given in Figure 3.12, where it is clear that a relatively large fraction of galaxy-type photometric redshifts (hatched bars) are clustered around $z = 1$. This is likely an artefact of the WISE-PS1 catalogue used, as redshift estimates in that work are made via a machine learning approach for which the training set of spectroscopic galaxy redshifts tails off sharply at $z = 1$ (see Fig. 1 in Beck et al. (2022) for reference). Nevertheless, we keep these redshifts as a useful first-order constraint on sources that otherwise have no distance or host information (that is, are not otherwise divided into galaxy or QSO type sources).

In addition to the cut-off in photometric redshifts, we see the distribution of spectroscopic galaxy (shaded, dark) and QSO (shaded, light) redshifts fall off at $z \sim 1$ and $z \sim 2$, respectively. In the case of the galaxies,

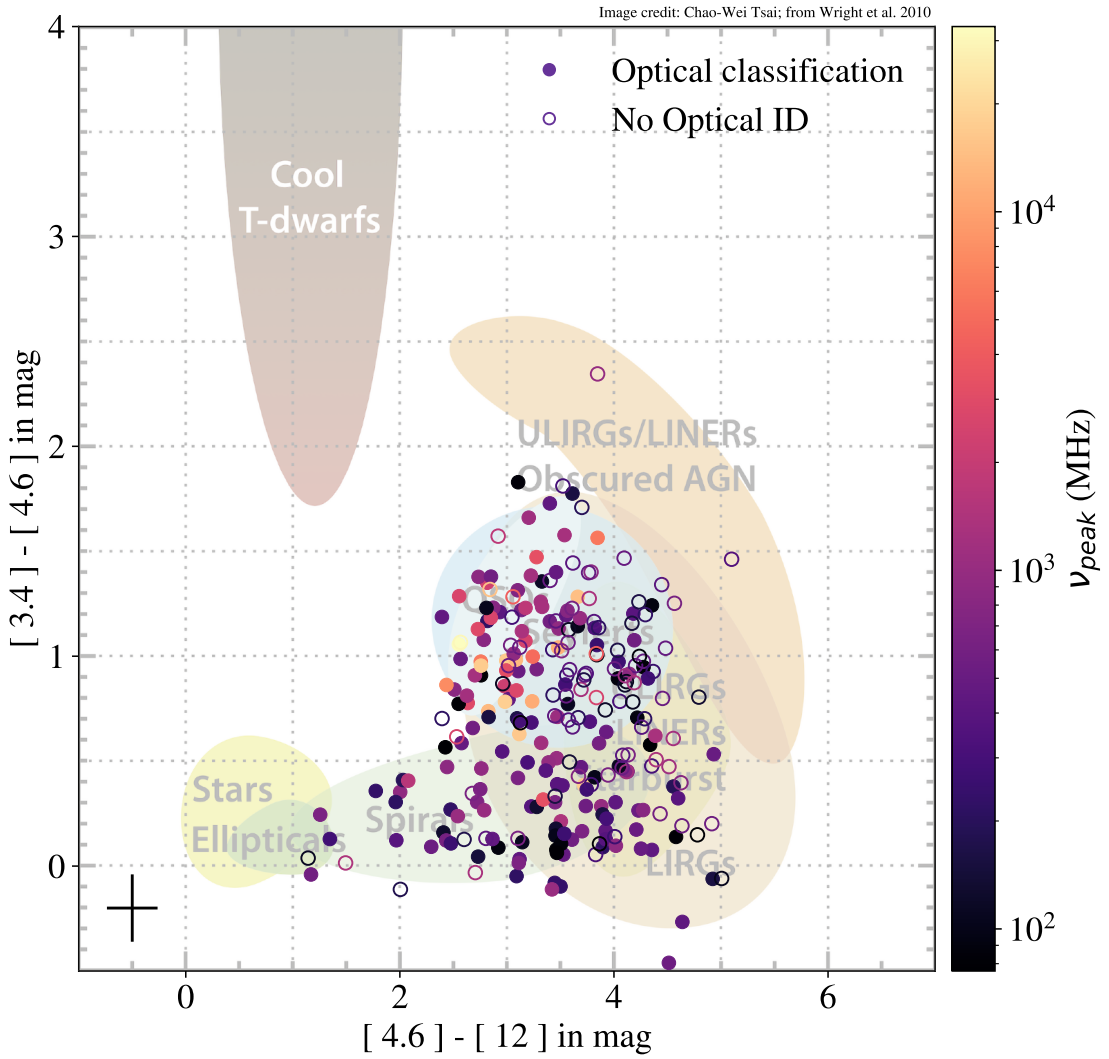


Figure 3.13: The distribution of our sample in the two colour WISE diagram, with the familiar classification groups from Wright et al. (2010) overlaid for comparison. Filled points have an optical classification from either spectroscopy or the photometric catalogues we referenced in Section 3.5.2, unfilled points do not. Points are coloured by the peak frequency of their broadband radio SED, $\nu_{\rm peak}$. The average uncertainty on these colours is indicated by the errorbars in the bottom left of the plot.

this is likely due to the underlying sensitivity limit of SDSS, while for the quasars it is a consequence of the targeting algorithm which, as of SDSS-IV, was designed to maximise completeness at $z \leq 2.2$ (Myers et al., 2015). Taking this into account alongside the discussion of the photometric training sample above, there are no obvious trends in the redshift distribution of our PS sample. What is perhaps more interesting then, is a consideration of the sources that lack redshift estimates, and indeed also optical photometry. For these, we turn to an analysis at infrared wavelengths in the following section.

3.5.3 WISE colours

For a fuller consideration of the host galaxies of our sample, we now analyse their redshift and optical classification in conjunction with IR photometry from the WISE survey. For this, we first match against the AllWISE catalogue (Cutri et al., 2021), again using a simple, radial constraint of 5 arcseconds. This returns a total of 250 unique matches from amongst our sample, corresponding to a match rate of $\sim$ 70 per cent.

The WISE colour-colour space

The locus of our sources in the WISE colour-colour plot is shown in Figure 3.13, with points coloured according to the turnover frequency in their radio SED. There are two striking features about the IR properties of these

sources which may in fact be linked; the distribution of peak frequencies, and the lack of Early-type hosts.

Beginning with the first of these, we see that the sources with the highest frequency spectral turnover are most closely clustered around the region typically inhabited by QSOs/Seyferts, with $\nu_{\rm peak}$ falling as the WISE colours of sources deviate further from this space. This is quite interesting, as even without accounting for the redshifting of the peak frequency, the QSOs in our sample would seem to have smaller jets than their galaxy-type counterparts, recalling that on the whole, the QSOs are more distant in our sample than the galaxies (meaning their peaks would be redshifted to comparatively lower frequencies than galaxy-type sources; see Figure 3.12). Furthermore, where available, the optical classifications of our sample support this finding, with the QSO-type objects falling comfortably within the QSO/Seyfert space of the WISE colour-colour plot. This reinforces the notion that the sources with the highest frequency peaks are indeed hosted by QSOs. The absolute $i-$band magnitudes of these sources (M_i) are also indicative of QSO activity; all of our sources possessing an optical redshift and WISE colour $W1-W2>0.5$ are brighter than $M_i<-22$ mag, which is the tail end of the $i-$band absolute magnitudes in the SDSS-DR16 Quasar Catalogue (Lyke et al., 2020). This may suggest that there is something unique about the environment of QSOs which tends to trigger the very smallest radio jets, whereas sources with galaxy-type optical hosts are less likely to do so, and indeed galaxy-type objects with radio peaks above a few gigahertz were not detected in our sample. In support of this interpretation, we note that Hancock et al. (2009) found the majority of PS sources identified at gigahertz frequencies in AT20G were quasar-type, while the recent work of Nyland et al. (2020) looked specifically for changing-look quasars in the Very Large Array Sky Survey (VLASS) at 3 GHz, and identified several with radio SEDs indicative of recent triggering (i.e. with high frequency peaks), suggesting a selection-philosophy broadly consistent with our interpretation here.

On the subject of the broader distribution of our sample in the WISE colour-colour space, we note that only ~ 3 per cent (8/267) of our PS sources detected in all four WISE bands would be considered 'Wise Early-type' ($W2-W3<2$), while some 47 per cent of optically-identified sources are hosted by QSOs. This latter fraction is similar to the proportion of QSOs in one of the first canonical samples presented in O'Dea (1998), but the low fraction of Early-type galaxies is worth a further moment of consideration. It is lower than the fraction of Early-type galaxies in the young radio galaxy sample of Kosmaczewski et al. (2020), though not by much considering the small size of that sample ($5/29\approx 17$ per cent are Early-type in that work), and the additional selection criteria imposed there (X-ray detection, spectroscopically-constrained redshift, constraint on linear size of the radio jets). What is perhaps more interesting is a comparison with the "Fanarof-Riley Type 0" (FR0) sample of Sadler (2016), of which almost 67 per cent exhibit a low star formation rate consistent with Early-type host galaxies. That sample was constructed based on the relatively sparse frequency-coverage available at the time. Therefore, the authors did not require their sources to exhibit broadband spectral turnover, only compactness at 1 GHz in addition to the 20 GHz selection frequency, making it a sample of *candidate* PS sources. Nevertheless, the only additional selection criterion imposed was a crossmatch with an optical counterpart, which makes it a useful point of comparison against our work. The discrepancy between the mid-IR colours of that sample and ours is therefore striking, and suggests we are perhaps probing some subset of that larger sample of compact objects, which does not include sources with Early-type hosts. Indeed, the WISE colours of our sample are more akin to those of the extended, lobe-dominated (FR-II) radio galaxy sample presented in van Velzen et al. (2015) — discussed further with reference to PS sources in Kosmaczewski et al. (2020). This once more raises the question, discussed in Sadler (2016), as to the connection between PS sources specifically, and the broader class of compact radio objects; how these two different classification methodologies overlap, and what can be said about how each type of radio source will evolve (or not) into more extended structures as it ages, a discussion recently revived in the series of papers beginning with Kiehlmann et al. (2024a).

WISE photometry and approximate source distances

We see also in Figure 3.13 a certain clustering of the the sources without optical redshifts, either spectroscopic or photometric. Again, these appear to be centered around the typical locus of QSOs/Seyferts and starburst galaxies, with just a very few sources possessing WISE colours indicative of an elliptical, or an obscured AGN. This may suggest many of our sources without redshift information are high-z QSOs. To explore this further, we redo the IR crossmatch with the catWISE 2020 catalogue (Marocco et al., 2021) to obtain greater W1 completeness for our sample (catWISE2020 is 99 per cent complete for $W1<15.1$ mag, whereas AllWISE is only 95 per cent complete at these magnitudes). Since catWISE2020 only contains information for the first

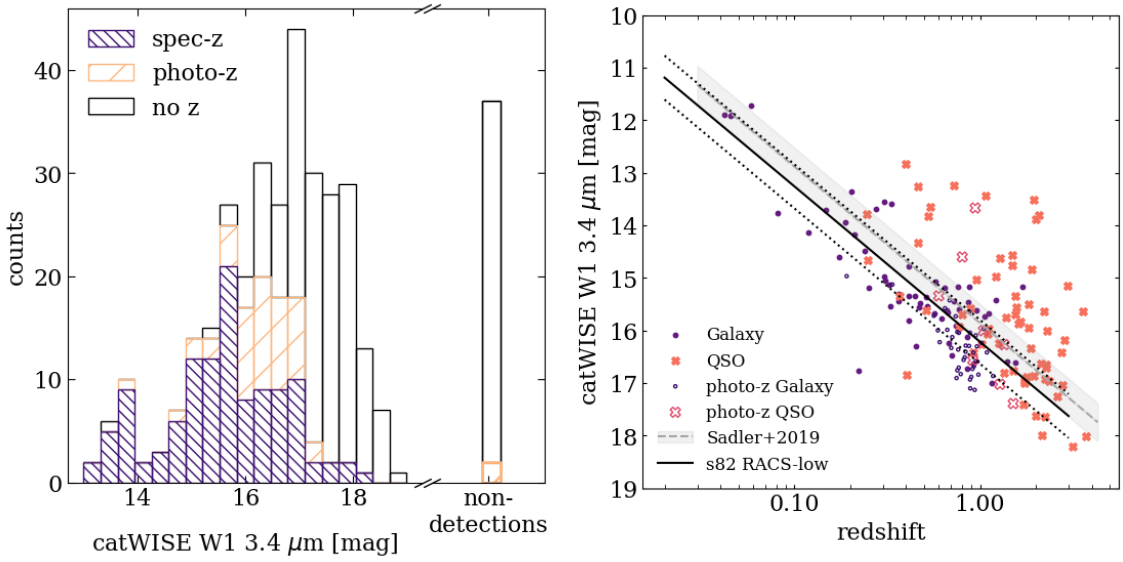


Figure 3.14: Left: The distribution of spectroscopic (densely hatched, purple) and photometric (sparsely hatched, orange) redshifts for our sample as a function of WISE W1 magnitude, as well as sources without any redshift constraint (unfilled). Some 38 of the sample have neither WISE photometry nor a redshift estimate, which are candidate high-redshift sources as discussed in-text. Right: Spectroscopic (filled points) and photometric (unfilled points) redshifts for our sample as a function of WISE W1 magnitude, divided into QSO-type (crosses) and Galaxy-type (circles) sources. Dashed grey line is the linear fit from Sadler et al. (2019), with shading indicating the median scatter in that sample. Solid black is fit to a full sample of RACS-low radio sources from Stripe 82, with black dotted lines indicating the median scatter in this sample. Some 38 of the sample have neither WISE photometry nor a redshift estimate, which are candidate high-redshift sources as discussed in-text.

two WISE bands, we perform a simpler analysis presented in Figure 3.14. In the left-hand figure, we show the distribution of sources as a function of W1, broken down into those with spectroscopic redshifts (right-hatched, navy bars), photometric redshifts (left-hatched, orange bars) and without redshift information (unfilled bars). It is clear that redshift completeness is a strong function of $W1$ magnitude, with only 9 per cent of sources (14/154) fainter than W1 = 17 mag possessing redshift information.

The WISE magnitudes of our sources can tell us something about their distances, even in the absence of redshift information, in a variation on the familiar $K - z$ relation. Although this relation only holds for galaxy-type hosts (it relies on a simple scaling of flux from the stellar population with redshift), it can also be used as a lower-limit on QSO-type hosts, where the optical-IR light is dominated by the central AGN but nevertheless will have some contribution of starlight. LaMassa et al. (2013) discussed this in relation to X-ray sources within Stripe 82, noting that sources undetected in WISE or optical photometry could be either highly dust-obscured, or candidate high-redshift sources. Closer to our frequencies of interest, Sadler et al. (2019) showed that the W1 magnitudes of a number of radio galaxies detected at 162 MHz with the MWA could be fit by the following linear equation:

$$W1 = 15.860 + 2.976 \log(z) \tag{3.4}$$

We plot the W1 magnitudes of our sample as a function of redshift in the right-hand panel of Figure 3.14, where filled markers indicate spectrosocpic redshifts, and unfilled photometric. Equation 3.4 is shown as a dashed grey line, with grey shading either side indicating an offest of ±0.36 mag, the median scatter in their sample. On top of this we have added a solid black line indicating a similar linear fit, but this time to all RACS-low point sources within Stripe 82 (i.e. sources with 1 Gaussian component in the radio catalogue) which are brighter than 10 mJy, have a spectroscopic, galaxy-type redshift from SDSS-DR16 and a catWISE counterpart within 3 arcseconds of the radio position. This fit has the form:

$$W1 = 16.218 + 2.958 \log(z) \tag{3.5}$$

and the median scatter in the sources about this fit is ±0.42 mag, as indicated by the dotted lines. This new fit corresponds fairly closely to the one derived previously for compact radio galaxies bright at 162 MHz, just shifted down by approximately $1-\sigma$, which suggests that radio galaxies selected at our chosen frequency (888 MHz) and sensitivty ($S \geq 10$ mJy) overall reside in slightly less luminous hosts than those in the Sadler et al. (2019) sample, which were selected to be highly luminous at radio wavelengths, and are also likely to have a strong non-thermal contribution to their WISE W1 magnitude. As is to be expected, many of the QSOs in our sample lie above both relations, since the central AGN can provide a strong excess of optical-IR emission over that provided by extended star formation. However, the galaxy-type sources in our sample lie largely below the Sadler et al. (2019) fit, and this offset still persists when considering our fit to a radio-matched sample, albeit to a much lesser degree, as the median offset of our PS sample is 0.23 mag fainter than the simple linear fit. From this we conclude that our sample of PS sources reside in somewhat less luminous hosts than the bright, strongly scintillating sources of Sadler et al. (2019), and also in slightly less luminous hosts than the majority of other radio galaxies in the field.

Even so, we can use our linear fit in Equation 3.5 to obtain a first estimate for the distances of the objects for which no redshift is available. As was stated in Sadler et al. (2019), we caution that this is only an estimate of the lower-limit on source redshifts, as the QSOs typically have an additional flux contribution from their core as discussed above, and are thus not well described by the fit we performed. Nevertheless, as stated above the vast majority of our sources without spectro-photometric information have catWISE magnitudes W1 > 17 mag, which would put them at redshifts $z \geq 1.8$ from our linear fit (or $z \geq 1.3$ if we were to take the lower bound, recognising that our PS sources tend to fall slightly below the fit), while the 37 sources un-detected in W1 could be as distant as $z \geq 3$ (using the 90 per cent completeness limit of catWISE2020, corresponding to 17.7 mag).

In summary, we obtain spectro-photometric redshift estimates for 184 of our 359 sources, divided fairly evenly between QSO-type (79/184) and galaxy-type (105/184) hosts, and largely restricted to $z \leq 1$. The WISE colours of these sources show many to be dominated by AGN emission at IR wavelengths, and we have comparatively few sources with early-type host galaxies. Following the method of Sadler et al. (2019), we find from the relation between W1$-z$ that many of our PS sources are likely distant, perhaps as high as a redshift of $z \sim 3$, which makes this an interesting population with which to study the cosmological evolution of compact radio jets.

3.6 The population of PS sources

We consider here the radio properties of our PS sample in the larger context of radio populations. We show that the luminosity distribution of our sources is typical of previous PS samples, though we are lacking some of the highest luminosity sources from earlier works. We finally consider briefly the nature of AGN environments conducive to jet launching.

3.6.1 Radio Luminosity

We calculate the 5 GHz radio luminosity of the 184 PS sources with redshift estimates as:

$$L_{5\mathrm{GHz}} = \frac{4\pi D_L^2 S_{5,\,\mathrm{rest}}}{1+z} \tag{3.6}$$

where D_L is the luminosity distance in metres, and z is the redshift of the source. We note here that instead of performing a $k-$correction according to the spectral index of the source, α, we have simply sampled the SED at a frequency corresponding to 5 GHz in the source restframe ($S_{5,\,\mathrm{rest}}$). A frequency of 5 GHz was chosen for purely historical reasons, to allow for easy comparison with samples from the literature.

The luminosity distribution of our sample is shown in Figure 3.15, along with a few key PS samples from previous work. Firstly, we note that, from amongst our full 359-source sample, the number for which we can calculate luminosities is directly comparable to the other samples highlighted here. This is despite the fact that this sample focuses on a much smaller area of sky; Callingham et al. (2017) covered the entire Southern equatorial sky, Slob et al. (2022) sampled 740 deg^2, whereas Stripe 82 is only 300 deg^2. This is likely due to both the deep optical coverage in our chosen field, but also the fact that we are maximising the use of legacy

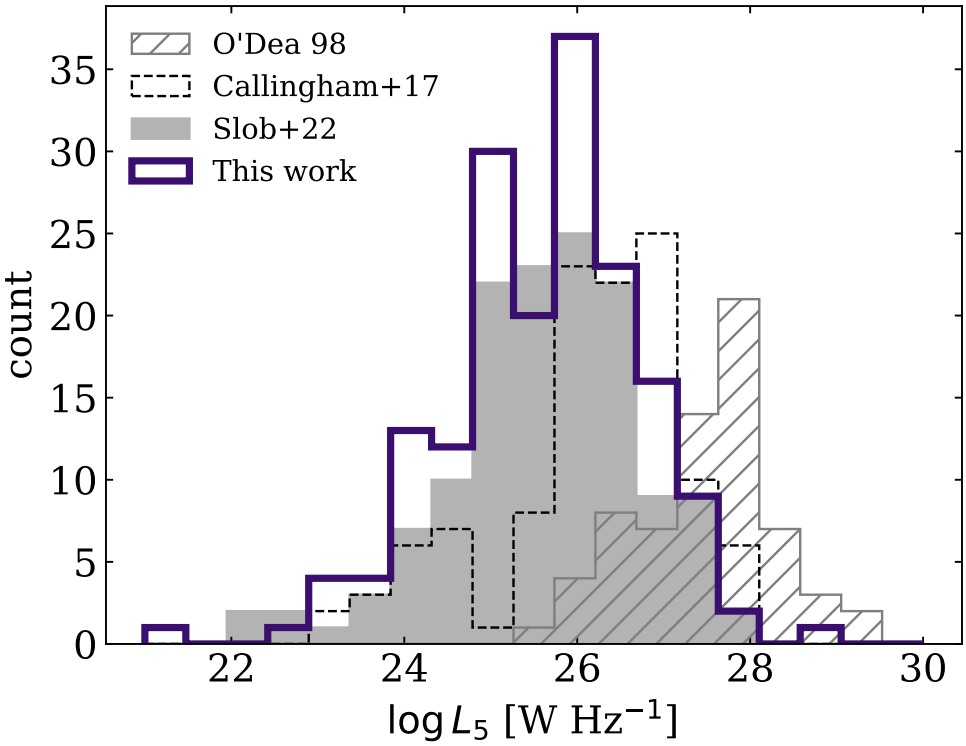


Figure 3.15: 5 GHz radio luminosity distribution of our sample compared to relevant ones from the literature.

datasets to extend our search down to fainter radio flux densities. Therefore, a judicious use of legacy radio frequency catalogues across larger areas of sky is likely to increase the sample of PS sources for which we can constrain distances by several orders of magnitude.

Beyond comparisons of sample size alone, we note that the luminosity distribution of our sample most closely resembles those more recent distributions shown in Figure 3.15. As in Callingham et al. (2017), we too lack the highest luminosity sources above $L_5 = 10^{28}\,\mathrm{W\,Hz^{-1}}$. However, where in that work the lack of high luminosity sources from a large-area search was interpreted as a "convolution of source evolution and redshift", here it could be simply that our on-sky area is not large enough to see the very rarest and brightest radio galaxies. Crucially, as was seen in the Slob et al. (2022) sample, we are now starting to probe the faint luminosity end of the radio AGN population, though we are still about two orders of magnitude away from the faintest end of the local radio luminosity function where star forming galaxies begin to dominate (Mauch & Sadler, 2007). It is very likely that future samples combining legacy radio surveys over larger sky areas than this one will greatly increase the number of such sources, and our capacity to study them will only grow with the greater sensitivity of upcoming surveys from the next generation of radio instruments.

3.6.2 The significance of Galaxy-type and QSO-type hosts

In performing our analysis on this new sample of PS sources, we have placed particular emphasis on the optical identification of the hosts as either galaxies or QSOs. It is well known that jet-driven or 'radio-mode' AGN emission is typically associated with radiatively-inefficient accretion, while radiatively-efficient QSOs exhibit strong, jet-driven radio emission in only ~ 10 per cent of cases, and it is still not known exactly why (Kellermann et al., 2016). Furthermore, this fraction of jet-dominant QSOs does not appear to evolve with cosmic time, even as far as $z \approx 6$ (Keller et al., 2024, and references therein). Studying the conditions in which we see the most compact radio jets may therefore provide interesting insights into the complex interplay between AGN fuelling and feedback mechanisms across cosmic time.

In Figure 3.16 we present the peak flux density, S_{peak}, as a function of the peak frequency in the rest-frame of the source, $\nu_{\mathrm{peak,\,rest}}$. Here, the unfilled markers indicate those sources with photometric constraint on redshift only, and filled points indicate spectroscopic redshifts are available. It is clear that the sources with the highest

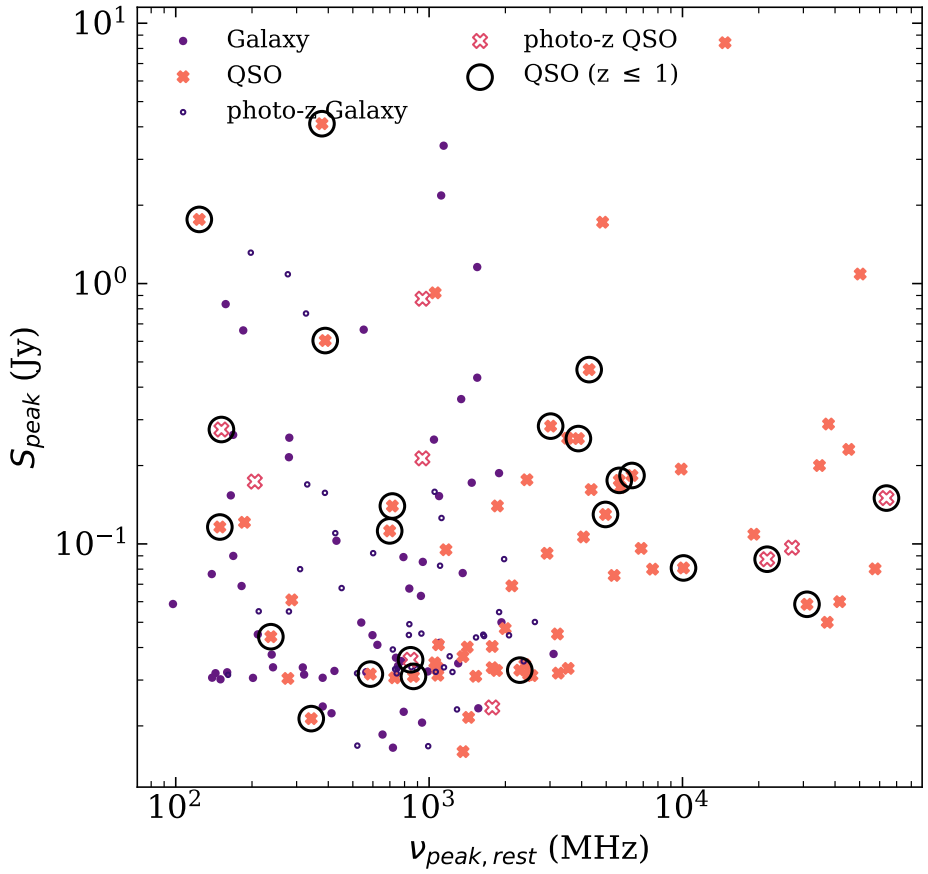


Figure 3.16: The distribution of peak flux densities ($S_{\rm peak}$) as a function of $\nu_{\rm peak,rest}$, the peak frequencies of our sample in the source rest-frame. As in Figure 3.14 panel b, marker shapes indicate optical identification with a galaxy (circle) or QSO (cross), with filled points having a spectroscopic identification, and unfilled points photometric classification. QSOs below a redshift of 1 are encircled with a larger, black marker.

frequency peaks in our sample, corresponding to the most recent triggering, are QSO-type, while no galaxy-type PS sources are found to peak above ~ 2 GHz in our sample. It is possible that this is partially a selection effect since quasars are optically brighter and can be seen out to higher redshifts. There may also be additional beaming effects at play, whereby intrinsically less luminous PS sources have their radio emission boosted in QSO-type hosts due to the approximate alignment of the jet-axis and our line of sight. However, if we restrict the sources in this plot to those with redshifts $z \lesssim 1$ (where the optical surveys we crossmatch against are sensitive to Galaxy-type hosts), there are still a number of QSO-type sources with peaks above a few gigahertz. These are indicated in the plot by the enclosing black circles.

Although not as marked, this phenomenon was apparent in the early, canonical sample of PS sources from O'Dea (1998), where only 20 per cent (4/18) sources peaking above 2 GHz had galaxy-type hosts. This also is in line with the philosophy behind the search for changing-look quasars conducted by Nyland et al. (2020), who found 14 radio sources peaking upwards of several gigahertz in the restframe, all hosted QSO-type objects, using a multi-epoch radio search. As already discussed in Section 3.5.3, this suggests that radio jets may be preferentially triggered in QSO-type objects, and as they grow and extend beyond the medium of their host the central AGN may transition from radiatively-efficient accretion into the inefficient accretion typically associated with extended radio structures. If the radio luminosity of these jets increases as they grow up to scales of a few kiloparsecs, as suggested by the models of Shabala et al. (2008) and Young et al. (2025), then such evolution of the hosts would be broadly consistent with the observed anti-correlation between radio-loudness and Eddington-scaled accretion rate (Ishibashi et al., 2014). A fuller analysis of radio jets on all-scales, coupled with detailed study of the hosts' optical spectra, will no doubt provide greater insight into this complex picture of jet triggering and growth, especially when coupled with insights from both idealised (Borodina et al., 2025) and cosmological simulations (Thomas et al., 2025).

3.7 Conclusions & Future Outlook

In this work we have presented a new, uniform sample of sources with compact radio jets, thought to be recently-triggered radio AGNs. We have shown that our RadioSED method for fitting and classifying broadband radio flux density measurements is able to increase the number of known PS sources within the Stripe 82 field by more than an order of magnitude (Section 3.3), especially when making use of the latest generation of surveys including GLEAM-X DRII (Ross et al., 2024) (Section 3.3.1). Estimates of the sensitivity of our method suggest that we are able to extend the completeness of our sample in both $\nu_{\rm peak}$ and $S_{\rm peak}$ by more than an order of magnitude compared to previous work (Section 3.3.2).

From amongst the 359 sources identified by RadioSED, we recover all 11 previously known PS sources (Section 3.3.3). Our sample is more compact than the general population of radio galaxies, at both the resolution of the continuum surveys used for fitting (Section 3.4.1), and at the sub-arcsecond resolution from widefield IPS measurements (Section 3.4.2). However, sources with 'retriggered' SEDs tend to be less compact at low frequencies, likely due to extended, lobe-dominated emission. Fewer of our sources are compact at 162 MHz than the PS sources studied in Chhetri et al. (2018). The 800 MHz variability of this sample is on the whole low, with the VAST pilot data only able to provide an upper limit on the debiased VI of most sources (Section 3.4.3). In those sources which are variable, some of this variability may be explained by RISS, especially where there is evidence of milliarcsecond cores from Petrov & Kovalev (2025).

The multiwavelength properties of our sources suggest that these PS sources are fairly evenly divided amongst galaxy- and QSO-type hosts (Section 3.5) in line with the divide amongst previous PS samples. However, where our sources have WISE colours fewer appear to belong to early-type hosts than in comparable FR0-type samples (Section 3.5.3), reinforcing the idea that FR0 and PS type sources are drawn from distinct (but perhaps overlapping) populations. While many of our sources that are not spectro-photometrically constrained are likely distant, potentially residing at $z \sim 3$ or higher (Section 3.5.3), those sources for which distances are well constrained show a luminosity distribution in good agreement with previous samples (Section 3.6.1). When examining the rest-frame peak frequencies (a direct proxy for source size and hence also age), we see that the most compact sources in our sample tend to be QSO-type, which we discuss in the context of AGN jet triggering (Section 3.6.2).

Overall, the sample of PS sources presented here represents a new step in the study of radio AGN jets in their early stages post-triggering. Thanks to the wealth of radio and multiwavelength data already available, we are now at the stage where we can assemble large, statistical samples of these sources using tools like RadioSED, and begin to probe their properties on a population-level. Our capacity for this large-scale analysis will only grow with the next generation of surveys from the SKA (Carilli & Rawlings, 2004) and ngVLA (Murphy et al., 2018), and with the rich datasets already available to us from decades of observing across dozens of instruments and several continents.

Acknowledgements

We would first like to thank our referee, Professor D.J. Saikia, for providing useful comments and additional references which greatly improved the overall clarity of the paper. This research was supported by an Australian Government Research Training Program (RTP) Scholarship, and by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. Parts of this research were also supported by the Australian Research Council Centre of Excellence for Gravitational Wave Discovery (OzGrav), project number CE230100016.

This research has made use of NASA's Astrophysics Data System Bibliographic Services, the crossmatch service provided by CDS, Strasbourg, and the NASA/IPAC Extragalactic Database (NED) which is funded by the National Aeronautics and Space Administration and operated by the California Institute of Technology. Furthermore, this research makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration. We also made use of Astroquery

(Ginsburg et al., 2019) and Astropy:[1] a community-developed core Python package and an ecosystem of tools and resources for astronomy (Astropy Collaboration et al., 2013, 2018, 2022).

This research used data obtained with the Dark Energy Spectroscopic Instrument (DESI). DESI construction and operations is managed by the Lawrence Berkeley National Laboratory. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High-Energy Physics, under Contract No. DE–AC02–05CH11231, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract. Additional support for DESI was provided by the U.S. National Science Foundation (NSF), Division of Astronomical Sciences under Contract No. AST-0950945 to the NSF's National Optical-Infrared Astronomy Research Laboratory; the Science and Technology Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation; the Heising-Simons Foundation; the French Alternative Energies and Atomic Energy Commission (CEA); the National Council of Humanities, Science and Technology of Mexico (CONAHCYT); the Ministry of Science and Innovation of Spain (MICINN), and by the DESI Member Institutions: www.desi.lbl.gov/collaborating-institutions. The DESI collaboration is honored to be permitted to conduct scientific research on I'oligam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. National Science Foundation, the U.S. Department of Energy, or any of the listed funding agencies.

This scientific work uses data obtained from Inyarrimanha Ilgari Bundara, the CSIRO Murchison Radio-astronomy Observatory. We acknowledge the Wajarri Yamaji People as the Traditional Owners and native title holders of the Observatory site. CSIRO's ASKAP radio telescope is part of the Australia Telescope National Facility (https://ror.org/05qajvd42). Operation of ASKAP is funded by the Australian Government with support from the National Collaborative Research Infrastructure Strategy. ASKAP uses the resources of the Pawsey Supercomputing Research Centre. Establishment of ASKAP, Inyarrimanha Ilgari Bundara, the CSIRO Murchison Radio-astronomy Observatory and the Pawsey Supercomputing Research Centre are initiatives of the Australian Government, with support from the Government of Western Australia and the Science and Industry Endowment Fund.

Data Availability

Archived data from ASKAP surveys, including VAST and RACS, can be obtained through the CSIRO ASKAP Science Data Archive, CASDA (http://data.csiro.au/).

Other survey data used in SED construction and multiwavelength analysis is all readily available through available through the CDS VizieR catalogue service.

Measurements derived from radio catalogues and not elsewhere published are available in the online version of this paper. Samples of these datasets are contained in Appendices D and E.

Any additional data is available upon reasonable request to the corresponding author.

[1] http://www.astropy.org

4

The neutral gas properties of PS AGN: a theoretical perspective

Emily F. Kerrison[1,2,3], Samuel R. Ward[1], Stephanie Tonnesen[1], Vanessa A. Moss[3,2], and Elaine M. Sadler[2,1,3]

[1]Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA
[2]Sydney Institute for Astronomy, School of Physics A28, University of Sydney, NSW, 2006, Australia
[3]ATNF, CSIRO Space and Astronomy, PO Box 76, Epping, NSW 1719, Australia

This chapter has been submitted as *Forward Modelling the ASKAP-FLASH 21cm Absorption Survey with SANGRiA: Geometry Alone Cannot Explain High HI Detection Fractions Towards Compact Radio Sources.* Kerrison E. F., Ward S. R., Tonnesen S., Moss V. A. and Sadler E. M., to the *Astrophysical Journal*

The question will have to remain unanswered until more data have been taken

Jansky (1935) on the discovery of extra-solar radio emission and its source

This chapter in context

The major scientific thread connecting PS sources to the ASKAP-FLASH survey is the question of H I detection rates. Is H I absorption really more common towards these young radio AGN than towards the general population of radio galaxies (as discussed in Section 1.2.3), or is there some underlying selection effect? Detection rates are currently higher towards PS sources (Geréb et al., 2014; Maccagni et al., 2017; Yoon et al., 2025), but until the current generation of H I surveys, performed with the technically advanced SKA precursor instruments (particularly ASKAP-FLASH; Allison et al. 2022 and MALS; Deka et al. 2024), it was not possible to perform a truly unbiased search for H I absorption. Samples were pre-selected based on colour, radio luminosity, size, or some combination of the three, and it was incredibly difficult to disentangle the complex input of selection biases from the equally complex output of detection rates. These ongoing H I surveys are therefore the first opportunity we have to examine the detection rates in PS sources across unbiased samples.

Although these surveys are still ongoing, we need not wait for their completion to begin investigating the PS-H I connection. Even while these new observational samples are small and growing, we can begin to understand them with the help of mock observations. Such mock observations are typically extracted from (magneto)hydrodynamic simulations by modelling the electromagnetic emissions of a given region and optionally folding in instrumental effects in order to create datasets that are directly comparable with observations (Haworth et al., 2018).

In this chapter, we use mock observations towards 'radio AGN' in the Simba cosmological simulations (Davé et al., 2019), to understand the impact of geometry on our detection rates of associated H I absorption in ASKAP-FLASH. Simba provides jet-like models for AGN feedback, and there is a strong precedent for its use in mock observations of radio AGN (Thomas et al., 2021; Thomas & Davé, 2022; Thomas et al., 2025), so it seemed a natural choice to extend to mock observations of 21 cm absorption. However, in this work we intentionally decouple our radio jets from all but the basic requirement that their 'host' galaxies are H I-rich, and currently host an AGN. This allows us to test whether we can reach the high ASKAP-FLASH detection fraction towards PS-type AGN with geometric arguments alone. If we cannot, it suggests that not only are PS sources embedded in a gas-rich medium, but that this medium must be *uniquely* gas rich due to some jet-gas relationship.

Statement of contribution

This chapter faithfully reproduces material from an article submitted to the *Astrophysical Journal* for publication, with minor stylistic adaptations to fit the thesis style. I am first author, with Samuel Ward, Stephanie Tonnesen, Vanessa Moss and Elaine Sadler as co-authors. The research problem was conceived of by me in collaboration with S. Ward, as part of an application to the Pre-Doctoral programme at Centre for Computational Astrophysics within the Flatiron Institute. The mock observation package used in this work, SANGRiA, was co-developed by S. Ward and myself, and analysis for this paper was carried out by me with close supervision from S. Ward. S. Tonnesen performed the morphological classifications, and provided guidance and feedback throughout both the code development and analysis. I wrote the majority of the paper, but S. Ward contributed text to sections 4.2.1, 4.2.4, 4.4.6, 4.6, and wrote Section 4.2.3. S. Tonnesen contributed text for sections 4.4, 4.6, and 4.7. Figures 4.1, 4.14, 4.15, and 4.16 were provided by S. Ward. All other portions of the paper are my own work, with co-authors providing comments and feedback throughout the preparation of the manuscript.

4.1 Forward Modelling ASKAP-FLASH detections: Abstract

The 21 cm hydrogen line is an important probe of the cool, neutral medium in and around galaxies, which is crucial to galaxy formation and evolution. However, a number of factors influence the detectability of H I in absorption beyond just the total gas mass of the system. We investigate here two such parameters: galaxy geometry and radio morphology, using a suite of mock observations produced with the package SANGRIA (Simulating Absorption of Neutral Gas for Radio Astronomy), with galaxy data drawn from the SIMBA cosmological simulations. These mocks are matched to the observational parameters of the First Large Absorption Survey in HI (ASKAP-FLASH), an all-sky, untargeted search for H I absorption currently being conducted with the Australian SKA Pathfinder. We focus on instances of associated absorption, where the continuum backlight is produced by the jets of an active galactic nucleus associated with the same galaxy hosting the H I, so that we can investigate the apparent over-detection of H I towards galaxies with compact radio jets ($\lesssim 1$ kpc) in ASKAP-FLASH. We find that, although simple geometric arguments can explain some of the preferential detection towards compact jets, they are not sufficient to explain the high detection rates seen in the ASKAP-FLASH Pilot Surveys. We interpret this discrepancy between the geometrically-driven detection rate in our mocks and the observed detection rate in ASKAP-FLASH to be evidence that the most compact AGN jets reside in uniquely gas-rich environments.

4.2 Introduction

Neutral atomic hydrogen (H I) is a fundamental ingredient in star formation and galaxy evolution, and is crucial to understanding the cosmic baryon cycle (Péroux & Howk, 2020). Fortunately, it is detectable in absorption out to arbitrarily high redshifts (provided the radio continuum backlight is sufficiently bright), making it an excellent tracer of how the interstellar medium (ISM) evolves over time (Morganti & Oosterloo, 2018). For this reason, the First Large Absorption Survey in H I was born (ASKAP-FLASH; Allison et al. 2022; Yoon et al. 2025), to trace the evolution of H I in the under-studied period after cosmic noon ($z = 0.4 - 1$) where a marked decline in star formation rate density is observed, likely caused by depleted cold gas reservoirs coupled with feedback mechanisms (Madau & Dickinson, 2014).

While the ASKAP-FLASH survey is ongoing, a trend is emerging amongst H I-detected sources, in which there are an over-representation of compact "Peaked Spectrum" (PS) radio AGN compared to their number amongst the general radio population (Kerrison et al., 2024a; Yoon et al., 2025). This PS classification is based on the fact that the radio emission in these objects peaks at some characteristic frequency (proportional to the size of the jet), before turning over at low frequencies due to absorption losses. PS sources are thus believed to be young or confined AGN, with their radio jets extending out to only a few kiloparsecs, making them intimately connected with the ISM of the host galaxy (O'Dea & Saikia, 2021). Already in targeted searches for H I absorption it has been noted that PS sources exhibit a high detection rate of H I (e.g. Conway, 1997; Geréb et al., 2014), but ASKAP-FLASH is the first time such a skew has been seen in a completely untargeted search of extragalactic radio sources. This presents an interesting puzzle; are these recently-triggered radio AGN inherently more likely to host the cool, neutral medium detectable in H I absorption, or is this trend simply the product of a geometrical selection effect related to their extreme compactness? One avenue for investigating these questions is to use mock observations of simulated galaxies, where the underlying distribution of the gas is known exactly.

Mock observations, where simulated datasets are post-processed to produce observable quantities, are a well-established tool for exploring the complexities of galaxy evolution and dynamics. At radio wavelengths, mocks of H I 21 cm emission maps have been used to study resolved rotation curves (Roper et al., 2023) and line profile asymmetries (Glowacki et al., 2022), drawing data from the SIMBA cosmological simulations (Davé et al., 2019). Other studies have used post-processing to examine the global properties of 21 cm emission in IllustrisTNG (Stevens et al., 2019) and EAGLE (Crain et al., 2017), and El-Badry et al. (2018) demonstrated that gas kinematics could be derived even from unresolved H I profiles using the FIRE simulations (Hopkins et al., 2018). At even higher resolution, Booth & Theuns (2007) studied 21 cm emission mocks from idealised simulations to understand the role of H I in galactic fountains. Sinigaglia et al. (2022) similarly used mock observations to better understand the systematics involved in spectral stacking, a technique used to recover statistical measures of H I-mass in large samples of galaxies. However, these H I studies remain focused on emission in the local Universe. At much higher redshifts, Geil & Wyithe (2008) used a semi-numerical approach

to understand how sensitive the current generation of low frequency instruments might be to H I emission at the epoch of reionisation, and others have been working on mock catalogues and tools to predict the detectability of H I with the SKA (e.g. Obreschkow & Meyer, 2014; Elson et al., 2016). Still, all the work in this space has focused heavily on H I emission, primarily at low $z \lesssim 0.2$ and high $z \gtrsim 6$ redshift. There remains a large fraction of cosmic history for which we have no H I mocks because we have no observational data to compare them to; it is observationally difficult (indeed, impossible with current instruments) to detect H I in emission at these redshifts. However, absorption studies are currently able to observe H I out to $z = 3.5$ (Aditya et al., 2021), and a number of surveys have been designed to search for H I in this way towards lists of pre-defined targets spanning various slices of redshift space including $0.7 < z < 1$ (with 29 targets; Murthy et al. 2022), $0 < z < 2$ (with 391 primary targets; Gupta et al. 2016), $0 < z < 2.74$ (with 260 targets; Grasha et al. 2020), and $0.8 < z < 3.5$ (with 22 targets; Kanekar et al. 2014). ASKAP-FLASH is the only existing survey to search for H I absorption in an untargeted manner towards *all* radio continuum sources (above a flux density limit $S \geq 30$ mJy) across a large ($24,000\,\mathrm{deg}^2$) region of sky and a wide redshift range ($0.4 < z < 1.0$). It is for this reason that we now have suitable datasets to compare with mock observations of H I absorption at cosmic noon.

In this paper, we present a first analysis of neutral hydrogen probed in absorption by the 21 cm line in a cosmological simulation. Using the Simulated Absorption of Neutral Gas for Radio Astronomy package (SANGRIA; Ward, Kerrison & Tonnesen *in prep.*), we inject radio jets into galaxies from SIMBA which serve as a backlight for the hydrogen permeating the galaxy (Section 4.3). We develop three, physically-motivated jet models (Sections 4.3.2 & 4.4.3) which are injected into galaxies in order to create mock spectral line observations (Section 4.3.4). Our suite of mock observations is drawn from a representative sample of gas-rich galaxies originating in a single SIMBA snapshot (Section 4.4). With this suite we test the effect of varying the radio jet parameters (Sections 4.5.1 & 4.5.2), the orientation of the host galaxies (Section 4.5.3), and the orientation of the radio jets (Section 4.5.4) using a uniform parameter search. We also consider the effect of galaxy morphology on our ability to detect H I (Section 4.5.5). Finally, we re-weight our uniform parameter search to better match the observed luminosity function of radio galaxies, in order to determine how much of the high detection rate towards compact radio AGN can be attributed to geometry alone (Section 4.6). We discuss our results and present our conclusions in Sections 4.7 & 4.8.

Throughout this work we assume a ΛCDM cosmology with parameters of $H_0 = 68\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$ and $\Omega_\mathrm{b} = 0.048$, following Planck Collaboration XIII (2016).

4.3 Radio observations of cosmological simulations

In this section, we outline the basic methodology behind the SANGRIA pipeline for producing mock H I absorption spectra. Further details will be described in a forthcoming companion paper (Ward, Kerrison & Tonnesen *in prep.*). Figure 4.1 shows an overview of the pipeline – on the left we show the gas particle data from a SIMBA galaxy (see Section 4.3.1) and an injected FRII jet model (Section 4.3.2). The middle panels show mock observations of this system (Sections 4.3.3 & 4.3.4), showing both a FWHM = 30″ beam matched to the parameters of the ASKAP-FLASH survey (bottom), and a high resolution beam of FWHM = 2″ (top) for illustrative purposes. The right column shows the radio spectrum around the 21cm H I line extracted from the brightest pixel. The following sections will now describe these steps in more detail.

4.3.1 The SIMBA simulations

The parameters of the ASKAP-FLASH survey place constraints on many of the characteristics required of our chosen simulation; it must have global gas masses well-matched to observations not only in the local Universe, but at intermediate redshifts ($0.4 < z < 1.0$), it should implement some well-motivated AGN formation, growth and feedback prescription (in order to produce galaxies likely to host radio jets), and it should contain a diverse population of galaxies including both mergers and isolated systems.

We construct our mock observations from the SIMBA[1] cosmological simulations (Davé et al., 2019), run with the meshless finite mass hydrodynamics solver from GIZMO (Hopkins, 2015). We use the fiducial $m100n1024$ run, which has a comoving box size of $L = 147$ cMpc and a baryonic mass resolution of $m_\mathrm{b} = 1.83 \times 10^7\,\mathrm{M}_\odot$.

[1] http://simba.roe.ac.uk

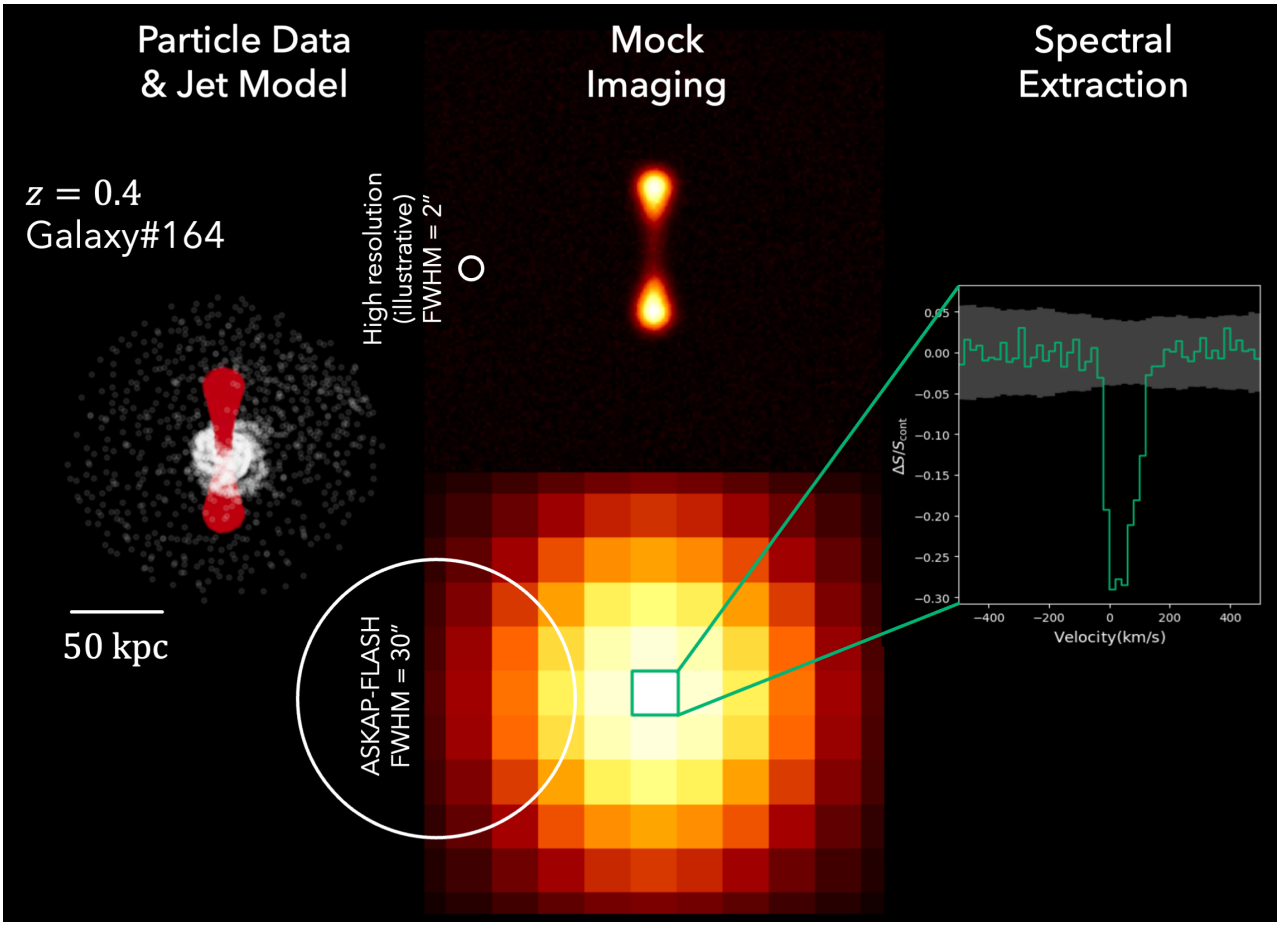


Figure 4.1: An outline of the SANGRiA pipeline, translating particle and ray data into mock H I spectra. *Left:* the Simba particle data (white points) from Galaxy #164 and our injected jet model (red points). *Middle:* mock radio observations of this system after beam convolution. We show two beam sizes: *top* an illustrative high resolution FWHM = 2″ beam and *bottom* the ASKAP-FLASH beam (30″). *Right:* the H I spectrum extracted from the brightest pixel of the ASKAP-FLASH spectral cube, showing a clear 21cm absorption line.

Gas cooling and photoionization heating are computed using the Grackle-3.1 library (Smith et al., 2017), assuming a spatially uniform UV background and including self-shielding following Rahmati et al. (2013). As cosmological simulations do not resolve the cold neutral medium, the neutral and molecular hydrogen fractions are estimated using subgrid prescriptions (Krumholz & Gnedin, 2011). The resulting H I and H_2 masses are calculated on-the-fly for each gas particle and stored in the snapshots, which we use directly in our ray-tracing procedure, as explained in the next section.

These simulations include a bi-directional, hydrodynamically-decoupled outflow model at low Eddington ratios ($f_{\rm Edd} < 0.2$) to approximate 'jet-mode' AGN feedback, and per-particle H I masses can be read directly from Simba snapshot files. Previous work has also shown that with sensible selection criteria and prescriptions for radio luminosity, Simba samples can be well-matched to observational samples of radio galaxies in the local Universe (Thomas et al., 2021; Thomas & Davé, 2022; Thomas et al., 2025), making an apples-to-apples comparison with ASKAP-FLASH data readily achievable using these simulations.

Other previous work has introduced mock observation pipelines for studying neutral hydrogen in emission (e.g. Martini, Oman 2019; Oman et al. 2019; Oman 2024). However, to study H I in absorption, additional models are needed for the background emission and radiative transfer. Since Simba does not directly model the luminosity or extent of the radio jets produced by the AGN, we must inject jets of appropriate size and luminosity before we can perform any analysis of 21 cm absorption. Motivated by this need for simple, flexible radio jet models within a larger mock observing framework, we have created Sangria (Simulating Absorption of Neutral Gas for Radio Astronomy), a robust end-to-end pipeline for producing mock 21 cm absorption line observations. While our software package has been built for flexibility and will be described in detail in Ward,

Kerrison & Tonnesen *(in prep)*, here we focus on one use-case: H I absorption against radio jets embedded in the hydrogen gas, matched to the parameters of the ASKAP-FLASH survey. In the following subsections we outline the methods for incorporating radio jets into the simulation, and then for extracting mock spectral line profiles from the associated galaxies.

We chose to use Simba in this study for the reasons outlined above. In principle though, any hydrodynamic galaxy simulation could be used by the Sangria pipeline that has saved per-particle H I masses, or which can be post-processed to obtain them. Expanding this study to other simulations is left for future work.

4.3.2 Radio Jets with Sangria

For an individual galaxy extracted from a Simba box, Sangria can flexibly insert one of three different jet morphologies at any arbitrary orientation and position relative to that galaxy, or the user may choose a point-like 'quasar' source of radio emission. For this work, we fix the origin of our jets at the centre of the galaxy, defined as the local minimum in the galactic potential, and for simplicity we only rotate the jet structure along a plane through each galaxy rather than in any arbitrary direction. The jets themselves can take one of three functional forms, depicted in Figure 4.2.

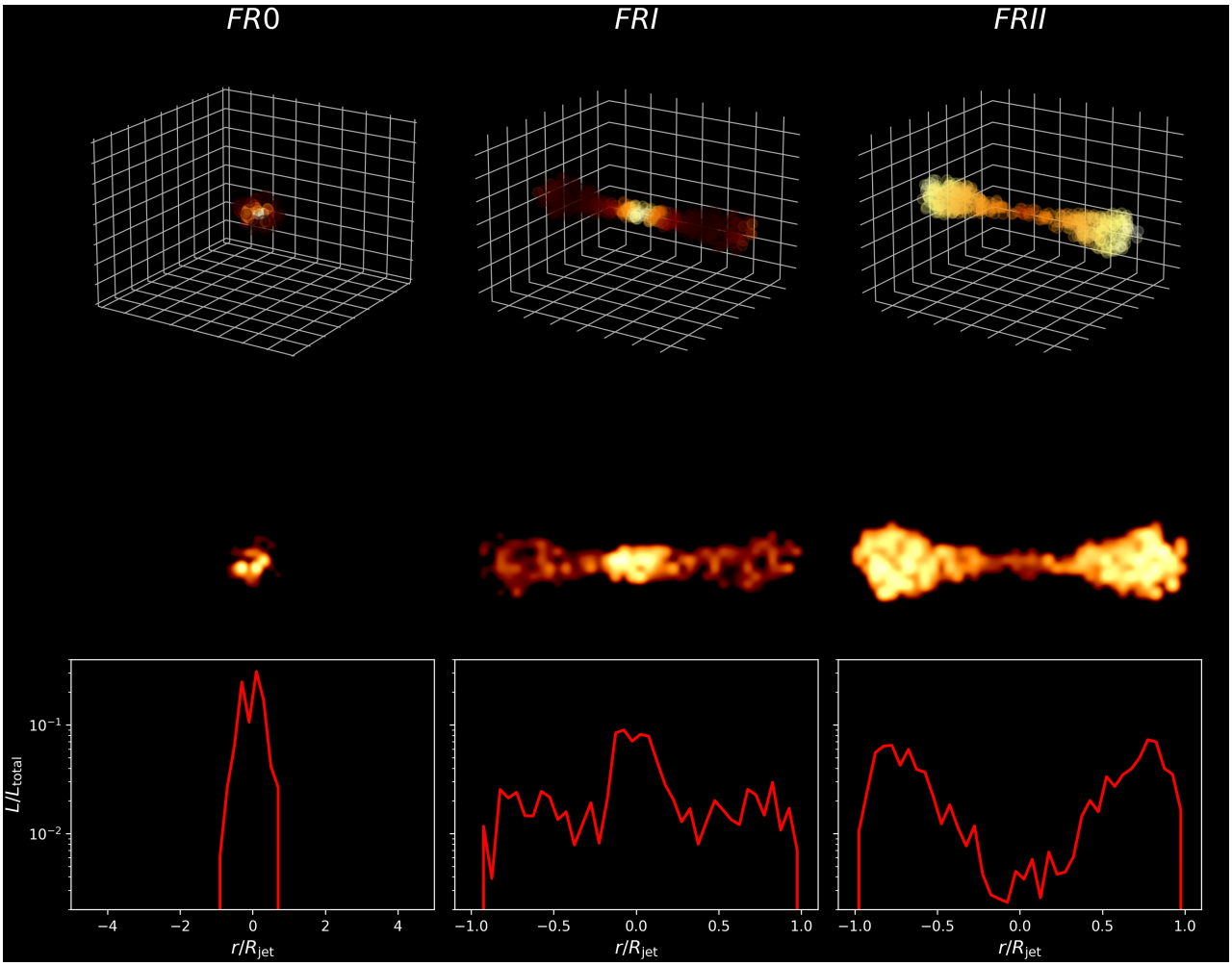


Figure 4.2: The luminosity distribution for the three jet types used in this work. These map onto the observational classes of FR0 or compact (left), FR-I or core-dominated (centre) and FR-II or lobe-dominated (right). The sizes and luminosities of these jet models are scaled to the required values before they are added into the Simba particle data and used as the backlight for the absorption ray tracing.

These models were chosen to map onto the main morphological classes of observed radio jets. The centre panel represents so-called 'core-dominated' radio sources with prominent jets that fade as a function of distance from the core, while the right panel represents 'lobe-dominated' emissions seen in large radio galaxies like Cygnus A. These are typically referred to as FR-I and FR-II galaxies respectively, after the authors of the original paper highlighting the dichotomy (Fanaroff & Riley, 1974). Radio 'emission' is distributed along these extended jet structures according to a radial power law $L(r) \propto r^a$, where $a = 0.8$ for FRI-type jets, and $a = 2.0$ for FRII. These values for a were manually selected to produced realistic-looking jets over a range of lengths and bolometric luminosities. We were unable to tie our choices to stronger observational constraints in this case, since the average radial emission profile of jets has only been measured observationally on sub-parsec scales, which are unresolved in our setup (e.g. Kravchenko et al., 2025). However, the seminal theoretical paper on AGN jet structures also assumes a power law scaling of this kind (Blandford & Königl, 1979).

Our third model represents the more recent FR-0 classification describing extremely compact radio AGN (Baldi et al., 2015), for which we use the model in the left-hand panel of Figure 4.2. Very Long Baseline Interferometry shows that many FR-0 sources are unresolved even on milliarcsecond scales (Baldi et al., 2025). Given that even in the local $z \leq 0.1$ Universe this corresponds to an angular scale of less than 1 arcsecond, these will be entirely unresolved in our mock images with matched ASKAP-FLASH resolutions. A simple, spherical emission model like that chosen here is therefore an appropriate representation of these most compact jets.

Both the FR-I and FR-II models are formed from bi-conical jets centred at the origin, while the FR-0 is a sphere. Within the three dimensional jet structure, the luminosity is discretised into individual rays distributed efficiently across the jets using a glass packing algorithm, which also allows us to emulate some of the irregularities of jet structures seen observationally as hot spots and knots (e.g. Miley, 1980). Our results are insensitive to the number of jet points used, as long as the local number density of jet particles is greater than the local number density of gas particles to ensure we do not undersample the gas, especially in the higher density disk.

4.3.3 Ray tracing

Each jet particle is used as the source point for ray casting. This allows each ray to begin at an arbitrary point in the box and with an arbitrary background luminosity ($L_{\rm bkg}$). The rays are then propagated in parallel towards the observer.

The rays are split into small bins along the line-of-sight, and the quantities from nearby Simba gas particles are averaged into each bin. The width of these bins is fixed to be smaller than the distance between neighbouring particles in the highest-density regions. Quantities are weighted both by their distance from the ray bin (using a cubic spline kernel and the smoothing length of each particle) and their H I mass.

The frequency-dependent optical depth for each ray is then calculated by integrating along the line of sight:

$$\tau_\nu = \sigma \phi_\nu \left(v_{\rm los}, T\right) \int n_{\rm HI} ds \tag{4.1}$$

where σ is the cross-section from the hyperfine 21cm transition, and the shape of the line is modelled as a Voigt function (ϕ_ν), which captures both Doppler and natural line broadening, and encodes the line-of-sight velocity ($v_{\rm los}$) and temperature (T) information of the particles along the ray. When calculating the cross section we must assume a spin temperature, which encodes the distribution of hydrogen atoms across the energy levels involved in the 21 cm spin-flip transition (Draine, 2011). We fix this to $T_s = 300$ K, the value for the Milky Way.

The frequency-dependent luminosity is then given by attenuating the background luminosity from the jet particle by the optical depth:

$$L\left(\nu\right) = L_{\rm bkg} e^{-\tau_\nu} \tag{4.2}$$

to give the shape of the absorption line for each ray.

When calculating the broadening of the line profile, we use the gas particles' kinetic temperature (T_k) as

given by Simba. An alternative, observationally motivated choice might be to use a spin temperature (T_s). However, T_s is not provided in the Simba particle data catalogues, and it likely varies with redshift and galaxy properties (Allison et al., 2022). Furthermore, we found that using an assumed, Milky Way like spin temperature (T_s = 300K) produced line widths much narrower than observed, whereas the kinetic temperature ($T_k \approx 10^{4-5}$K for gas particles with high H I) better matched realistic line profiles. This could be because using T_s neglects any gas turbulence that is unresolved by Simba (sub-kpc scales). However, assuming the kinetic temperature for the H I gas is a good estimate of the combined thermal and turbulent broadening.

4.3.4 Mock observations

After the ray tracing step, we now have a collection of rays with arbitrary (x, y) coordinates (having been integrated along the z-axis) and H I -attenuated luminosity profiles (L_ν). In this section, we describe how to transform this into an observed radio spectrum. In this paper, we use ASKAP-FLASH as our survey target (Allison et al., 2022; Yoon et al., 2025) and match the observational parameters to those used in the Pilot Survey. However, we note that our pipeline can be applied to any instrument or project that can target the 21 cm line.

Motivated by this survey choice, we fix our galaxies at $z = 0.4$ in this work to match the lower bound of the region in which ASKAP-FLASH is sensitive to H I absorption. The radio luminosity assigned to each ray is then converted into observer units of flux density:

$$S_{\nu,\mathrm{obs}} = \frac{L_\nu}{4\pi D_L^2}(1+z)^{1+\alpha} \tag{4.3}$$

where L_ν is the radio luminosity of the source (in W Hz^{-1}), $S_{\nu,\mathrm{obs}}$ is the observed flux density (in Jy), D_L is the luminosity distance in parsecs, and α is the spectral index of the source, which we fix at a fiducial value of $\alpha = 0.7$. The $(1+z)^{1+\alpha}$ term combines a relativistic correction and a k-correction for the spectral bandwidth.

The rays are then binned into pixels corresponding to an angular scale of 5 arcseconds (Fig. 4.1, second column). This pixelated image is then convolved with a 30 arcsecond Gaussian beam to mimic the observation with the ASKAP telescope (third column). As long as the beam is well-sampled ($\gtrsim$ 6 pixels/beam) the pixel scale is unimportant. Gaussian noise of σ_{noise} = 5 mJy/beam/channel is added to the data to match the per-channel sensitivity of the ASKAP-FLASH survey (Yoon et al., 2025).

For many combinations of jet and orientation, this leads to unresolved, point-like sources in mock images at $z = 0.4$, as can be seen in Figure 4.1. This agrees well with observations, since around 90 per cent of FLASH continuum sources are expected to be compact and unresolved in ASKAP images.

A spectrum is extracted from the cube towards only the brightest pixel in the image (last column), and with a spectral resolution of 18.5 kHz (corresponding to $\sim$ 6 km s^{-1} at $z = 0.4$). Although ASKAP-FLASH spectra use the full 288 MHz bandwidth of the instrument to search the $z = 0.4 - 1$ redshift range, this is not necessary for our mock spectra. Even extreme outflows of 1000 km/s or more would have a width of no more than $2 - 3$ MHz at our fiducial redshift. We therefore restrict the spectrum to a 6 MHz chunk, centred on the bulk velocity of the galaxy. Alongside the spectrum, we generate a channelised 'noise spectrum' containing Gaussian noise at the same level as that seen in the ASKAP-FLASH pilot surveys (5 mJy/beam per spectral channel).

The spectrum and associated noise are then fed to the same Bayesian linefinder used on ASKAP-FLASH observations (Allison et al., 2012). This searches for the presence of a Gaussian absorption line profile anywhere in the spectrum, and once found, it returns the parameters of the best fit model, as well as the likelihood of the line model over a pure noise model, expressed as a Bayes factor. Following the ASKAP-FLASH Pilot Surveys, we consider detections with a Bayes factor $\ln(B) \geq 30$ to be statistically significant (Yoon et al., 2025).

From the parameters of the FLASHfinder model, an observed column density of neutral Hydrogen can be derived as follows:

$$N_{HI} = 1.823 \times 10^{18} \frac{T_s}{f} \int \tau(\nu)\, d\nu \tag{4.4}$$

where T_s is the spin temperature of the gas, and f is the covering fraction of the radio source which is subtended by the absorbing gas.

For observers, assumptions need to be made about both T_s and f in order to calculate N_{HI}, which is taken as an estimate of the true column. Although we can directly measure the covering fraction f in our simulations, we too need to make assumptions about the spin temperature, which is not the same as the kinetic temperature of the gas recorded in the Simba particle datasets. In order to compare our simulated H I absorption features to the observations of ASKAP-FLASH as fairly as possible, in this work we therefore rely on direct comparison of the optical depth τ, for which we obtain measures of width (Δv), depth or 'peak optical depth' (τ_{peak}) and overall strength or 'integrated optical depth' (τ_{int}) directly from the FLASHfinder. Ward, Kerrison & Tonnesen *(in prep.)* will discuss measures of the column N_{HI} in more detail.

4.4 Sample Construction

Our main motivation is to determine under what conditions H I is detectable given the ASKAP-FLASH survey parameters. Accordingly, we draw our simulated galaxies from a Simba snapshot at $z = 0.4$ (snapshot #151), corresponding to the lower bound of the redshift space in which ASKAP-FLASH is sensitive to H I. Each galaxy is cut out from the box using a sphere of 100 kpc radius, and aligned by default so that its average angular momentum vector is vertical and perpendicular to the observed line of sight. This alignment is achieved using the `L_align` function also used by MARTINI.[2] We outline below additional properties of this sample as relevant to our investigation.

4.4.1 Simba galaxy selection

From the 46,575 galaxies within the Simba snapshot, we focus on 266 which are both likely to host a (radio) AGN, and which are H I-rich. In this way, we are testing the *detectability* of H I in ASKAP-FLASH-like mocks. This sample, and our selection process, are described in more detail below.

Thomas et al. (2021) defined radio AGN within Simba as those galaxies with massive, central black holes ($M_{\rm BH} > 10^8\,{\rm M}_*$) and full-power kinetic jets ($0 < f_{\rm Edd} < 0.02$). We follow this prescription, but we relax the requirement on $f_{\rm Edd}$ here to allow for any amount and type of feedback (Davé et al., 2019). This choice was made because observationally, ASKAP-FLASH detections are made towards both quasar-type AGN, and AGN with extremely compact ($\lesssim$ 1 kpc) jets, potentially indicative of AGN transitioning between 'radiative-mode' and 'jet-mode' accretion as implemented in Simba.

In summary, our selection criteria were as follows:

1. Galaxy is H I-rich : $M_{\rm HI,30\,kpc} > 10^{9.5}\,{\rm M}_\odot$
2. Galaxy is massive : $M_* > 10^{10.5}\,{\rm M}_\odot$
3. Galaxy hosts AGN : $M_{\rm BH} > 10^8\,{\rm M}_\odot$ and $\dot{M}_{\rm BH} > 0$

We show our sample in Figure 4.3, where the shaded box indicates the cuts that were performed to ensure the selected galaxies are both H I-rich (1) and massive (2). This figure also motivates our threshold for H I-rich, since a cut-off of $M_{\rm HI,30\,kpc} > 10^{9.5}\,{\rm M}_\odot$ selects the high-mass mode of the bimodal distribution seen in the AGN kernel desnity estimate (dotted line, y-axis). The threshold in stellar mass ensures that our galaxies have not undergone any significant mass loss events which would put their stellar mass below the black hole seeding threshold ($M_* \approx 10^{9.5}\,{\rm M}_\odot$), but is a more conservative limit than that imposed by Thomas et al. (2021). The scattered points coloured by $f_{\rm Edd}$ in Figure 4.3 indicate those galaxies hosting AGN (3), and the kernel density estimates along each axis show the distribution of each sub-population.

4.4.2 Sample morphologies

Our full sample of 266 galaxies were visually classified based on their morphology in the $z = 0.4$ Simba snapshot. Of these, $\sim$ 30 per cent (107/266) have a regular gas disk structure, $\sim$ 11 per cent (30/266) show spiral structure, and $\sim$ 16 per cent (43/266) have a spherical, bulge-like gas distribution. A further $\sim$ 15 per

[2]Found at `https://github.com/kyleaoman/kyleaoman_utilities`

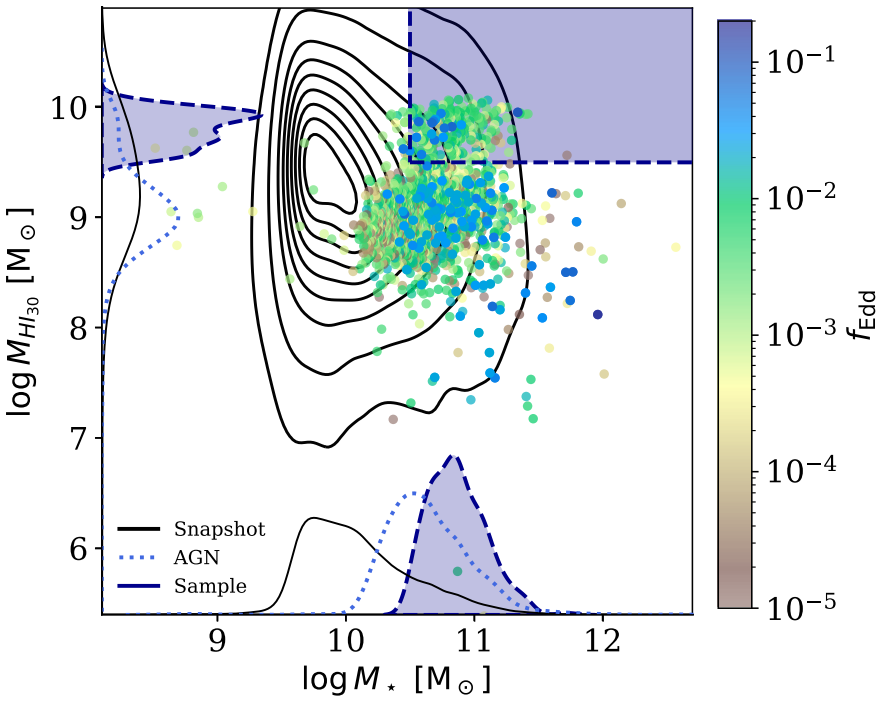


Figure 4.3: The parameter space used for selecting our Simba galaxies. We selected a sample of HI-rich galaxies with moderate to high stellar masses, and which span a representative range of f_{Edd} values. Contours are all galaxies within the Simba snapshot, scattered points are those which are considered AGN using the criteria outlined in Section 4.4.1, and the shaded box indicates the selection criteria used for our sample. Along the $x-$ and $y-$axes are kernel density estimates showing the distribution of each sub population.

cent (41/266) are strongly interacting (potentially merging) with a nearby ($d < 100$ kpc) companion, and ~ 11 per cent (31/266) have a disturbed gas profile which may indicate interactions with a more distant companion not visible in our cutouts. We combine these last two categories, of observed and potential interactions into one category of 'mergers and rogues' in Section 4.5.5. A representative subset of the full sample is shown in Figure 4.4, to give the reader a sense of their diverse (gaseous) morphologies, where it is clear that some (e.g. #446, #1958) have the majority of their H I centrally concentrated within $r \lesssim 20$ kpc, while others show H I structure on larger scales, up to $\lesssim 100$ kpc.

What does this mean in the context of observational samples? Optical imaging of ASKAP-FLASH sources reveals that 18% of the Pilot Survey H I detections have at least one companion within a projected distance of 20 kpc on sky at the redshift of the H I feature (Sadler et al. *in prep*). In addition to this, some targeted searches from the literature seem to suggest that mergers comprise a significant fraction of HI-detected sources. Dutta et al. (2018) showed that in a sample of mergers, the detection rate is much higher than in other samples (83 ± 17 per cent), and that the absorption lines themselves are much stronger in these systems. By contrast, the optically-selected sample of Geréb et al. (2015) contained only 5/32 mergers, with a further 7/32 with extended optical morphology, and the remainder appearing compact. It is likely that optical selection criteria used for many previous H I searches will influence these fractions, for example by missing strongly dust-obscured companions (Langan et al., 2024). Furthermore, the fraction of mergers in an H I-selected sample is not a fair comparison to this set of simulated galaxies because we have not yet determined which of these we would see in H I absorption, we can only say that we ought at least to include *some* mergers in our selection based on these observational results. Amongst the more general population of powerful (jetted) radio galaxies, Chiaberge et al. (2015) show that a very high fraction (90 per cent) are associated with recent or ongoing merger activity. The fraction of non-breakout (PS) radio AGN which show evidence of recent mergers or interactions is somewhat more modest, at $40 - 70$ per cent (O'Dea et al., 1996; Randall et al., 2011). In all cases, observers tend to classify these galaxies as 'interacting' if they show evidence of streams or tidal tails (even without a clear identification of the interacting companion), so a fairer comparison to our sample might be the ~ 30 per cent which we classify as mergers and rogues, as introduced above and discussed further in Section 4.5.5.

In summary, our sample shows a range of gas morphologies indicative of both isolated galaxies and merging/interacting systems that broadly reflects the diversity of galaxies hosting radio AGN, and detected in H I absorption.

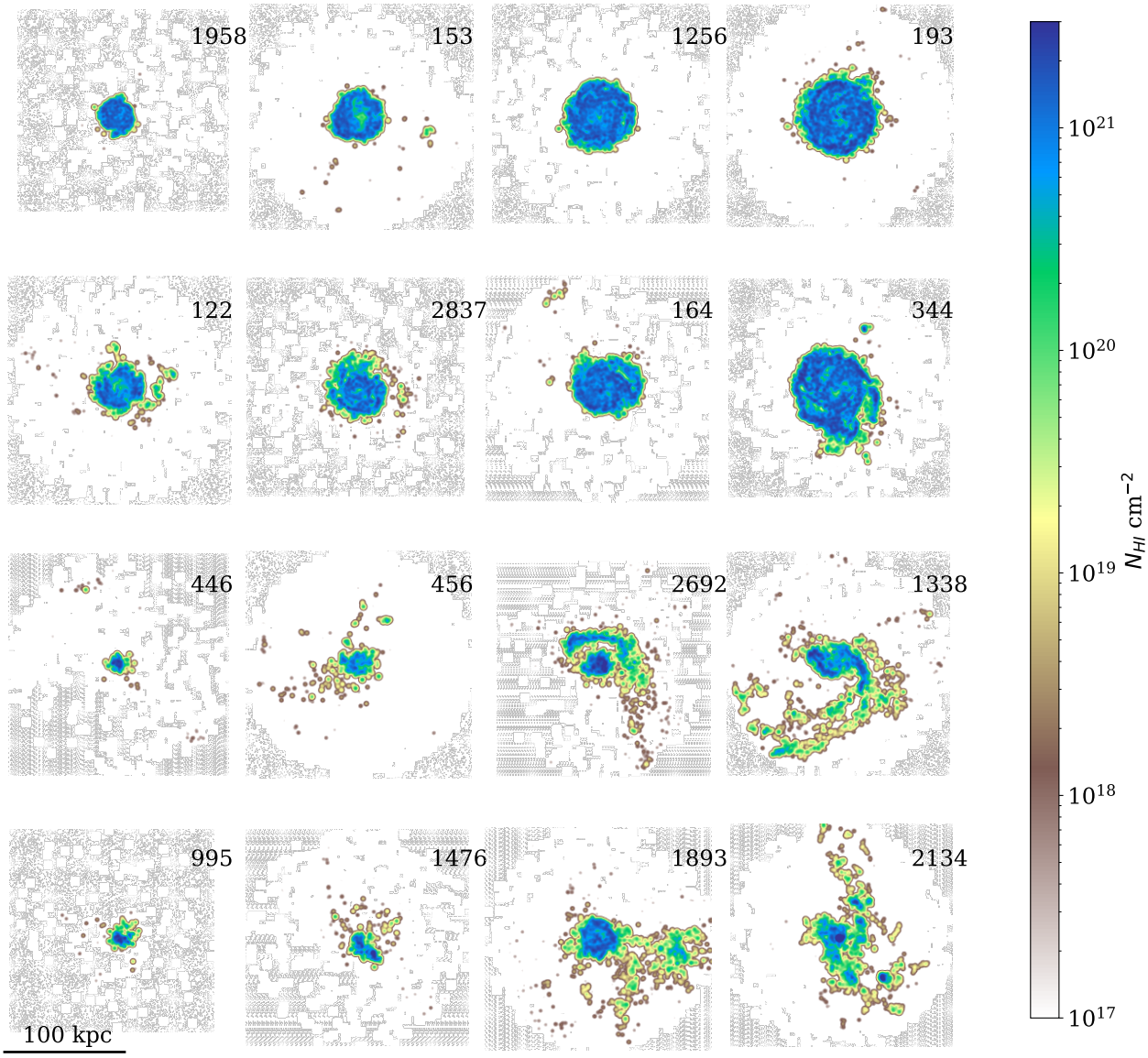


Figure 4.4: A selection of galaxies from our full sample, showcasing the range of galaxy morphologies. Ordered approximately from small to large (left to right) and from isolated disks to interacting systems (top to bottom) The colorscale indicates the column of H I. All galaxies are oriented such that their angular momentum vector (averaged over all star and gas particles) is oriented out of the page, such that disks are seen face-on. Note that galaxy 164 is often referred to in this paper as an example.

Rotation type	Jet Type	Jet Length [kpc]
Galaxy rotation (Gal.)	FR0	0.5
(jet fixed $\perp$ to galactic disk)	FRI	1.0, 5.0, 10.0
	FRII	5.0, 10.0, 50.0
Jet rotation (Jet)		
(galaxy inclined 60° to LoS)	FRI	1.0, 5.0
	FRII	5.0, 10.0, 50.0

Table 4.1: A summary of the jet-galaxy parameters used in this work. Each of these jet type/radius combinations is modelled with 17 luminosities regularly spaced in log between $L_{1.4\mathrm{GHz}}$=10^{24} W Hz^{-1} and $L_{1.4\mathrm{GHz}}$=10^{28} W Hz^{-1}.

4.4.3 Radio jet luminosity & morphology choices

As mentioned in Section 4.3.2, for each galaxy drawn from the Simba box we inject radio jets for which we have a number of free parameters including morphology, length and total luminosity. The luminosity of radio AGN is well characterised in the local Universe (Mauch & Sadler, 2007) and even out as far as $z \sim 1$ (Best et al., 2014; Hardcastle et al., 2025), the upper limit at which ASKAP-FLASH is sensitive to H I. In all cases there is an exponential decay in the high luminosity tail of radio power, so we use these observational constraints to place an upper limit of $L_{1.4\mathrm{GHz}}$=10^{28} W Hz^{-1} on the integrated luminosity of our radio jets. Although the radio luminosity function shows that there exists an abundance of AGN with jet luminosities $L_{1.4\mathrm{GHz}}\leq 10^{24}$ W Hz^{-1}, ASKAP-FLASH is only expected to be sensitive to H I towards the most powerful radio galaxies (Allison et al., 2022). Our pipeline independently confirms that, with ASKAP-FLASH survey parameters, we are not sensitive to H I in absorption with jets fainter than $L \sim 1 \times 10^{24}$W Hz^{-1}, so this was set as the floor in the integrated luminosity of our jet models for the remainder of this work.

As discussed in Section 4.3.2, with Sangria we allow for the injection of any arbitrary jet morphology and luminosity into our Simba galaxies. This is in contrast to Thomas et al. (2021) who explicitly link the radio luminosity of their AGN to the BH properties contained in the Simba catalogues. Our reasoning for decoupling the jets from the BH properties in this work is twofold; first, due to boxsize limitations the fiducial Simba run does not contain more than a handful of AGN with $L_{1.4\mathrm{GHz}}\geq 10^{27}$ W Hz^{-1} using the Thomas et al. (2021) prescription. Since these powerful jets are preferentially where ASKAP-FLASH detects H I (recalling that it is a wide and shallow survey), using BH-linked radio luminosities would lead to a prohibitively small sample of FLASH-like H I detections. Secondly, as stated elsewhere throughout this paper, by decoupling the jet morphology and luminosity from the BH (and galaxy) properties, we can perform a test of geometric affects in isolation. For the remainder of this work we thus restrict the injected radio luminosities to $10^{24} \leq L \leq 10^{28}$ W Hz^{-1}, over which range we will perform a uniform parameter study in Section 4.5, before more closely matching observed distributions in Section 4.6.

A summary of the different combinations of jet and galaxy parameters used in the remainder of this work is given in Table 4.1. Thus, in the following section when we refer to the "full sample" we are referring to all jet types, radii, and luminosity combinations for every galaxy. The 'Gal.' sample corresponds to those mocks where the jet and galaxy disk are fixed perpendicular to each other, and rotated through 12 viewing angles (the first row in Table 4.1). The 'Jet' sample is those mocks where the galaxy is fixed at 60 degrees to the line of sight, and the jet is rotated through 12 viewing angles (the second row in Table 4.1).

4.5 Testing Detectability in Mock Observations

With our parameters established in previous sections, we here perform a uniform parameter sweep to explore the conditions under which H I is detected in absorption. In particular, we focus on whether the jet type and luminosity, or the geometrical orientation of the jet and galaxy effect whether ASKAP-FLASH would detect an absorption line (4.5.1-4.5.4). We then consider the effect of galaxy morphology (4.5.5), before ending this section with a discussion of overall detection fractions (4.5.6).

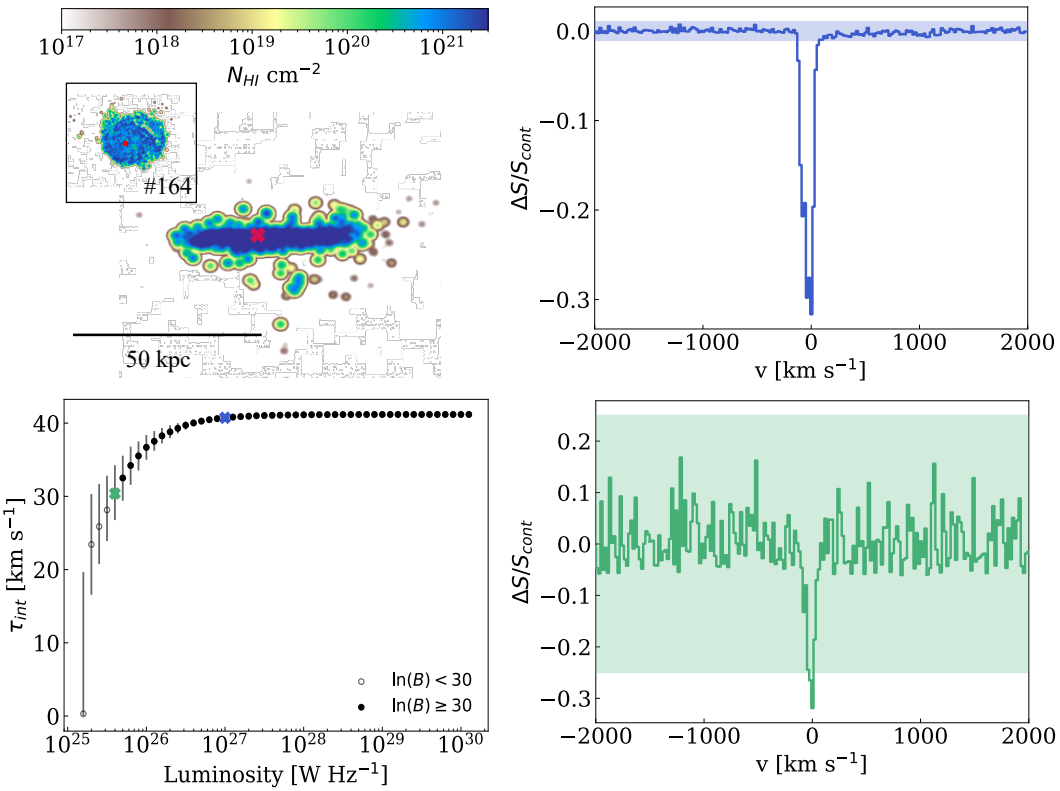


Figure 4.5: *Left, top*: The arrangement of Simba galaxy #164 used to test the effect of radio jet luminosity, aligned with the disk perpendicular to line of sight with an injected quasar (point source) of radio emission at the centre (red cross). *Left, bottom*: The integrated optical depth (i.e. strength) of the H I absorption feature as a function of radio luminosity, detected for this setup using the FLASHfinder. Filled circles indicate a significant detection ($\ln(B) \geq 30$). Two points are marked by crosses for which spectra have been extracted. *Right*: Two mock H I spectra extracted from the setup at a luminosity $L_{1.4} = 1.00 \times 10^{27}$ W Hz^{-1} (top), and $L_{1.4} = 3.98 \times 10^{25}$ W Hz^{-1} (bottom). The shaded region in each of these plots indicates 5× the root mean square noise across the spectrum.

4.5.1 The effect of jet luminosity

As part of jet injection, we specify the integrated radio luminosity at 1.4 GHz in the source restframe ($L_{1.4}$), decoupled from the black hole mass and accretion rate, as discussed in Section 4.4.3.

To show the effect of luminosity on the detection of H I in absorption, we examine a simple disk galaxy (Simba galaxy #164) oriented perpendicular to the line of sight, and with an injected quasar (point-like radio emission) at its centre, as shown in Figure 4.5. We vary the luminosity of this 'quasar' within the range $L_{1.4\mathrm{GHz}} \sim 10^{24} - 10^{30}$ W Hz^{-1} (larger than the range of our "full sample" to determine the thresholds for non-detection and saturation), creating a mock FLASH spectrum for each luminosity step, which is then fed to FLASHfinder to identify absorption signatures. We see that, although FLASHfinder detects absorption at just a few $\times 10^{25}$ W Hz^{-1} (where detection here means simply that $\tau_{int} > 0$), the integrated optical depth τ_{int} of the detected line is low with a high degree of uncertainty and the Bayes factor does not reach the threshold for significance ($\ln(B) \geq 30$). For reference, we add two crosses to this plot over luminosities for which we show the extracted spectra on the right of Figure 4.5. The first, shown in the lower right panel, is at the lowest luminosity for which the detection of H I would be considered significant by the FLASH team, with a Bayes factor $\ln(B) \geq 30$, corresponding to an injected $L_{1.4\mathrm{GHz}} = 3.98 \times 10^{25}$ W Hz^{-1} 'quasar'. The detected features are evidently noisy in this spectrum, given the large uncertainty on τ_{int}, although it contains the same structures at approximately the same depth as the spectrum towards the blazing $L_{1.4\mathrm{GHz}} = 1 \times 10^{27}$ W Hz^{-1} 'quasar', with a Bayes factor two orders of magnitude larger (Figure 4.5, upper right). It is only at this second, higher luminosity where the optical depth approaches the limiting value of 40 km s^{-1}. This limiting optical depth corresponds to

the H I column measured from the particle data if we fix the spin temperature to $T_s = 324$ (K) in our conversion to N_{HI}, though this is slightly higher than the $T_s = 300$ K input into SANGRIA to determine the 21 cm cross section. This may indicate that the Gaussian fit by FLASHfinder fails to recover some complexities in the visible H I profile, or that there is some very low optical depth gas that is strongly Doppler shifted away from the main feature in the spectrum. A more in depth discussion of these differences is left to Ward, Kerrison & Tonnesen (*in prep.*). The key takeaway from this investigation though, is that while some FLASH detections with $\ln(B) < 30$ are undoubtedly real, they may fail to recover the full line profile, and hence the full H I column, until a much higher Bayes factor, corresponding to a much higher signal to noise.

Observationally, Aditya et al. (2024) found in a flux-limited subsample of ASKAP-FLASH pilot data that the only detections of H I were made against intrinsically luminous radio sources, with $L_{1.4\mathrm{GHz}} \geq 10^{27}$ W Hz^{-1}. We find significant detections of H I at much lower luminosities than that in this test ($L_{1.4\mathrm{GHz}} \sim 10^{25.5}$ W Hz^{-1}). We interpret this as an effect of radio morphology; the radio luminosity in these mocks is concentrated in a point-like region at the centre of a high-density gas disk, whereas many radio jets have some spatial extension (more on this below). Overall though, our detection of H I as a function of radio luminosity follows the same trend as seen in ASKAP-FLASH observations, namely that brighter radio jets produce stronger H I features. Physically, this is explained by the fact that radio photons are not sufficiently energetic to photo-ionise ambient neutral gas, so we expect increasing radio luminosity to monotonically increase the signal-to-noise of any spectral line detected in H I absorption.

4.5.2 The effect of jet morphology

Moving from simple, quasar-like radio emission to more complex conditions, we now consider the effects of jet morphology on the detectability of H I absorption. These effects are twofold, due to both the jet length and typology, which we consider separately below.

Jet length

Radio AGN jets have been observed with lengths spanning more than six orders of magnitude in scale, from a few parsecs to more than a megaparsec. However, ASKAP-FLASH is preferentially detecting H I absorption against jets no more than a few kiloparsec end-to-end (Kerrison et al., 2024a; Yoon et al., 2025). To begin exploring this computationally, we again examine galaxy #164 from SIMBA at $z = 0.4$, this time injecting jets of varying lengths.

We fix the luminosity at $L_{1.4\mathrm{GHz}}$=10^{27} W Hz^{-1}, corresponding to a Bayes factor of $\ln(B) \sim 300$ for pointlike, quasar emission in Section 4.5.1. The jet length is varied across two orders of magnitude, from an unresolved FR0 ($2 \times l_{\mathrm{jet}} = 0.2$ kpc) to an extended FRI/FRII ($2 \times l_{\mathrm{jet}} = 200$ kpc) that spans the full width of the simulation cutout (which has a radius $r = 100$ kpc). For this setup, we again align the galaxy disk edge-on to our line of sight, where the orientation of the disk is identified using the averaged angular momentum vector of gas and star particles. We then align the jets perpendicular to the galactic disk. The setup is shown as an inset in Figure 4.6.

As shown in Figure 4.6, when the jets remain confined to within the disk ($l_{\mathrm{jet}} \lesssim 2$ kpc) the integrated optical depth of H I absorption (a proxy for the H I column density) remains constant to within statistical uncertainties though slightly lower than the point source optical depth shown in Figure 4.5. Once the jets begin to break out of the disk, the integrated optical depth drops exponentially until it falls below the significance threshold for the FLASH survey ($\ln(B) = 30$) at a total length $2 \times l_{\mathrm{jet}} \geq 20$ kpc for an FRII, or $2 \times l_{\mathrm{jet}} \geq 40$ kpc for an FRI. This trend reflects the radial profile of the gas in galaxy #164, where larger jets subtend more of the low density, circumgalactic medium. This is essentially an extension of the radial profiles mapped observationally with intervening Mg II absorbers (Bordoloi et al., 2011), and in cosmological mocks of Ly α absorption (Weng et al., 2024) down to the lowest impact parameters, and to radio wavelengths. If we perform this same exercise with the disk oriented face-on to our line of sight we find the same trend, but none of the Bayes factors are above 30, and $\tau_{int} > 0$ is only found for a few jets where $2 \times l_{\mathrm{jet}} < 1$ kpc.

For this test we have been careful to describe total jet lengths ($2 \times l_{\mathrm{jet}}$), to show that our threshold for jet breakout corresponds to that seen observationally in PS sources ($\sim$ a few kpc), but for the remainder of this work we will discuss our results directly as a function of l_{jet}.

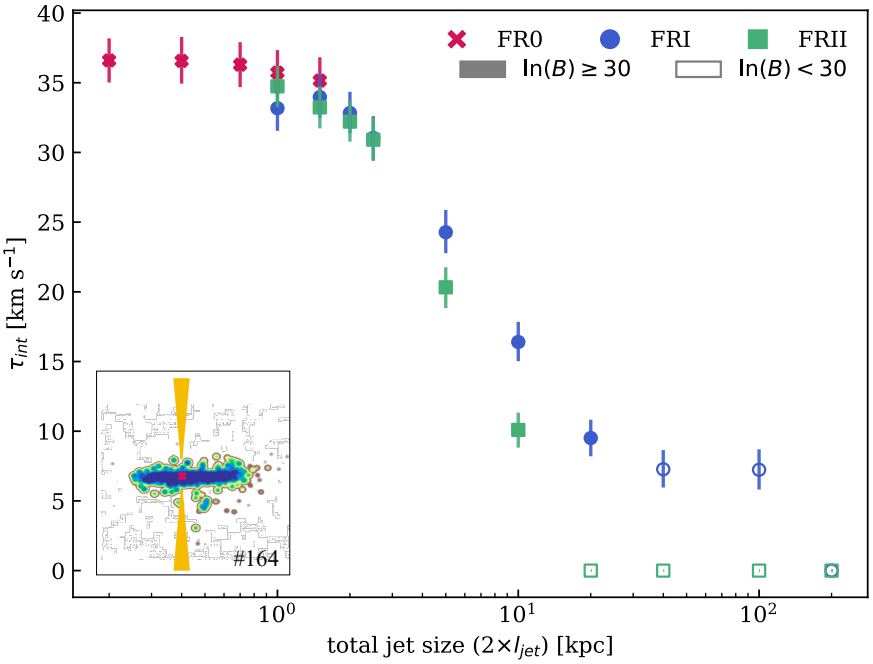


Figure 4.6: The effect of jet length on the detectability of H I in absorption. The most compact jets are modelled with an unresolved, FR0 morphology (pink crosses), but beyond 1 kpc we insert both FRI (blue circles) and FRII jets (green squares) into this setup. Non-detections ($\ln(B) < 30$) have unfilled markers. We fix the luminosity at $L_{1.4\mathrm{GHz}}$=10^{27} W Hz^{-1}.

The jet type - N_{HI} connection

Associated H I absorption has been detected towards both FRI and FRII-type AGN, both in front of the compact radio core (Morganti et al., 2001), and as jet-driven outflows (Mahony et al., 2013), but the impact of morphological differences on detection rate is difficult to pinpoint observationally. As seen already in Figure 4.6, there is a difference between the strength of H I absorption detected towards an FRI and an FRII once the jet length extends beyond the disk. For the same total luminosity and size, the FRII-type jets produce systematically weaker absorption signatures. In the setup presented in Section 4.5.2, the discrepancy in τ_{int} at a jet length 2.55 kpc is $4 \pm 3\,\mathrm{km\,s^{-1}}$, this extends to $6 \pm 3\,\mathrm{km\,s^{-1}}$ at a jet length of 5 kpc. To ensure this is not simply a product of the galaxy chosen, we extend this analysis to our Gal. sample of 266 galaxies in the 'Galaxy rotation' setup, but we keep the angle fixed such that line of sight is into the plane of the disk. The results of this analysis are shown in Figure 4.7, where we present a kernel density estimate (KDE) for the integrated optical depth across this sample, broken down into detections made against FRI/FRII jets (line colour), each at 5 and 10 kpc lengths (linestyle).

It is clear that though modest, this difference in τ_{int} persists across the whole sample. At $l_{\mathrm{jet}} = 5$ kpc detections against FRI-type jets have a median optical depth $1.86\,\mathrm{km\,s^{-1}}$ larger than those towards FRII-types (comparing the dotted lines), and this difference extends to $3.1\,\mathrm{km\,s^{-1}}$ in $l_{\mathrm{jet}} = 10$ kpc jets (comparing the solid lines). This must therefore be a product of the different luminosity distributions in the jets, with core-dominated FRIs preferentially illuminating the dense gas of the inner galactic regions. We note here that all jets below $l_{\mathrm{jet}} = 50$ kpc are unresolved by the ASKAP-FLASH beam, but from this figure it is clear that the jet length and morphology make a significant difference to what is happening on 'sub-beam' scales, which manifests as differences in the observed H I line profile.

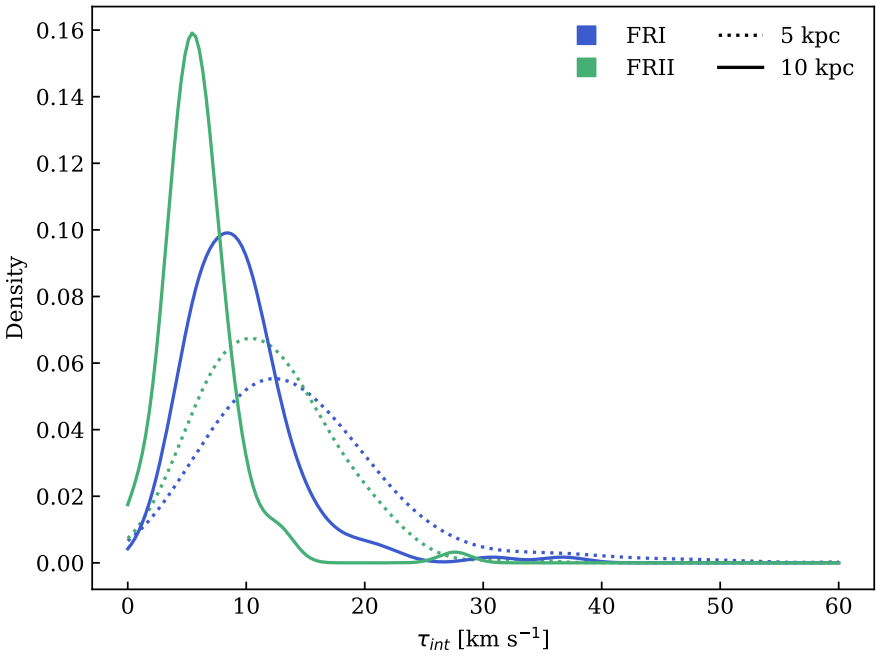


Figure 4.7: The distribution in integrated optical depth (τ_{int}) across all 266 galaxies in our sample when observed with an FRI (blue) and FRII type jet (green). We show the results for both 5 kpc (dotted) and 10 kpc (solid) jets with an integrated luminosity $L_{1.4\mathrm{GHz}}$=10^{27} W Hz^{-1}. The distributions are plotted using a gaussian kernel density estimation (KDE).

4.5.3 The effect of galaxy orientation

So far our experiments have maintained a fixed line of sight through the galactic disk(s). However one of the advantages of mock observations is that this need not be the case, and we are free to choose whatever viewing angle we please. We consider here the effect of viewing angle on our ability to detect H I absorption.

A single galaxy

In Figure 4.8, we show the strength of H I absorption observed towards galaxy #164 illuminated with a $L_{1.4\mathrm{GHz}}$=10^{27} W Hz^{-1} quasar and rotated through 360° (red shaded region) in 5° steps (note this is a much finer sampling of angles than that used for our full 'Galaxy rotation' and 'Jet rotation' sets of mocks). The strongest H I features are seen when the disk is edge-on to our line of sight (0° and 180°), reflecting the fact that these sightlines trace the highest column of gas. We note that even though both edge-on views show more absorption than those closer to face-on, the τ_{int} values still differ by about a factor of 2 between 0° and 180°. Thus the galaxy we have chosen here, although classified as a well-ordered disk, shows strong evidence of disk asymmetry in its H I absorption profile. This phenomenon is well documented observationally (Jog & Combes, 2009), and in other cosmological simulations (Łokas, 2022).

If the jets extend far beyond the H I disk, an edge-on inclination does not show significantly stronger absorption features than any other inclination angle. This is represented by the blue shaded region in Figure 4.8, which describes the H I profile seen towards an FRII with l_{jet} = 10 kpc. In this case, the strength of the H I feature shows a much weaker dependence on the inclination angle of the galactic disk, and it does not reach an optical depth greater than τ_{int} = 10.5 km s^{-1} for any inclination. At this luminosity, these detections are statistically significant, but towards fainter jets this would quickly become a non-significant detection based on the results from Section 4.5.1. Of course, this is but one of our 266 galaxies, we consider next the results across the full galaxy rotation ('Gal.') sample.

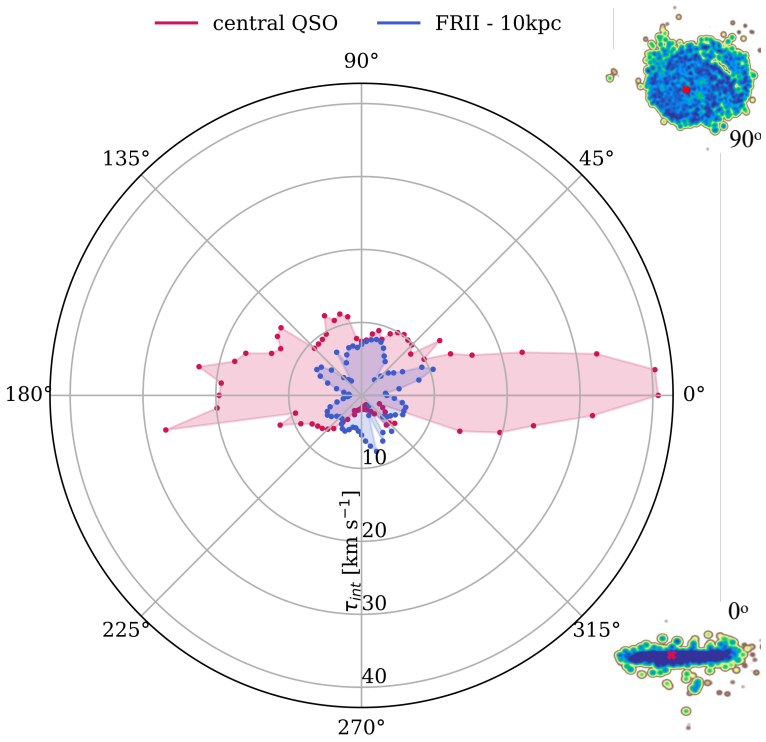


Figure 4.8: The integrated optical depth τ_{int} as a function of inclination angle of the galaxy, where 0° corresponds to the disk edge-on, 90° to face on. Red indicates the rotational absorption profile seen towards a central 'quasar' (pointlike emission), while the blue shaded region is that of an extended FRII. We fix the luminosity at $L_{1.4\mathrm{GHz}}$=10^{27} W Hz^{-1}.

The entire galaxy population

Across the full 'Galaxy rotation' sample the trend is much the same as in the case of Galaxy #164, namely that galaxy inclination has the strongest effect on the absorption profile when the jets are fully embedded within the gas disk ($l_{\mathrm{jet}} \leq 1$ kpc). This is shown in Figure 4.9, where the top half of the figure shows the angular profile of all FRI-type jets in our simulations, and the bottom half is reserved for FRII-type jets. Since we want to see how the H I profile changes statistically across a variety of galaxies with different gas masses and distributions, we have first normalised the H I absorption strength (τ_{int}) across each galaxy by the maximum optical depth obtained for any jet configuration within that galaxy ($\tau_{int,max}$) to produce what we call a 'normalised optical depth strength' ($\langle \frac{\tau_{int}}{\tau_{int,max}} \rangle$). As we have shown in Figures 4.6 & 4.8, a quasar viewed edge-on will often result in the most absorption. This, in combination with the asymmetry of most of our galaxies (e.g. Figure 4.8), is why even our smallest FRI jets do not often have high $\frac{\tau_{int}}{\tau_{int,max}}$ values.

Figure 4.9 reinforces the impact of jet morphology as discussed above in Section 4.5.2, where the solid (dotted) lines are the mean values of the sample for FRI (FRII) jets of each length, and the shaded regions denote the 1 σ standard deviation across the sample. Core-brightened FRIs produce deeper absorption line profiles when the galaxy is close to edge-on, even at the same overall jet length, than do their edge-brightened counterparts (FRIIs). At the extreme, the 50 kpc FRIIs show much weaker lines across the 'Galaxy rotation' sample than any other jet type, since so little of the integrated luminosity is in a region where it backlights dense H I. This means that even for extremely H I-rich galaxies like those in our sample, jet morphology plays a role in the detectability of H I absorption across a range of galactic viewing angles, and the effect is most exaggerated in edge-on disks. As the disk approaches a face-on viewing angle, the impact of morphology shrinks to be almost negligible, since the half of the jet furthest from the observer will always illuminate the galactic gas disk with the same integrated luminosity, just distributed slightly differently due to jet morphology.

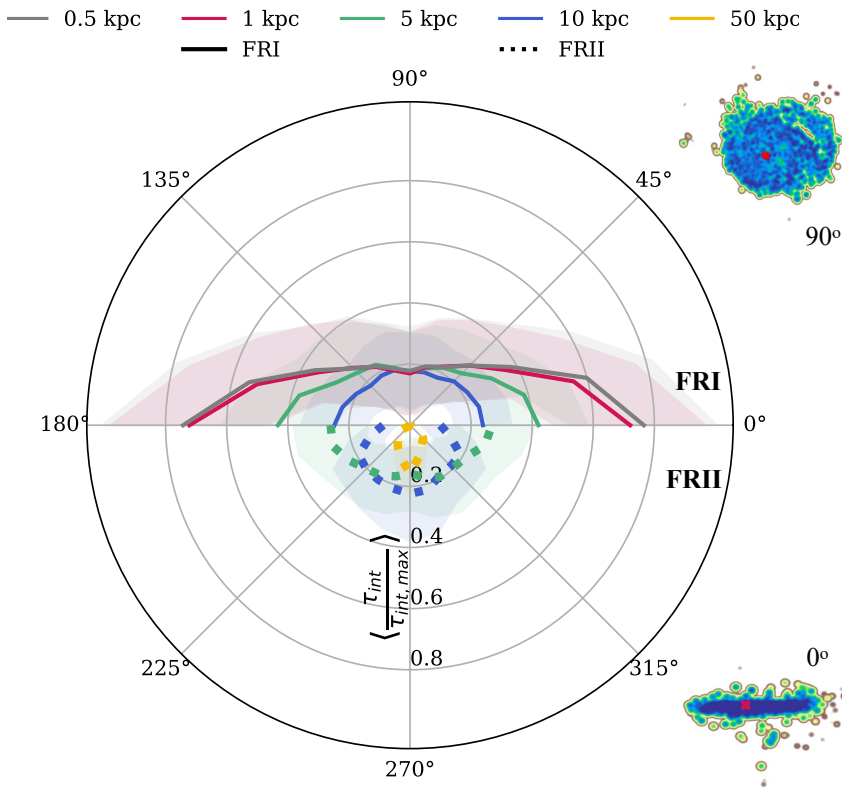


Figure 4.9: Similar to Figure 4.8 but for the full sample and all extended jet types. The top half of the plot is for FRI-type jets, the bottom half for FRII-type. Measured is the 'normalised optical depth strength', $\langle \frac{\tau_{int}}{\tau_{int,max}} \rangle$, calculated as the optical depth for each angle, normalised by the maximum optical depth attained for that galaxy. Lines are the mean across the full sample. As in Figure 4.8, 0° corresponds to the disk edge-on, 90° to face on. We fix the luminosity at $L_{1.4\mathrm{GHz}}$=10^{27} W Hz^{-1}.

4.5.4 The effect of jet orientation

A galaxy need not always have radio jets aligned perpendicular to the galactic disk. Observationally, Zheng et al. (2024) found that the degree of jet-galaxy alignment was dependent on galaxy properties, and Jung et al. (2025) showed that jet-galaxy alignment is more randomised close to cosmic filaments, with the jets preferentially aligned along the filament rather than the galaxy minor axis (likely a product of merger activity). Moreover, Mukherjee et al. (2018b) demonstrated in idealised simulations that sub-kiloparsec scale jets initially aligned close to the plane of the galaxy could still drive kiloparsec-scale outflows akin to those seen observationally in multiphase gas. We now consider cases where our injected jets are not aligned perpendicular to the gas disk of the galaxy, and are instead allowed to vary.

In performing the following analysis, we fix the orientation of our disks to an inclination angle of 60° (where 0° is edge-on to the observer, as in Section 4.5.3), as to first order this is the median inclination angle for a random, uniform distribution of galaxy orientations in three dimensions (it is also the approximate peak of the distribution of SDSS galaxies studied in Morselli et al. 2016).

A single galaxy

As with galaxy orientation in Section 4.5.3, we consider first the effect of jet orientation on a single galaxy, taking Simba Galaxy #164 as our test case. This time, we test all jet lengths used in our full suite of mock observations, shown in Figure 4.10, again fixing the jet luminosity to $L_{1.4\mathrm{GHz}}$=10^{27} W Hz^{-1} to ensure a high signal to noise on all possible detection configurations.

To orient ourselves, we can first compare the τ_{int} values in Figure 4.10 to those in Figure 4.8. Because the galaxy is inclined far from edge-on with respect to our line of sight, the τ_{int} values are much lower than the maximum values in Figure 4.8. In addition, rotating the jet while keeping the galaxy fixed leads to much

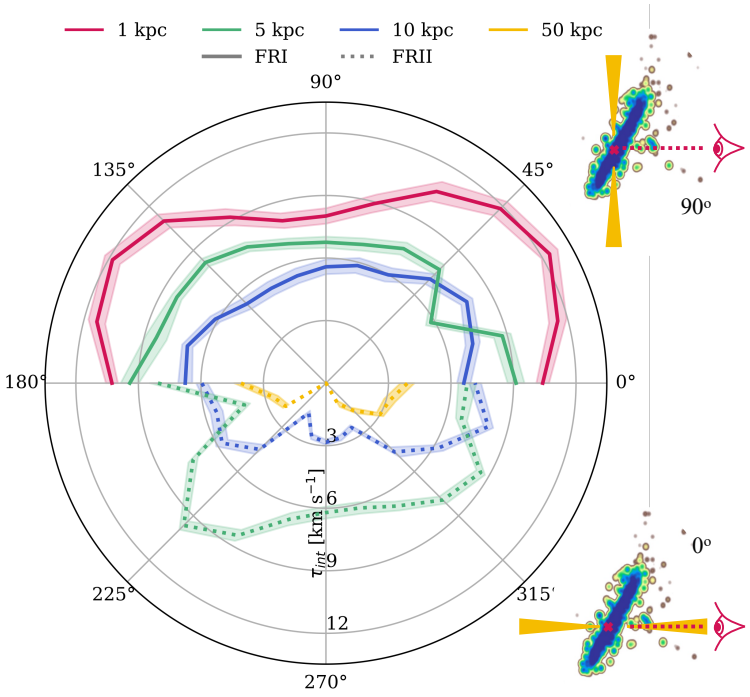


Figure 4.10: The integrated optical depth τ_{int} for Galaxy #164 as a function of inclination angle of the jet, where 0° corresponds to the jet viewed 'down the barrel', 90° to the jet projected across the field of view. Shaded regions around each line indicate the uncertainty on τ_{int}, as reported by the FLASHfinder. We fix the luminosity at $L_{1.4\mathrm{GHz}}$=10^{27} W Hz^{-1}.

smaller variations in the τ_{int} values than rotating the galaxy and jet in the FRI and smaller FRII jets, while the larger 10 kpc FRII jet actually varies by a factor of two depending on its' orientation. This is likely because changing the orientation of a longer jet means that very different column density gas in the disk will be the brightest backlit. As with our conclusions about jet type and galaxy orientation, the amount of gas between the illuminating source and the observer is the key factor driving variations in τ_{int}. In this setup, the FRI and FRII jets of the same length (same colour in the plot) produce different H I profiles even at inclinations of 0° and 180°, when the jets are viewed down the barrel. This is a result of the luminosity being concentrated within a smaller cross-section in core-dominant FRIs compared to edge-brightened FRIIs, which in galaxy #164 reveals variations in H I across the disk, and is an effect we would expect to wash out across all 266 galaxies.

We also note that there are asymmetries in the H I strength over the range of inclinations tested here, especially towards compact jets (FRI, l_{jet} = 1 kpc, red solid line in Figure 4.10). We interpret this again as a signature of the gas disk morphology in galaxy #164, since these asymmetries wash out as the jets extend in length and the integrated luminosity backlights a progressively larger area of the disk. Indeed at a jet length of 50 kpc (yellow dotted line) the original asymmetry is gone and we see instead that the strongest H I absorption is down the barrel of the jet (0°) in a configuration where the entire back half of the jet illuminates the galaxy. If the initial asymmetry seen for compact jets towards Galaxy #164 is a unique product of the H I distribution in this galaxy, we also would not expect to see any such asymmetry across the full galaxy sample. This is investigated below.

The entire galaxy population

When considering the effect of jet inclination angle across the full galaxy sample as seen in Figure 4.11, the asymmetry at small jet lengths is not apparent. This is despite the fact that all galaxies are inclined identically at 60° to the line of sight. Indeed when averaged across the full sample, jet orientation shows only a very minor variation with angle in any FRI-type configuration (top half of Figure 4.11). Across the FRIIs there is more variation across 180° of jet rotation, but the 5 and 10 kpc configurations agree to within the 1 σ spread (shaded regions) shown on the plot, Furthermore, there is little variation in the median of normalised optical depth strength across the range of jet rotations tested. The most pronounced difference appears for the 50 kpc, FRII

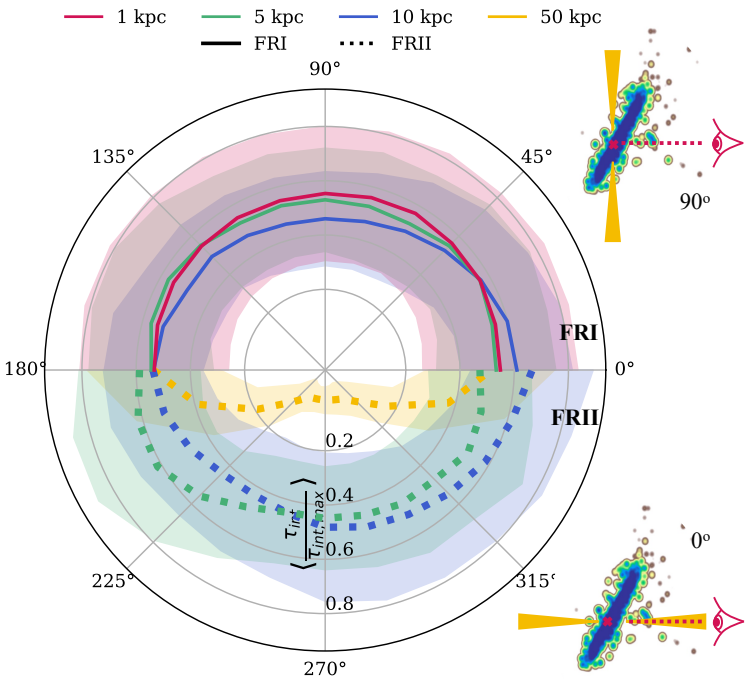


Figure 4.11: Similar to Figure 4.10 but for the full Jet sample. This time, lines are the mean across the full galaxy sample, and shaded regions are $1-\sigma$ standard deviation. The $y-$axis is also the relative H I absorption strength per galaxy, calculated as the strength of any given H I detection for that galaxy, divided by the strongest measured H I line, $\langle \frac{\tau_{int}}{\tau_{int,max}} \rangle$. We fix the luminosity at $L_{1.4\mathrm{GHz}}=10^{27}$ W Hz^{-1}.

jets, which show a marked drop in normalised optical depth once the jet moves more than $\sim 20°$ away from a 'down-the-barrel' configuration, parallel to the line of sight. Again, this makes sense as a 50 kpc jet viewed at an angle is likely to illuminate low column density gas far from the galaxy centre due to both their length and the large size of the radio lobes.

4.5.5 The Impact of Galaxy Morphology

While our previous results considered all 266 galaxies in the 'Galaxy rotation' or 'Jet rotation' sample at once, here we compare the mock observations of galaxies separated by morphology. In characterising our sample we have taken care to separate out galaxies with ordered disks from those which might observationally be considered bulge-dominated ellipticals, or morphologically unusual systems caused by recent galaxy-galaxy interactions. Our sample selection ensures all such systems we study are H I-rich, so if we see any differences in our detection fraction across the sample it must be due to the spatial distribution of H I.

Sadler et al. *in prep.* analyse the optical properties of ASKAP-FLASH detections from the Pilot Surveys. They identify a higher fraction of associated H I detections in bulge-dominated systems as opposed to disk-dominated ones, and these detections tend to have higher H I velocity widths (Δv), though these findings are across a small (30-detection) sample. We test here whether similar trends are observed in our high gas mass sample of Simba galaxies.

In Figure 4.12, the left panels present the linewidths of statistically significant ($\ln(B) \geq 30$) H I detections made towards our entire Simba sample with the galaxy-jet system rotating ('Gal.' setup, top) and galaxy inclination fixed ('Jet' setup, bottom), for all jet variations, and a radio luminosity $L_{1.4\mathrm{GHz}}=10^{27}$ W Hz^{-1}. The data are divided into galaxies classified morphologically as disks (light histogram), those with more spherical gas distributions, which might observationally be classified as 'bulge-dominated' (dark histogram), and 'mergers and rogues' which are actively merging, interacting, or have an extended asymmetric distribution setting them apart from narrow disks or spherical bulges (gray histogram). This final category of 'mergers and rogues' most closely resembles the 'interacting' category discussed in Section 4.4.2. As a reminder, 18% of ASKAP-FLASH Pilot Survey detections appear to have nearby ($\lesssim 20$ kpc) companions, and previous observational studies have found a high H I detection rate towards mergers (Dutta et al., 2018).

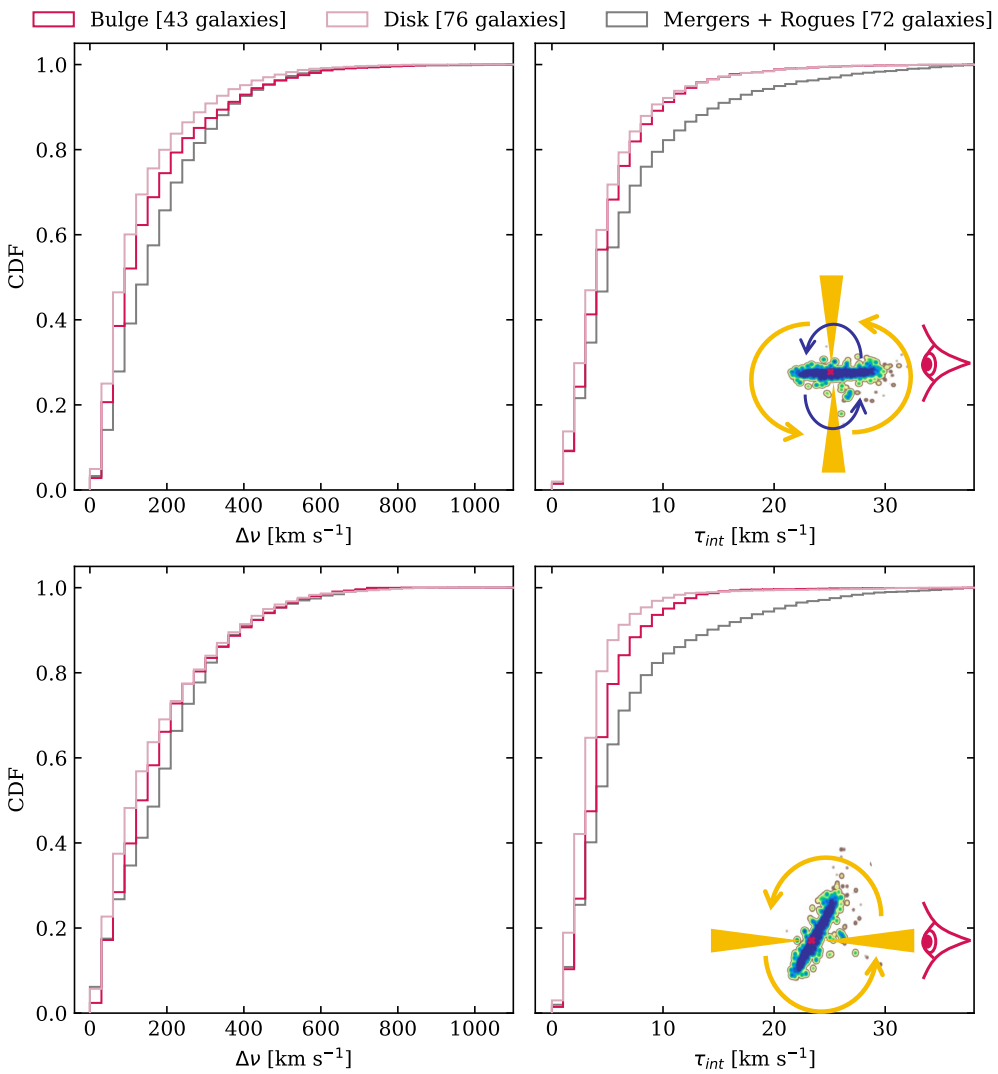


Figure 4.12: *Left:* The linewidths of statistically significant ($\ln(B) \geq 30$) H I detections made towards the entire SIMBA sample with all jet variations, and a radio luminosity $L_{1.4\mathrm{GHz}}$=10^{27} W Hz^{-1}. *Right:* The same but for the overall H I absorption strength (τ_{int}). *Top:* For the 'Galaxy rotation' set of mocks, indicated in the cartoon by both the galaxy (blue arrows) and jet (yellow arrows) rotating together. *Bottom:* for the 'Jet rotation' set of mocks, in which only the jet (yellow lines) rotates, while the galaxy remains inclined by 60 degrees to the line of sight..

	ASKAP-FLASH mean Δv [km s^{-1}]	ASKAP-FLASH median Δv [km s^{-1}]	SIMBA Gal. mean Δv [km s^{-1}]	SIMBA Gal. median Δv [km s^{-1}]	SIMBA Jet mean Δv [km s^{-1}]	SIMBA Jet median Δv [km s^{-1}]
Bulge-dominated	126.4 ± 14.1	114.6	165.9 ± 2.7	114.8	191.3 ± 2.7	150.0
Disk-dominated	93.2 ± 13.7	97.7	142.1 ± 3.4	96.7	177.8 ± 2.1	125.9
Mergers + rogues	−−	−−	190.2 ± 2.0	154.7	206.3 ± 2.1	185.0

Table 4.2: Summary statistics for the width of absorption lines (Δv) across the ASKAP-FLASH Pilot Surveys, and our two SIMBA samples, with the galaxy-jet system rotating uniformly ('Gal.') and the jet rotating with respect to a galaxy fixed at a 60° inclination ('Jet).

We note here that the sub-samples in Figure 4.12 do not sum to 266, because some galaxies in our sample could not be straightforwardly classified as bulge-dominated, disk-dominated or obvious mergers and rogues by eye.

We perform a two-sample Kolmogorov-Smirnov test with the hypothesis that H I widths in disk galaxies are equal to those in 'bulge-like' galaxies. Even assuming a strict threshold of 0.01, we obtain a p-value orders of magnitude less than this for both our 'Galaxy rotation' sample and the 'Jet rotation' sample. Therefore, we reject the null hypothesis, with our alternative being that H I widths tend to be different in bulge-like Simba galaxies compared to disks. In the left-hand panels of Figure 4.12, the cumulative distribution function (CDF) of bulge-dominated galaxies is systematically found to the right of disk-dominated ones, which we further interpret as evidence that the H I widths of bulge-like Simba galaxies are larger than those of disks, as was observed in the ASKAP-FLASH Pilot sample. The mergers and rogues also have slightly wider lines again. We give the mean and median of the linewidth distributions Δv for the sub-populations in the ASKAP-FLASH Pilot Survey, and in our simulations, in Table 4.2. These are in line with the results of the KS test.

Performing the same KS analysis on overall H I absorption strength (τ_{int}) rather than width, we once more reject the null hypothesis that the H I line profiles are drawn from the same underlying distribution (p-values again order of magnitude below our cut-off of 0.01 for both the 'Galaxy rotation' sample and the 'Jet rotation' sample). The CDF for H I absorption strength is given in Figure 4.12, right, which shows that the mergers and rogues tend have the strongest lines of the three populations, followed by the bulge-dominated galaxies, and then finally the disks.

In performing this analysis, we acknowledge that by eye the CDFs in Figure 4.12 have similar shapes. However, since each galaxy has been observed across 12 different angles with 7 different jet morphologies, the samples of 43 / 76 / 72 galaxies indicated in the figure legend actually translate to 2,961 / 5,218 / 4,947 significant H I detections which contribute to the CDF. The observed difference in these distributions is therefore statistically significant. Furthermore, each of these categories has a similar rate of non-detections, falling within 11.8 ± 0.1 per cent, so the difference in shape is unlikely to be due to a difference in overall detection rate. We note that this non-detection rate is comparable to that found in Dutta et al. (2018), who had a non-detection rate of ~ 17 per cent. However, we have chosen to fix our jet luminosities high ($L_{1.4\mathrm{GHz}}$=10^{27} W Hz^{-1}), so we caution against over-interpreting this figure here.

Not shown in these CDFs is the upper limit on the integrated optical depth τ_{int} detected in our Simba samples. This was done to better highlight the discrepancies between different sub-populations at low τ_{int}. However, we note the strongest H I absorption profile detected to date was towards MRC 0531-237, detected as part of the ASKAP-FLASH Pilot Survey sample, with $\tau_{int} = 143.80 \pm 0.35\,\mathrm{km\,s^{-1}}$ (Aditya et al., 2024). This is near the limit of H I absorption strength detected in our mock observations, bearing in mind that we have designed our sample to contain those galaxies with the highest H I masses from the Simba box. Only one Simba galaxy (#3055) produces mock line profiles as strong as MRC 0531-237, and it is classed as a merger or rogue. It will be interesting to see just how rare these extremely strong H I profiles are in the full FLASH survey; their occurrence rate may be an interesting test for the distribution of cool, neutral gas in cosmological simulations.

Overall then, our mock observations tell us that compared to disks, we are more likely to detect H I absorption profiles that are both wider and deeper towards galaxies with spherical gas distributions, which might observationally be classified as 'bulge-dominated', and profiles that are wider and deeper again towards 'mergers and rogues'. If merging/interacting galaxies do tend to have wider and stronger H I profiles, it may explain the high detection rate towards these objects in targeted samples (e.g. Dutta et al., 2018, and the discussions in Section 4.4.2 of this work). The differences between bulges and disks found here are in line with the tentative findings in Sadler et al. *in prep.*, but an analysis of the optical counterparts in the full ASKAP-FLASH survey will be necessary to better characterise this dependence on host-galaxy morphology with greater statistical significance.

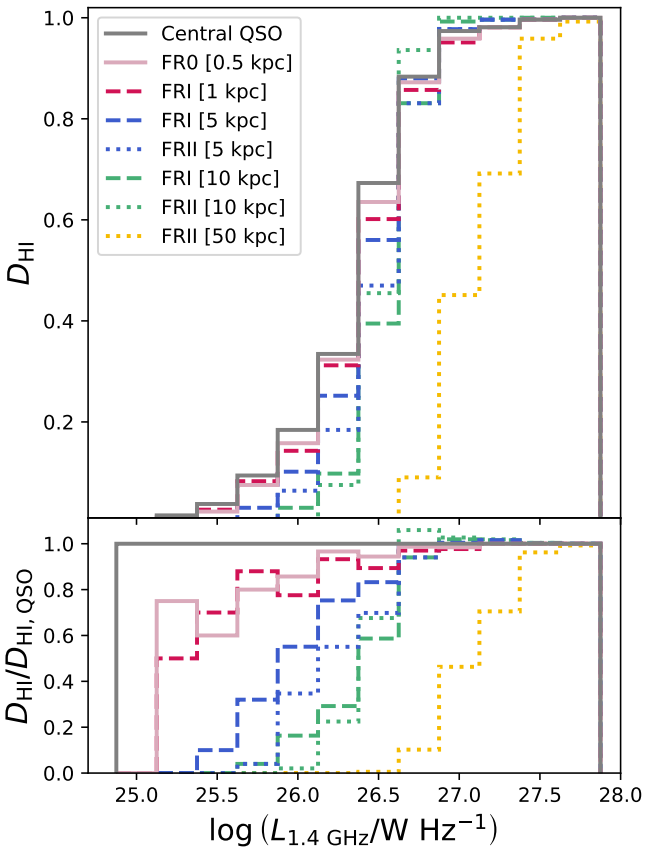


Figure 4.13: The H I detection fraction as a function of jet luminosity for the 'Jet Rotation' setup. *Top panel:* fraction of all galaxies and sightlines that result in a detection of H I , split into the different jet types (solid lines: central quasar/FR0; dashed lines: FRI; dotted lines: FRII) and jet sizes (pale pink: 0.5 kpc; pink: 1 kpc; blue: 5 kpc; green: 10 kpc and yellow: 50 kpc). *Bottom panel:* the H I detection fraction for each jet type relative to the quasar detection fraction. The compact ($\leq$ 1 kpc) jets are detected in H I absorption at a higher rate at low luminosity ($\leq 1 \times 10^{26}$ W Hz^{-1}) than more extended jet morphologies.

4.5.6 Where can FLASH detect HI?

To summarise what we have learned so far; ASKAP-FLASH is most sensitive to H I absorption against bright, $L_{1.4GHz} \geq 10^{27}$ W Hz^{-1} jets (Section 4.5.1), which are compact ($l_{jet} \lesssim 3$ kpc; Section 4.5.2) or at least core-dominant (FRI; Section 4.5.2). Towards these compact jets, edge-on galaxy disks produce the strongest absorption lines (Section 4.5.3), though at orientations approaching 90° jet morphology has less of an impact on H I profile depth, and jet inclination has the strongest effect at the largest scales ($l_{jet} \gtrsim 50$ kpc; Section 4.5.4). Across our full sample, we tend to see wider and stronger lines towards bulge-dominated galaxies than disk-dominated galaxies, and the widest and strongest lines are seen towards merging or highly-interacting systems (Section 4.5.5). This last observation is in line with the early results from the ASKAP-FLASH Pilot Surveys (Sadler et al. *in prep.*).

To synthesise some of this information, we now consider the detection *fractions* as a function of jet morphology, recalling that our original motivation for this investigation was to see whether we could explain the high detection rate towards compact jets in the ASKAP-FLASH sample. In Figure 4.13, we show the fraction of galaxies and sightlines at a given luminosity with significant H I detections (D_{HI}). We split this into the various jet types and lengths, as denoted by the coloured lines. The lower panel shows the detection fraction relative to the case with a central quasar to more easily distinguish between the different jet morphologies.

We can see from this plot that the probability of detecting H I rises rapidly as a function of luminosity, from $D_{HI} \lesssim 10$ per cent for $L_{1.4GHz} \lesssim 10^{25.5}$ W Hz^{-1} to $D_{HI} \gtrsim 90$ per cent for $L_{1.4GHz} \lesssim 10^{27}$ W Hz^{-1}. However, within this luminosity range, the size and luminosity distribution of the illuminating radio source plays a large role in determining H I detection. For example, at $L_{1.4GHz}$=10^{26} W Hz^{-1}, around $D_{HI} \simeq 40$ per cent of galaxies

with a central quasar are detected in H I . Centrally-concentrated FR0s are slightly less efficient, with the lower panel showing a relative detection of $\simeq$ 90 per cent compared to quasars. Extended sources are even less likely to be detected, with FRI-5kpc being detected at rates of 40 per cent compared to quasars; FRII-5kpc at 22 per cent; and the large FRII-10kpc having only a 7 per cent relative detection rate, corresponding to only $D_{\rm HI}\lesssim$ 2 per cent of galaxy sightlines with this jet arrangement having a positive H I detection. These differences decrease at higher radio luminosities, with all jet types being detected at $D_{\rm HI}$= 90 per cent rates in systems with $L_{1.4\rm GHz}\gtrsim 10^{26.75}$ W Hz^{-1}.

The striking conclusion from this figure is how steeply the probability of detecting H I rises with radio luminosity. Increasing the intrinsic radio luminosity of the source by just over a single order of magnitude ($L_{1.4\rm GHz} \simeq 10^{25.5-27}$ W Hz^{-1}) changes the detection fraction from almost 0 to 100 per cent. This suggests a sharp, low luminosity cut-off for H I detections at around $L_{1.4\rm GHz} \simeq 10^{26}$ W Hz^{-1}, which agrees with the high radio luminosities seen for ASKAP-FLASH Pilot detections (Aditya et al., 2024).

4.6 HI detections across jet populations

We have tested the effects of jet morphology, luminosity, and viewing angle in a uniform way across our full sample of H I-rich, AGN-hosting Simba galaxies. In order to make meaningful statements about detection rates in ASKAP-FLASH and, crucially, return to the question of over-detection towards compact sources, we need to incorporate a more physically-motivated distribution of radio luminosities and jet types into our suite of mock observations. For the following work, we focus on the 'Jet rotation' set of mock observations, where the galaxy inclination is fixed at 60° to the line of sight, to reflect the 'average' galaxy seen in an observational survey, and to allow the jet direction to vary relative to the plane of the galaxy.

4.6.1 Matching observed populations

We first re-weight our jet luminosities, so that the distribution of jets in our sample matches the observed luminosity function of so-called 'jet-mode' radio AGN at $0.5 < z < 1.0$ described in Best et al. (2014). This is achieved by binning our mock observations by luminosity, and then scaling the bins by a weight, $w_L = N_{\rm target}/N_{\rm bin}$, where $N_{\rm target}$ is derived from Table 2 of Best et al. (2014). In Figure 4.14 we show this parent distribution of radio luminosities (gray line) and the resulting distribution of H I detections (gray histograms). We also break down the distributions by jet type and length, showing the scaled parent distribution by coloured dotted lines, and the predicted H I detections in coloured histograms, with the top panel containing the luminosity distribution of compact (FR0 and 1 kpc FRI) jets, the middle the broken-out FRI (5 kpc and 10 kpc) jets, and bottom the broken-out FRII (5 kpc, 10 kpc, and 50 kpc) jets. We see that the largest effect is on the high luminosity end of our detections, for which the effective number density is significantly reduced. By contrast, below $L_{1.4\rm GHz}\sim 10^{26.75}$ W Hz^{-1} the luminosity distribution of H I detections remains unchanged, and as we move towards lower luminosities more compact sources tend to dominate. We currently make the assumption that each jet type follows the same distribution, i.e. the ratio between the different jet types is constant as a function of luminosity. A more finely-tuned re-weighting is left for future work, since this would required that the independent luminosity functions of FRIs and FRIIs be well constrained at $z \sim 0.5$. This may soon be possible using future releases of LOw-Frequency ARray (LOFAR) Two-metre Sky Survey (LoTSS; Shimwell et al., 2017), and the Evolutionary Map of the Universe (EMU; Norris et al., 2011). In the following section we motivate our split between the jet types and show the distribution in Table 4.3.

Next we must determine our threshold for a simulated jet to be 'broken out' of its host galaxy. Observationally, the broadband radio spectral energy distribution (SED) is a proxy for this, if we assume that any radio AGN with broadband absorption must still be embedded in the dense, ambient medium of its host galaxy. Typically, such sources are found to have jet lengths $\ll 10^1$ kpc (O'Dea & Saikia, 2021). In Section 4.5.2, we showed that for a typical disk-like Simba galaxy, the optical depth strength τ_{int} drops rapidly for jet lengths beyond $\sim$ 2 kpc (see especially Figure 4.6). Therefore, we consider here any jets with lengths $\leq$ 1 kpc in our simulations to be still confined to their host galaxy (specificaly, the FR0 and FRI-1 kpc models).

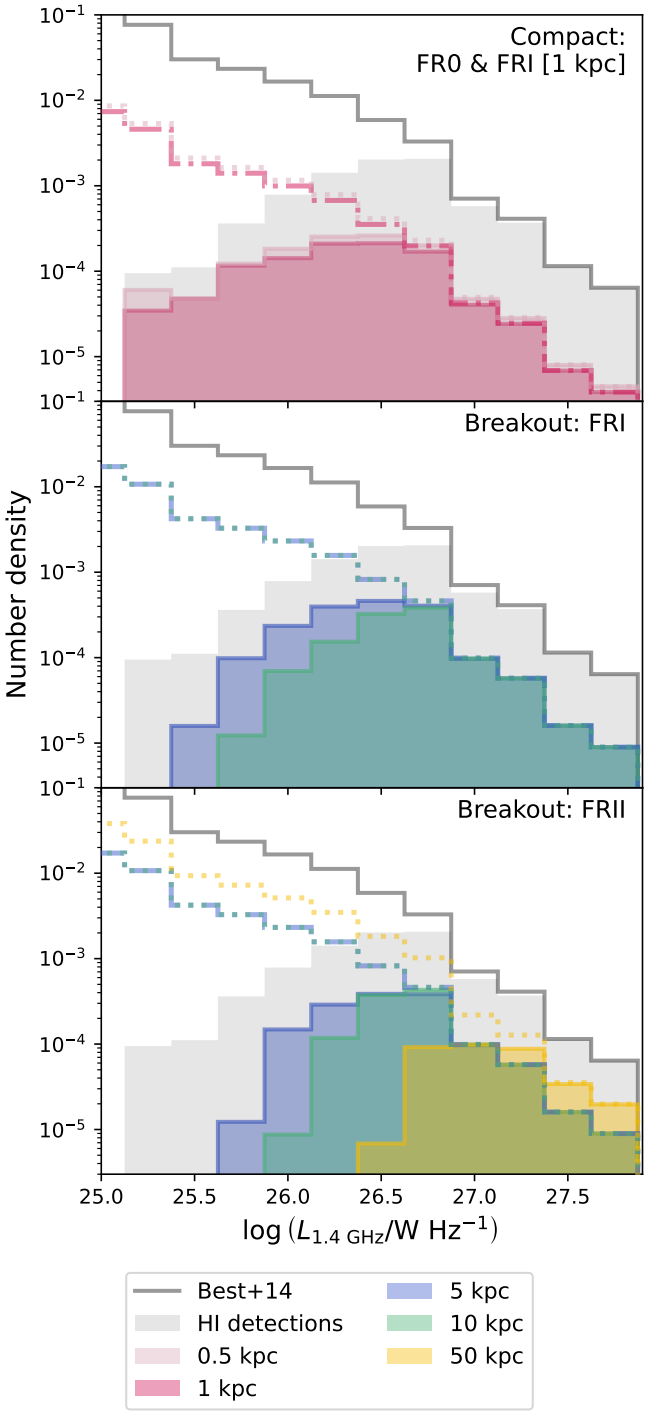


Figure 4.14: Population weighting of the H I detections. The solid, gray line in each subplot shows the Best et al. (2014) radio luminosity function for $0.5 < z < 1$ in the luminosity range of our sample. The gray, shaded region shows the luminosity distribution of all H I detections after weighting by this function. This total population is split into the different jet types, with the overall re-weighted population for each jet type given by the lines, and the H I detected population after re-weighting shown by the shaded histograms, with colours and linestyles as in Figure 4.13. The lowest luminosities still preferentially select out compact jets, but the number density of high luminosity detections is tapered by the application of the observed luminosity function.

Jet Type	Fraction [%]
Compact	
FR0 [0.5 kpc]	7
FRI [1 kpc]	6
Breakout	
FRI [5 kpc]	14
FRII [5 kpc]	14
FRI [10 kpc]	14
FRII [10 kpc]	14
FRII [50 kpc]	31

Table 4.3: The ratio of different jet morphologies and lengths used as the fiducial parent population in our analysis. The compact fraction (FR0s and 1 kpc FRIs) is set to 13 per cent to match the observational constrains from the ASKAP-FLASH pilot sample (Kerrison et al., 2025). The ratio of breakout FRIIs to FRIs is set to 2:1, based on LOFAR data from Horton et al. (2025) and ASKAP data from Norris et al. (2025), where we place an additional flux density constraint $S \geq 30$ mJy on the latter to match ASKAP-FLASH constraints.

Finally, just as we specify the parent luminosity distribution of our jets, we also need to specify the parent length distribution, as our simulations currently contain an equal number of each jet length-morphology combination. Again, we focus here on the confined/breakout dichotomy, since this is most central to our comparison with ASKAP-FLASH. Observational samples put the fraction of confined jets anywhere between about 4.5 to 30 per cent of the general radio population (Callingham et al., 2017; O'Dea, 1998), but we follow Kerrison et al. (2025) as our reference in this work, as their sample of compact AGN was initially selected at ~ 0.8 GHz using ASKAP survey data and a flux density limit of $S_{0.8_{\rm GHz}} \geq 5$ mJy. Their sample is therefore the most well-matched to the samples observed by ASKAP-FLASH. They found a compact fraction of 12.9 per cent (with a 95% confidence interval [11.7%,14.2%]). The remaining 87 per cent of the population is distributed equally amongst the remaining lengths, but with a weighting of 2:1 for FRII:FRI type jets. This ratio of FRII:FRI type jets is reflected in the observational samples from both LOFAR (Horton et al., 2025) and ASKAP surveys, provided we apply a flux density cut to their data comparable to the cut used for ASKAP-FLASH ($S \geq 30$ mJy) (Norris et al., 2025). The length distribution used for re-weighting is listed in Table 4.3. However, the exact length distribution of jets within the confined and breakout categories (not the split between confined/breakout itself) is the main source of uncertainty in our re-weighting scheme, and we consider some alternatives in Appendix F.

4.6.2 Geometry alone cannot explain HI detections

After fixing the parent luminosity distribution, length (compactness) distribution, and our threshold for what constitutes a breakout jet, we are finally ready to consider the H I detection fraction amongst different sub-populations of our jets. This is shown in Figure 4.15, where the left-hand bar gives the parent population of radio AGN after the population re-weighting discussed in Section 4.6. It is divided into the different jet length/morphology combinations used in this work. The relative fractions of each of these populations have been fixed according to the ratios in Table 4.3. For clarity, we highlight the sources considered 'compact' in our simulated sample with a thick, magenta outline (both FR0s and the 1 kpc FRIs). This fraction should be directly compared to the filled diamond, which is the fraction of compact sources derived from observations with a 95 per cent confidence interval derived using a Wilson score interval. In this left-hand bar, the two are identical by construction. In the right-hand bar, we give the population fractions within the subset of radio AGN which are H I detected using the ASKAP-FLASH criterion $\ln(B) \geq 30$. Again, the compact jets in our simulation are highlighted by the thick, magenta outline, and should be directly compared to the filled diamond, which represents the fraction of compact jets detected so far in ASKAP-FLASH.

With our current assumptions, compact sources comprise only ~ 30 per cent of our mock ASKAP-FLASH detections. This is in contrast to the results from the ASKAP-FLASH survey, where 70 per cent (21/30 95% CI [52.1%,83.3%]) of sources show a clear peaked spectrum, interpreted as their being towards non-broken-out jets (Kerrison et al., 2024a; Yoon et al., 2025). If we consider only those sources in the Pilot Survey which are

classified as associated detections using the logistic regression of Curran (2021) (as was done in Yoon et al. 2025), the fraction of associated ASKAP-FLASH detections which are not broken out is even higher, at 12/14 ($\sim$ 86 per cent). This suggests that geometry alone is not enough to explain the high detection rate towards peaked spectrum sources. Instead, they most likely reside in galactic hosts which are uniquely rich in the cold, neutral medium traced by H I absorption. For reference, we show in Appendix F, Figure F.1 the composition of the population required to match the observational fraction of H I detections towards compact sources, and it is only possible if we construct a jet population entirely from the two extremes of morphologies - compact FR0s and 50 kpc FRIIs. Also for reference we supply Figure G.1, applying the same re-weighting scheme to our 'Galaxy Rotation' simulations, which shows a modest boost to the compact detection fraction that is still insufficient to match observations. We consider the implications of this result more fully in the following section.

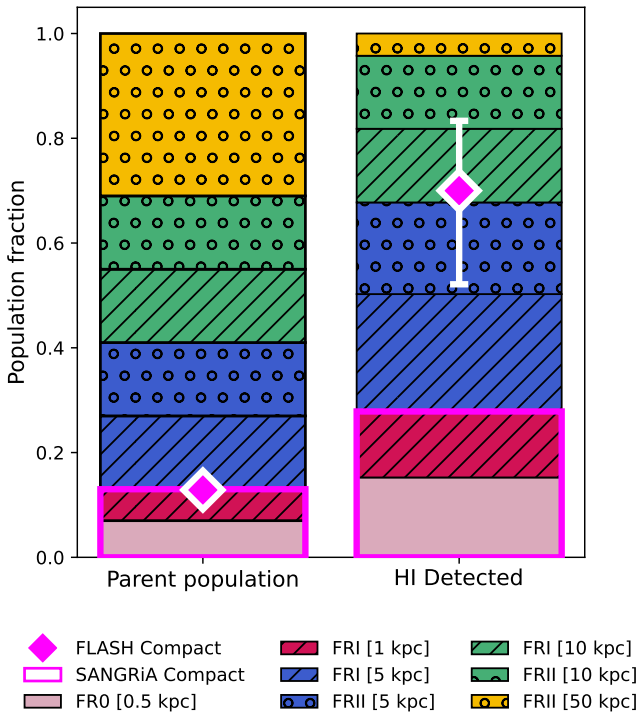


Figure 4.15: *Left:* The distribution of jet lengths and morphologies after re-weighting our full 'Jet rotation' sample by the observational constraints. The total fraction of sources classified as compact is highlighted by the thick, magenta outline. The filled diamond indicates the fraction of compact (PS) sources amongst the general radio AGN population as found in Kerrison et al. (2025). *Right:* The distribution of jet lengths and morphologies amongst our H I detections. The total fraction of compact sources is again given by the thick, magenta outline. The filled diamond indicates the fraction of H I detections in the ASKAP-FLASH Pilot Survey which are compact (PS). For both sides, the uncertainty on the observational fraction of PS sources is expressed as the 95 per cent confidence interval using a Wilson score interval. The fraction of compact sources H I detected in our Simba sample is far lower than the compact fraction in the ASKAP-FLASH sample.

4.7 Discussion

We can now directly address the original motivation of this work: small AGN jets are only $\sim$ 10 per cent of the population but $\sim$ 70 per cent of ASKAP-FLASH Pilot Survey detections. Can this high detection rate amongst compact sources be explained by a solely geometric argument that for any jet or galaxy orientation, smaller jets are more likely to be observed in absorption because they will always be backlighting a high-column density region of the galaxy?

The observational evidence is not (yet) sufficient to answer this question. Perhaps closest is the work of Pihlström et al. (2003), who first identified an anti-correlation between jet size and H I column densities derived from H I absorption. They argued that this relationship was due to a radially decreasing gas density profile, so

that longer jets probe statistically less dense regions of the ISM. However, they did not draw any conclusions about whether this gas density profile is *different* in compact-jetted sources compared to galaxies with extended radio jets. Curran et al. (2013) built on this work, implementing simple models based on the size and geometry of both the radio continuum emission and typical H I clouds to explore the observed anti-correlation between H I column and jet size. They concluded that this relationship could be fully explained by geometry alone, even in extended sources ($l_{\rm jet} \geq 1$ kpc). However, while geometry might explain the observed anti-correlation, their models were not tested against *detection rates* for compact and extended sources, because they were comparing with observational samples in which this was unremarkable.

Recently though, Morganti et al. (2026) argued that the high H I detection rate towards compact AGN jets is likely a combination of geometry and their gas-rich environment. Their analysis was focused on a targeted, observational sample of 17 radio AGN, out of which two detections of associated H I were made, both towards AGN with compact jets. Their interpretation is supported by our findings here.

To our knowledge, this work is the first test of geometrical effects on the detectability of H I absorption using cosmological simulations. In designing our experiment, we intentionally decoupled the ISM conditions and the properties of our injected radio jets, so that any difference in detection rate must be *purely* geometric. By failing to match the observed compact fraction, we have ruled out the purely geometric hypothesis, showing that jet properties and ISM conditions must be linked, which is to say that longer jets are, on average, hosted by H I-poor galaxies and/or that compact jets are found in galaxies which are preferentially H I-rich. The compact geometry of the jets alone provides only a factor of 2 increase to the detection rate against these sources, but this is still far from the factor of ~ 6 required to match the observations.

Although it is beyond the scope of this experiment to determine why compact jets could be preferentially found in H I-rich host galaxies, there are a number of interesting possibilities which other simulations are well-placed to explore. Observations show that compact jets are capable of driving multiphase outflows, including those comprising cold atomic and molecular gas, over parsec to kiloparsec scales (e.g Vermeulen et al., 2003; Jarvis et al., 2019; Morganti et al., 2023; Girdhar et al., 2024). This suggests radio jets could expel some fraction of nuclear H I as they propagate out to kiloparsec scales, making galaxies with extended jets potentially H I-poor. Idealised simulations have also shown that jet-ISM interactions can drive outflows, but that their properties are highly dependent on the nuclear gas density, the jet power, the direction of the jets relative to the galactic gas disk, and local gas turbulence (Mukherjee et al. 2018b; Tanner & Weaver 2022; Borodina et al. 2025, and see review by Mukherjee 2025). Therefore both observations and simulations suggest that breakout jets may have evacuated hosts of gas (ejective feedback).

Alternatively, if the galaxies which host compact radio jets are H I - or more generally gas-rich, it is reasonable to think that the broadband spectral turnover (which is the observational signature of their compactness) might also be caused by free-free absorption in this dense, circumnuclear medium (e.g Bicknell et al., 1997). Under this assumption, the idealised simulations of Young et al. (2025) can not only explain the broadband spectral signature of compact radio jets, but also provide a sensible model for the evolution of their spectrum as they propagate out into the less-dense, outer regions of their host galaxy.

These considerations of jet-ISM interactions remain beyond the scope of this work because of limitations in our current set up. First of all, the spatial resolution of Simba is not high enough to adequately resolve AGN-driven outflows at the level where their multiphase composition could be mapped, as is possible in the idealised simulations mentioned above. Furthermore, the atomic hydrogen content of the gas in Simba is determined via sub-grid models, and although these allow for a two-phase ISM that is calculated on-the-fly, the 'cold' gas is typically at $\sim 10^3 < T < 10^5$, still at least an order of magnitude hotter than the atomic hydrogen traced observationally with H I absorption (Davé et al., 2020).

As we discussed in Section 4.3.3, our approach is to impose a spin temperature of $T_s = 300$ K when calculating the 21 cm cross section, and allow the line to be broadened by the kinetic temperature of the gas particle. This acts as an effective thermal and kinetic broadening component to account for unresolved turbulence below the resolution of Simba and generates absorption lines of depths and widths that match well to those observed with ASKAP-FLASH. To further investigate these assumptions, we could run the Sangria pipeline on simulations with an explicitly resolved cold ISM, such as the idealised simulations mentioned above, or the FIRE suite (Hopkins et al., 2018). This would allow a more detailed study of the H I kinematics and

absorption line profiles, albeit for a much smaller sample of galaxies.

Finally, we do not link our jet parameters to the BH properties recorded in the SIMBA catalogues by choice, but by doing so we lose any information about possible jet-driven outflows, implemented in SIMBA as kinetic-mode AGN feedback. Likewise, the luminosity distribution of our jets, while physically motivated, is simplistic. Coupling jet length to luminosity could be achieved by making use of pre-established semi-analytic models that track these properties over the lifecycle of AGN jets (e.g. Hardcastle, 2018; Turner et al., 2023). And ideally, this could all be joined with insights from high-resolution, idealised simulations like those mentioned above to inform the next iteration of 21 cm absorption mocks from cosmological simulations.

There are several interesting options for future work building on our setup though, which would allow us to address at least some of the limitations in the current work, and which could be pursued independently or combined in a number of useful ways. Perhaps the most natural is to link our injected jet properties with the BH properties of the SIMBA galaxies, following the prescriptions of Thomas et al. (2021) where appropriate. This would also be an opportune way to explore whether compact jets are preferentially launched in gas-rich environments within cosmological simulations, linking back to our findings here. For an even more careful match to ASKAP-FLASH, a larger study of H I absorption in the full SIMBA box, incorporating both associated and intervening detections, would be most useful. Doing so would, however, limit the number of bright radio jets in the box, as discussed in Section 4.4.3. Nevertheless, such an approach would no doubt also benefit from more physically motivated jet modelling also, incorporating semi-analytic approaches like those of Hardcastle (2018) and Turner et al. (2023). These modifications would also pave the way for a full mock lightcone, that stacks together portions of the SIMBA box from different redshifts to emulate a full, 3 dimensional observation, as Thomas et al. (2025) produced for a mock observation of the MIGHTEE survey (Jarvis et al., 2018). Such mocks then open up many posibilities for studying the cosmological evolution of H I in simulated galaxies.

4.8 Conclusions

In this work, we have applied SANGRIA , a new framework for mock 21 cm absorption spectra, to a collection of AGN-hosting galaxies in the SIMBA cosmological simulations. Motivated by observational findings, especially those of the ASKAP-FLASH Pilot Surveys, our aim was to test whether H I absorption is preferentially detected towards compact radio jets due to geometry alone. For this reason, we intentionally decoupled our injected radio jets from the BH properties of our chosen SIMBA galaxies, with the result that any observed variation in detection rates could be purely attributed to jet and galaxy geometry, rather than jet-ISM interactions.

We demonstrate that our pipeline reproduces expected results from ASKAP-FLASH when varying the physical parameters of the jet-galaxy configuration. The strength of H I absorption scales inversely with jet length for a given luminosity because the same amount of radio flux density is spread over a wider area (Section 4.5.2). The significance of H I features also scales with the luminosity, as this is effectively increasing the signal to noise on detections (Section 4.5.1). Overall the strongest lines appear when a galaxy is viewed edge-on (Section 4.5.3), but jet orientation has a minimal effect when averaged across the full sample, except for the longest, 50 kpc jets (Section 4.5.4). Furthermore, as seen in the ASKAP-FLASH Pilot Surveys (Sadler et al. *in prep*), H I line profiles are wider in bulge-dominated galaxies than their disk-dominated counterparts (Section 4.5.5). We also find that our line profiles are widest in systems undergoing active merging, strong interactions or which have an extended, asymmetric gas distribution, suggesting that these galaxies have the most kinematically disturbed gas. We suggest that such a population of complex galaxies traced by wide absorption profiles will likely be detected in larger samples of H I lines from the full ASKAP-FLASH survey.

Putting all these findings together, in Section 4.6 we modify the distribution of our simulated jet-galaxy configurations to better match observed samples. We weight the whole population such that the luminosity distribution follows the observed luminosity function of radio galaxies at $z \sim 0.5$. We additionally weight different jet configurations so that our distribution of jet morphologies matches both the fraction of compact/broken-out morphologies in the literature, and the observed ratio of FRI:FRII type jets. As a result, we can directly compare the fraction of compact jets in our semi-realistic population of mock H I detections to the same fraction in the ASKAP-FLASH Pilot Survey. We find that even with the conservative assumption that all detections in the ASKAP-FLASH Pilot Survey are of H I gas physically associated with the radio source (and not intervening), geometric configurations alone cannot reproduce the observed detection rate towards compact sources within

our framework. We interpret this result as evidence that compact, peaked spectrum sources must be significantly more gas rich than the hosts of extended radio jets in order to produce such a high detection rate in an untargeted survey like ASKAP-FLASH.

In this paper we presented a framework to compare new large observational surveys to cosmological simulations in an apples-to-apples manner which can give us insight into the physical connection between AGN jets and their galaxy hosts - namely that pre-breakout AGN are more likely to reside in gas-rich galaxies. With larger observational datasets, SANGRIA can go far beyond the use in this paper and directly test the cold gas fraction in cosmological simulations beyond the local Universe, probe observational biases in calculating the observed the cosmic H I mass density, Ω_{HI}, and make predictions about the high$-z$ H I forest (an analogue to the Lyα forest) that may soon be detectable with the SKA.

Acknowledgements

EFK acknowledges support from the CCA Pre-doctoral Program, during which this work was initiated, and an Australian Government Research Training Programme stipend (RTP), under which this work was finished. The computations reported in this paper were performed using resources made available by the Flatiron Institute, a division of the Simons foundation.

This scientific work uses data obtained from Inyarrimanha Ilgari Bundara, the CSIRO Murchison Radio-astronomy Observatory. We acknowledge the Wajarri Yamaji People as the Traditional Owners and native title holders of the Observatory site. CSIRO's ASKAP radio telescope is part of the Australia Telescope National Facility (https://ror.org/05qajvd42). Operation of ASKAP is funded by the Australian Government with support from the National Collaborative Research Infrastructure Strategy. ASKAP uses the resources of the Pawsey Supercomputing Research Centre. Establishment of ASKAP, Inyarrimanha Ilgari Bundara, the CSIRO Murchison Radio-astronomy Observatory and the Pawsey Supercomputing Research Centre are initiatives of the Australian Government, with support from the Government of Western Australia and the Science and Industry Endowment Fund.

Facility: ASKAP

Software: SANGRIA (Ward, Kerrison & Tonnesen *(in prep.)*, FLASHfinder (Allison et al., 2012), `astropy` (Astropy Collaboration et al., 2013, 2018, 2022), `matplotlib` (Hunter, 2007), `numpy` (Harris et al., 2020), `python` (Van Rossum & Drake, 2009) and `scipy` (Virtanen et al., 2020; Gommers et al., 2025). This research has made use of the Astrophysics Data System, funded by NASA under Cooperative Agreement 80NSSC21M00561.

5

A new discovery: Intervening HI towards PKS 0405–385

Emily F. Kerrison[1,2,3], Hyein Yoon[1], Elaine M. Sadler[1,2,3], Yijung Kang[2,1,3], Phil G. Edwards[2], Artem Tuntsov[8], Joshua P. Pritchard[2], Vanessa A. Moss[2,1], Elizabeth K. Mahony[2], Hayley Bignall[8], J.N.H.S. Aditya[9,1,10], James R. Allison[11], Steve J. Curran[12], Ron D. Ekers[2], Marcin Glowacki[13,14], Jamie Stevens[2], Renzhi Su[9] and Matthew Whiting[2]

[1]Sydney Institute for Astronomy, School of Physics A28, University of Sydney, NSW, 2006, Australia
[2]ATNF, CSIRO Space and Astronomy, PO Box 76, Epping, NSW 1719, Australia
[3]Institute for Data Innovation in Science, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea
[4]Astronomy Program, Department of Physics and Astronomy, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea
[5]Korea Astronomy and Space Science Institute, 776 Daedeokdae-ro, Daejeon 34055, Republic of Korea
[6]ARC Centre of Excellence for Gravitational WAve Discovery (OzGrav), Australia
[7]Kavli Institute for Particle Astrophysics and Cosmology, SLAC National Accelerator Laboratory, Stanford University, 2575 Sand Hill Road, Menlo Park, CA 94025, USA
[8]Manly Astrophysics, 15/41-42 East Esplanade, Manly, NSW 2095, Australia
[9]Shanghai Astronomical Observatory, CAS, 80 Nandan Road, Shanghai 200030, P. R. China
[10]State Key Laboratory of Radio Astronomy and Technology, A20 Datun Road, Chaoyang District, Beijing, P. R. China
[11]First Light Fusion Ltd., Unit 9/10 Oxford Pioneer Park, Mead Road, Yarnton, Kidlington OX5 1QU, UK
[12]Victoria University of Wellington School of Chemical and Physical Sciences, Ground floor Laby Building, Kelburn Parade, Wellington, NZ 6012
[13]Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh, EH9 3HJ, United Kingdom
[14]Inter-University Institute for Data Intensive Astronomy, Department of Astronomy, University of Cape Town, Cape Town, South Africa

This chapter is published as *One sightline, many systems: a FLASH discovery of HI towards scintillating quasar PKS 0405–385* Kerrison, E. F., *et al.*. 2026. *Publications of the Astronomical Society of Australia* 43, e058. doi:10.1017/pasa.2026.10190.

If the history of science is any guide, the most important insights and facts will be those that are unpredictable.

(Drake, 1982), on SETI

This chapter in context

It is an interesting paradox that unexpected discoveries are a routine part of doing science, and this chapter presents one such discovery. The blazar PKS 0405–385 is not a PS source (indeed it cannot be, by definition), and so it was not within the original scope of this thesis. However, in the course of checking incoming data from the ASKAP-FLASH survey, I noticed a strong, intervening H I absorption feature towards it, which was the genesis of this chapter. What followed was an investigation into the archival data available towards this source which provided some insights into the background radio blazar, but little on the intervening gas. This prompted us to propose for new optical observations (which we were awarded), and consult with a range of observers and theorists who have previously worked on either this particular source, or objects that are similar in some respect. In the following manuscript we present a synthesised version of all this information, from the history of radio monitoring, through the new observations of the intervening gas, to theoretical constraints on Galactic scattering.

Quasar sightlines of the kind presented here, which pass through several intervening galaxies at a range of impact parameters, are critical to our understanding of the diffuse, circumgalactic medium. Optically-selected sightlines are already being studied systematically under the MUSE-ALMA haloes programme (Péroux et al., 2019), which traces the cold, molecular phase of the CGM primarily through CO emission detected with ALMA. ASKAP-FLASH can uniquely contribute to this field by tracing the cold, *atomic* CGM in H I absorption, provided only that the quasar is radio bright, as in the case of PKS 0405–385.

Statement of contribution

This chapter faithfully reproduces Kerrison et al. (2026), which was accepted for publication in the *Publications of the Astronomical Society of Australia*, Vol. 43, id. e058 (April 2026), with minor stylistic adaptations to fit the thesis style. I am the first author, with 17 co-authors including my supervisors, Elaine Sadler and Vanessa Moss. The research problem was conceived of by E. Sadler and myself soon after the initial discovery of the H I detection in ASKAP-FLASH data. New optical imaging and spectroscopic data were obtained by co-author H. Yoon, and archival data was provided in ready-to-use format by co-authors J. Stevens, H. Bignall and P. Edwards. Co-author J. Pritchard aided with additional reduction of archival radio data to produce a lightcurve used in Figure 5.5, and co-author A. Tuntsov performed the theoretical calculations in Section 5.5. I performed the calibration and reduction of the new optical data and synthesis and analysis of all the disparate datasets. I wrote the paper, with all co-authors contributing helpful comments which added to the final version of the manuscript.

5.1 A FLASH discovery of HI towards scintillating quasar PKS 0405–385: Abstract

We report the discovery of an intervening 21 cm absorption line at $z = 0.882$ towards the $z = 1.284$ quasar PKS 0405−385, identified in the First Large Absorption Survey in H I (FLASH). This quasar once displayed the most rapid known intraday variability at radio frequencies, from which it earned the title of 'the smallest radio quasar'. Although its size was revised upwards soon after based on updated scattering theory, PKS 0405−385 remains an important probe of Galactic plasma, and now also of intervening gas discovered through H I absorption. We present new long-slit spectroscopy spanning both PKS 0405−385 and the candidate host of the intervening H I gas. We identify Mg II and Fe II absorption lines in this spectrum consistent with the redshift of the intervening H I, as well as two additional, independent metal-line systems at $z = 0.907$ and $z = 0.966$, but we cannot accurately pinpoint the host(s) of this intervening gas in current data. We revisit the radio variability of PKS 0405−385 in light of advances in scintillation theory, as well as extended monitoring with the Australia Telescope Compact Array and the Australian SKA Pathfinder, and find a revised linear size ≥ 0.3 pc, but no new evidence of repeating intraday variability.

5.2 Introduction

Sightlines towards background AGN are important probes of the circumgalactic environment, providing a pencil beam sample of all the multi-phase gas along the line of sight. In particular, Lyman-α absorption lines in the restframe UV provide most of our current knowledge about neutral atomic hydrogen (H I) in the distant Universe (e.g. Wolfe et al., 2005).

Radio measurements of 21 cm absorption can also provide information about H I in the distant Universe, particularly at $z < 1.7$ where Lyman-α is not yet redshifted enough to be detectable by ground-based optical instruments, complicating the study of H I-rich Damped Lyman-α (DLA) systems with $N_{\rm HI} \geq 2 \times 10^{20}\,{\rm cm}^{-2}$ (e.g. Kanekar & Briggs, 2004; Morganti et al., 2015). The optical depth of the 21 cm absorption line is inversely related to the gas excitation (spin) temperature, so the H I line is most sensitive to cold neutral gas with a spin temperature below ~ 300 K (Morganti & Oosterloo, 2018).

In this paper, we report the discovery of redshifted 21 cm H I absorption along the line of sight to an intellectually (and physically) scintillating radio quasar, PKS 0405−385. The layout of the paper is as follows. In Section 5.3 we summarise what was known about PKS 0405−385 before the discovery of the intervening H I presented here, including estimates of its angular size, and historical importance. In Section 5.4 we present the new H I detection towards this source, and briefly discuss its line characteristics. Section 5.5 offers a more complete analysis of the nature of PKS 0405−385 and the intervening gas based on new optical data (Section 5.5.2) and an up-to-date radio lightcurve (Section 5.5.3). We discuss a few theoretical considerations on the propagation of the light from PKS 0405−385 through foreground matter in Section 5.6. Throughout this work, we adopt a flat, Λ cold dark matter (ΛCDM) cosmology in line with values from Planck Collaboration (2020); $\Omega_{\rm m} = 0.315$, $\Omega_\Lambda = 0.685$, and $H_0 = 67.4\,{\rm km\,s^{-1}Mpc^{-1}}$.

5.3 The history of PKS 0405–385

PKS 0405−385 (J0406−3826) is an $m_i = 18.10$ mag quasar at $z = 1.28$ primarily known today as an early and extreme example of an Intraday Variable (IDV) radio source. In observations taken with the Australia Telescope Compact Array (ATCA) in 1996, it showed hour-to-hour variations as high as 50% at 4.8 GHz, an order of magnitude more extreme than the variability seen in the previously most variable IDV source, OJ 287 (Kedziora-Chudczer et al., 1997). At the time, scattering theory suggested such variations could only be produced towards a background source with an angular size less than 5 μas, which would have made this the smallest known radio quasar (Kedziora-Chudczer et al., 1997). The initial IDV activity lasted several months, with a second period of IDV observed in 1998 (Kedziora-Chudczer et al., 2001; Kedziora-Chudczer, 2006). During the second period of IDV, 4 nights of optical observations were conducted by H.B. on the ANU 40 inch Telescope at Siding Springs Observatory. The data were also reduced by H.B. using standard IRAF procedures (Tody, 1993). A standard deviation in relative photometry between PKS 0405−385 and a comparison star of

similar magnitude at R-band was 0.08 magnitudes based on 11 measurements taken between 7 December - 10 December of that year. This variability was not deemed significant at the time and the data went unpublished, it will be discussed more thoroughly in a future work, along with other archival data on PKS 0405−385. No further episodes of IDV were seen in a monitoring program that continued until 2002 April (Kedziora-Chudczer, 2006), and a few years later, Rickett (2002) presented a revised distance to the Galactic scattering screen responsible for IDV, increasing the source size estimation.

IDV notwithstanding, PKS 0405−385 is a strong ($S_{5\,\mathrm{GHz}} \sim 2\,\mathrm{Jy}$) source used as part of the International Celestial Reference Frame (Charlot et al., 2020), and at high energies, it is a Fermi GeV gamma-ray blazar. Gong et al. (2022) noted quasi-periodic outbursts on a ∼2.8 year timescale in Fermi data between 2008 August and 2021 November, however that trend has not continued in more recent years (Abdollahi et al., 2023). A visual examination of both long term radio monitoring with the Australia Telescope Compact Array, and v-band optical monitoring with the ASAS-SN network also shows several outbursts at both radio and optical wavelengths over this period (Stevens et al., 2012; Kochanek et al., 2017). These outbursts are physically unrelated to any IDV at radio frequencies; such multiwavelength flaring is often seen in blazars, and is typically explained by shocks forming in the core and propagating out along the radio jet (e.g. Beaklini et al., 2017).

In the optical-IR, PKS 0405−385 is not red by the definition of Ross et al. (2015) ($r_{\mathrm{AB}} - W4_{\mathrm{Vega}} > 14\,\mathrm{mag}$), having $r_{\mathrm{AB}} - W4_{\mathrm{Vega}} = 11.3\,\mathrm{mag}$, nor is it red using the more relaxed definition of Glowacki et al. (2019) (W2−W3 > 3.5 mag), as it has a WISE colour W2−W3 = 2.56 mag. Thus although there is a high H I detection rate towards red quasars, this is not one such source (Carilli et al., 1998; Glowacki et al., 2019; Dutta et al., 2020). Also in the optical, this quasar has only one published spectrum, from Véron et al. (1990), in which it is identified as a $z = 1.285$ quasar based on Mg II and C III] emission. An earlier spectrum was discussed by Savage & Wright (1981), who incorrectly placed the source at $z = 2.04$ based on (mis)identifications of C IV and Ly α. However, this earlier spectrum was not published. Véron et al. (1990) note the presence of absorption lines in their spectrum, but state that the resolution is insufficient to attempt identification. There is passing mention in Kedziora-Chudczer et al. (1997) of an intervening absorption system at $z = 0.875$ in the Véron et al. (1990) spectrum, identified by R.W. Hunstead in 1996, presumably based on the association of absorption features in the Véron et al. (1990) spectrum with Fe II and Mg II. We present in the following section the first secure identification of this intervening system.

5.4 A New Discovery : Intervening HI

Two intervening 21 cm H I lines were detected in a radio spectrum of PKS 0405−385 taken on 21 March, 2024 as part of the First Large Absorption Survey in H I (FLASH; Allison et al., 2022; Yoon et al., 2025) conducted with the Australian SKA Pathfinder (ASKAP). FLASH is an untargeted search for H I at redshifts $0.42 < z < 1$ towards all bright ($S \geq 30\,\mathrm{mJy}$) radio sources in the southern sky excluding the Galactic plane. The FLASH spectral cubes have a ∼ 30 arcsec spatial and 18.5 kHz spectral resolution, and each FLASH spectrum uses the full 288 MHz instantaneous bandwidth of the ASKAP radio telescope at 712–1000 MHz (Hotan et al., 2021). A one dimensional spectrum is automatically extracted towards each source above the chosen flux density threshold, averaged over the beam, and continuum subtraction is performed in both the visibility and image plane as part of the ASKAPsoft pipeline (Allison et al., 2022).

The spectrum of PKS 0405−385 is available from the public archive[1], where it is listed as component 3a of scheduling block SB 60306 (FLASH field 212). The segment of the spectrum containing the intervening H I detection is shown in Figure 5.1, where two narrow components are clearly visible with a peak-to-peak velocity separation of approximately $45\,\mathrm{km\,s^{-1}}$. The characteristics of these lines are outlined in Table 5.1, and were obtained using `FLASHfinder` (Allison et al., 2012).

The estimated H I column densities suggest that this is a DLA system, and we note that its column only increases if we assume $T_s > 100\,\mathrm{K}$, as is likely at large galactic radii. This quasar-DLA pair is therefore a potential analogue to the intervening system detected in H I towards PKS 1127-2145 (Kanekar & Chengalur, 2001), where the H I profile was seen to vary in optical depth over the course of 6 months, later attributed to scintillation caused by Galactic scattering (Macquart, 2005). It should also be measured against the quasar-DLA

[1]The CSIRO ASKAP Science Data Archive (CASDA); https://research.csiro.au/casda/

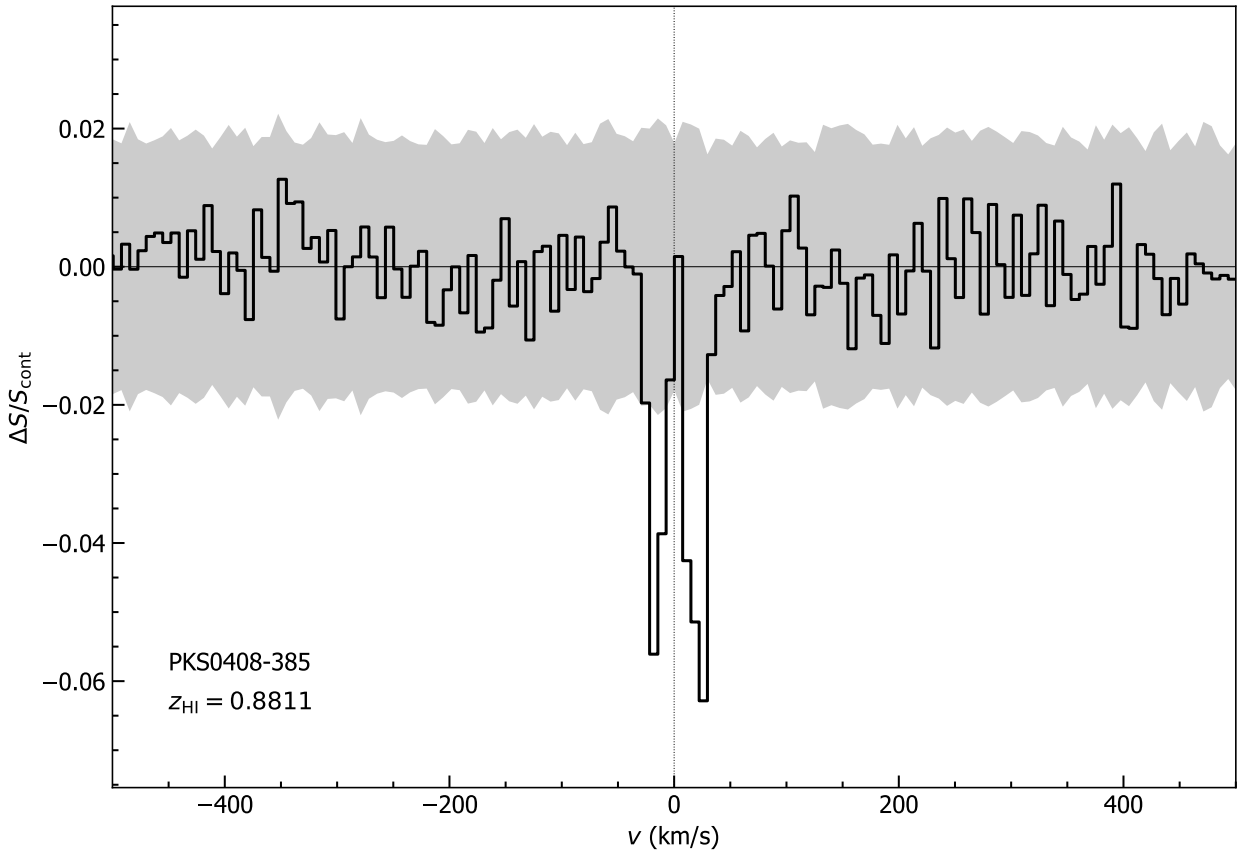


Figure 5.1: ASKAP spectrum of the intervening H I lines towards PKS 0405−385 . The velocity scale is relative to the systemic redshift of $z = 0.88115$. The y-axis indicates the absorption strength as a fraction of the continuum flux density. The grey band indicates 5× the per-channel noise, taken from a blank sky spectrum around the target.

FLASH H I parameters	Component 1	Component 2
Redshift z	0.8809853 ± 0.0000004	0.881207 ± 0.000003
$\tau_{\rm peak}$	0.059 ± 0.007	0.063 ± 0.004
$\tau_{\rm int}$ (km s^{-1})	0.97 ± 0.09	1.29 ± 0.07
Δv (km s^{-1})	16.44 ± 3.5	20.4 ± 1.1
ln (B)	120	230
$N_{H\,I}$ (T_s = 100 K)	1.8×10^{20} cm^{-2}	2.4×10^{20} cm^{-2}
$N_{H\,I}$ (T_s = 1000 K)	1.8×10^{21} cm^{-2}	2.4×10^{21} cm^{-2}
peak-to-peak separation: 44.1 km s^{-1}		

Table 5.1: H I linefinder measurements for PKS 0405−385, derived from fitting a simple Gaussian profile to each component. The first five rows correspond to output from the linefinder, the redshift (z) peak and integrated optical depths ($\tau_{\rm peak}$, $\tau_{\rm int}$), the velocity width (Δv) calculated as $\Delta v = \tau_{int}/\tau_{peak}$ and the logarithm of the Bayes factor, a statistical measure of the preference for a line existing at this location in the spectrum (ln (B)). The column density in the last two rows is derived using the familiar equation, $N_{\rm HI} = 1.823 \times 10^{18} T_s \times f^{-1} \int \tau(v) dv$ and assuming covering factor $f = 1$ and two different spin temperatures for the gas.

pair seen towards PKS 2355-106. There, a second, intervening H I absorption component separated from the first by $\sim 55\,\mathrm{km\,s^{-1}}$ appeared between initial GMRT observations in 2006, and follow up with both MeerKAT and GMRT in 2010. The variability in the H I profile there was interpreted as the result of proper motion of the background source, since the quasar is canonically compact at VLBI resolution and showed insufficient variability in its radio continuum for scintillation to explain the observed variations in H I optical depth (Srianand et al., 2022). The lensed system PMN J0134-0931 is another useful comparison, shown in Kanekar & Briggs (2003) to exhibit a H I absorption profile with two strong, narrow components separated by $\sim 250\,\mathrm{km\,s^{-1}}$ which is reproducible in models with a single, intervening galaxy disk sampled at discrete locations by separate, high surface brightness components of the background radio source. Moreover, recent high resolution, L-band follow up of twelve FLASH detections with the Very Long Baseline Array (VLBA) revealed eleven sources with complex or extended structure on milliarcsecond scales, including the only detection in that sample to have a two component profile, PKS 2007−245 (Aditya et al., 2025). The velocity separation between the components there was $25\,\mathrm{km\,s^{-1}}$. Although not an exact analogue (PKS 2007−245 has lobes spanning ~ 40 mas), the velocity separation between the two H I components here may similarly suggest an underlying complex or core-jet structure in the radio continuum source, where each component draws a discrete sightline through the intervening H I gas.

Most crucially though, the redshift of this intervening system ($z \approx 0.881$) corresponds nicely with the redshift of the intervening line ($z \approx 0.875$) which was identified some 30 years ago in the original Véron et al. (1990) spectrum, but never followed up.

5.5 PKS 0405–385 revisited : new observations

Spurred on by this new detection of intervening H I from an untargeted search, we have revisited PKS 0405−385 to see what can be learnt about both the quasar and this intervening system with additional radio and optical data.

Since the redshift of the newly-discovered H I gas aligns closely with that of the absorption line reported in the original optical spectrum, it offers potential new insight into the multi-phase interstellar (or circumgalactic) medium of the intervening system, if this system can be more securely identified.

Images from DR10 of the Legacy Survey (Dey et al., 2019) reveal five nearby galaxies labelled counter-clockwise A–E in Figure 5.2, left, which have DR9 photometric redshifts broadly commensurate with the FLASH detection given their uncertainties ($0.8 < z_{\mathrm{phot}} < 1.2$). Out of these candidates, Galaxy A has the closest redshift to the FLASH detection at $z = 0.8 \pm 0.3$, making it the most likely host of the H I gas. Furthermore, Galaxy A has the bluest optical colours of these candidates ($g - i = 0.58$ mag from the Legacy Survey DR10) and is therefore likely starforming, so a high H I mass would not be surprising. Should this indeed be the host, the quasar sightline is passing through gas at an impact parameter of ~ 36 kpc (5.07 arcsec), probing the circumgalactic medium at a distance where strong absorption lines are common, at least at earlier cosmic times (Adelberger et al., 2005). Alternatively, the H I may exist in the ISM of another galaxy for which the light in the legacy images is entirely blended with that of the background quasar, due to its extremely small impact parameter. As a third and final alternative, galaxies A–E may form a foreground group, in which case the H I detected in FLASH may sample a clumpy, extragalactic medium, evidence of galaxy-galaxy interactions (e.g. Weng et al., 2022). Unfortunately, the Legacy Survey photometric redshifts are not accurate enough to distinguish between these pictures; optical spectroscopy is required. To achieve this, we obtained fast turnaround time on the 8.1 m Gemini South telescope, using the Gemini Multi-Object Spectrograph (GMOS) for both optical imaging and spectroscopy under project GS-2024B-FT-215 (P.I. Yoon). An analysis of this new optical data is presented below.

5.5.1 GMOS imaging

Our imaging observations comprise 3×100 s in each of the r and i bands, and 3×80 s in the z-band, all obtained on December 10th, 2024. Data pre-processing and reduction were performed using DRAGONS (Data Reduction for Astronomy from Gemini Observatory North and South; Labrie et al., 2023). We re-identified galaxies A–E in the resulting three colour composite image, shown in Figure 5.2, right. No additional nearby galaxies were identified in the GMOS images. Also shown in Figure 5.2, right is the alignment used for longslit

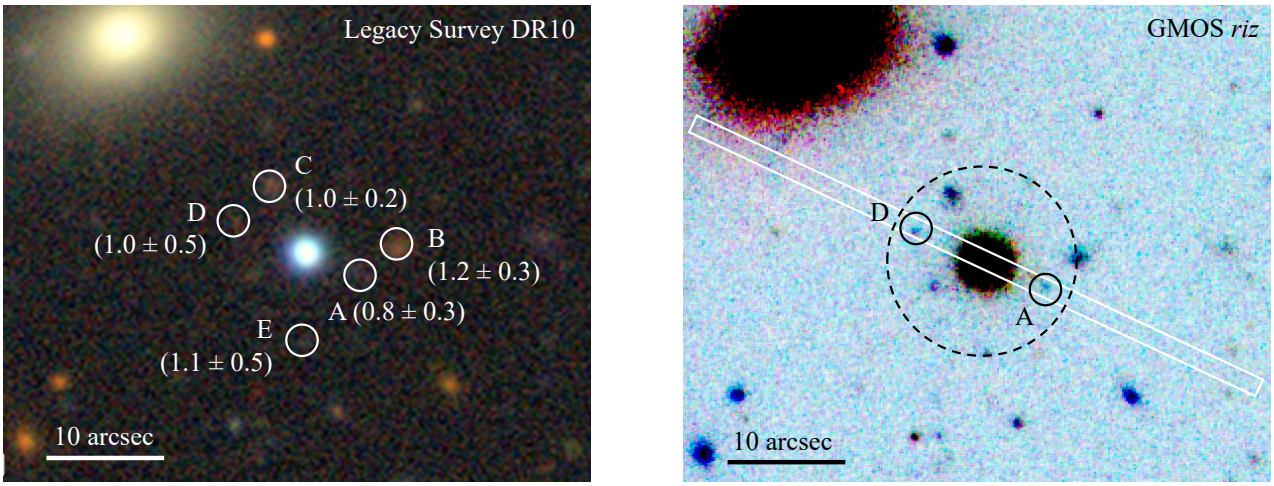


Figure 5.2: Left: three colour image taken from DR10 of the Legacy Survey (Dey et al., 2019) of a region centred on PKS 0405−385. Five nearby galaxies visible in the image are identified as A–E. Galaxies A–E all have photometric redshifts from DR9 of the Legacy Survey within the range [0.8, 1.2] as indicated in the image, with Galaxy A closest to the redshift of the FLASH detection at $z = 0.8 \pm 0.3$. Right: three colour image from Gemini GMOS obtained as part of follow up on this source. The white rectangle indicates the positioning of the slit used to obtain spectroscopy, aligned to span both PKS 0405−385 and Galaxy A (coincidentally also spanning Galaxy D). The circle indicates a region of radius 50 kpc at $z = 0.881$, the redshift of the FLASH detection, centred on PKS 0405−385.

spectroscopy (253.0 deg E of N), to ensure both PKS 0405−385 and the most likely candidate, Galaxy A, would be captured. The circles centred on PKS 0405−385 represent a region of radius 50 kpc at $z = 0.881$, the redshift of the FLASH detection.

5.5.2 GMOS spectroscopy

Spectroscopic observations were carried out on December 9th, 2024 under arcsecond seeing conditions using the B600 grating ($R \sim 1688$, or ~ 0.8 nm FWHM at 510 nm) and a slit width of 1.5 arcseconds. Data acquisition was split into 4×900 s exposures, binned 2×2 in both directions, with two sets using a central wavelength of 510 nm and two set to 520 nm, to fill the gap between detectors. The spectra have a resulting wavelength coverage of approximately 3550 − 6780 Å.

Data pre-processing and reduction were once again performed with DRAGONS, and RV correction to the heliocentric frame was performed using Astropy's `SpecCoord` class. Initial line and redshift identification was made with a modified version of MARZ (Manual and Automatic Redshifting Software; Hinton et al., 2016) with additional, restframe-UV lines. Equivalent Width (EW) measurements were performed with the `Specutils` package within Astropy (Astropy-Specutils Development Team, 2019). The co-added, reduced spectrum has a median signal to noise ratio $SNR = 107$ per pixel across the full spectrum.

The new GMOS spectrum extracted towards PKS 0405−385 is shown in Figure 5.3 alongside the original spectrum from Véron et al. (1990). The full list of lines identified in the new spectrum is given in Table 5.2. The signal-to-noise ratio was not sufficient to extract usable spectra towards the intervening system(s). In both spectra, the Mg II doublet and C III] from PKS 0405−385 are visible in emission, indicated by the red solid lines in Figure 5.2. Given the width of the C III] feature, it is likely in fact to be a blend of C III] (λ1908.83), Si III (λ1892) and Al III (λ1860), the ratios of which can provide insight into the Eddington accretion rate of the quasar (Marziani & Sulentic, 2014; Martínez-Aldama et al., 2018). However, we leave further, detailed discussion of the emission line properties of PKS 0405−385 to future work, ideally with a higher resolution spectrum in which these blended features can be better resolved and subsequently modelled.

The GMOS spectrum clearly resolves the Mg II doublet in absorption at $z = 0.882$ (blue dot-dashed lines), which is unresolved but visible in the Véron et al. (1990) spectrum also. Furthermore, the GMOS spectrum shows several other absorption lines at this redshift, including iron lines at ~ 4400 Å, and iron and manganese

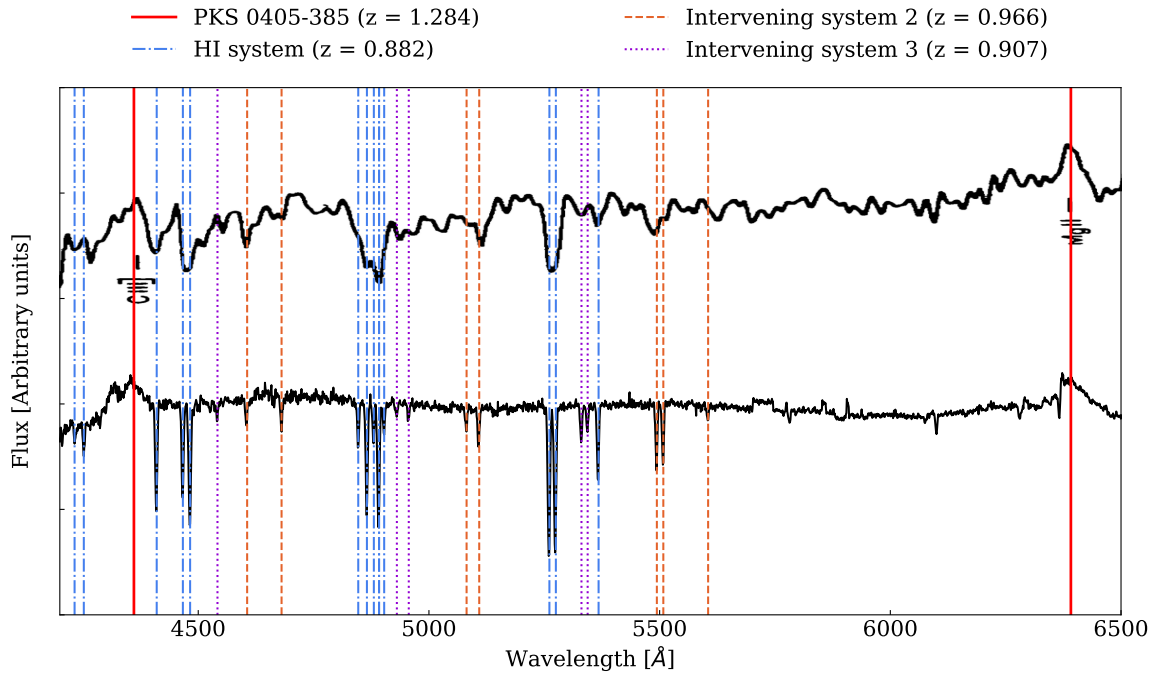


Figure 5.3: The original optical spectrum from Véron et al. (1990) (top) compared to our new spectrum taken with GMOS-S (bottom). Vertical lines indicate emission lines associated with background quasar PKS 0405−385 (red, solid), absorption lines associated with the intervening galaxy detected in FLASH data (blue, dot-dashed), and two further, previously unidentified intervening galaxies (orange, dashed and violet, dotted). Lines were identified using MARZ and the new Gemini spectrum only. A number of lines from both intervening systems are visible in the original Véron et al. (1990) spectrum.

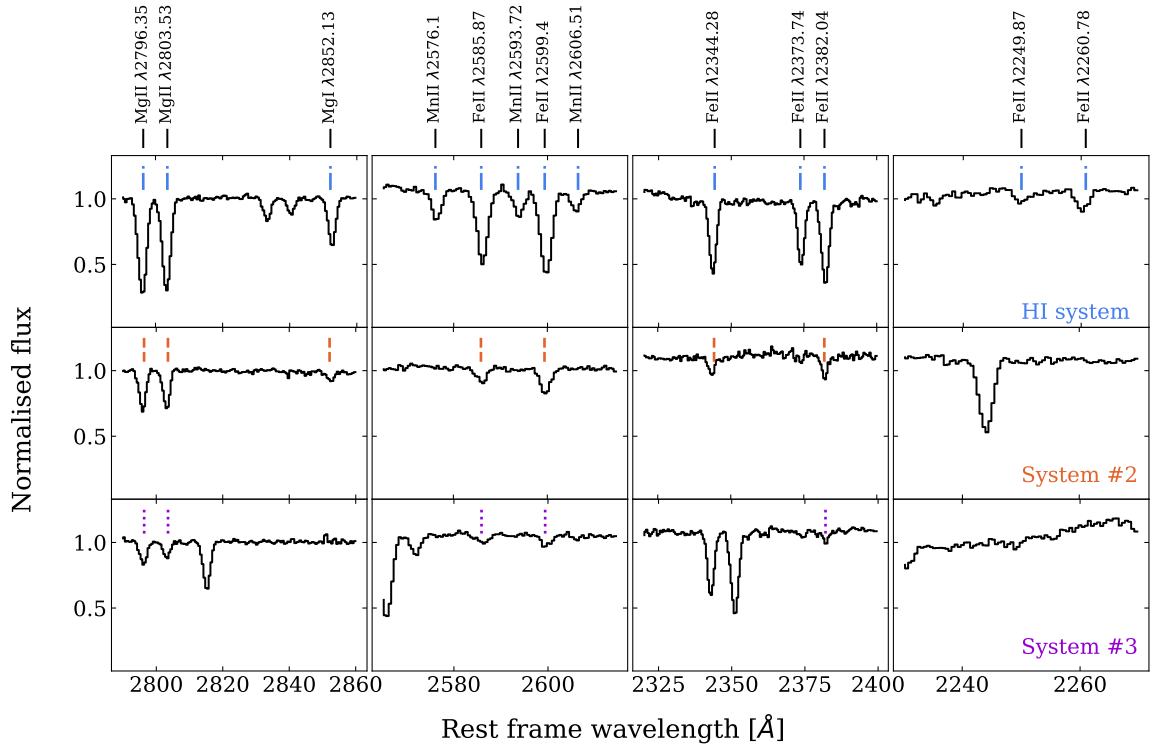


Figure 5.4: Cutouts from the continuum-subtracted GMOS spectrum presented in Figure 5.3 centred on the regions in which absorption lines are seen at the redshift of the H I system (top row) the second, intervening system at $z = 0.966$ (middle row) and the third at $z = 0.907$ (bottom row). Vertical lines in each subplot indicate the detection of an absorption line corresponding to the labels at the top of the figure.

lines at ∼ 4880 Å, both of which also align with unresolved features in the Véron et al. (1990) spectrum. These suggest our original H I detection is towards at least one galaxy at $z = 0.882$ which we refer to henceforth as the 'H I system'. There are possibly also several more lines at the redshift of the FLASH H I line, including Zn II + Cr II(λ2062.66), and Zn II + Mg I(λ2026.14). However, these fall at an observed wavelength $\lambda_o < 4300$Å, where the noise in our new GMOS spectrum is higher, so these detections are currently of low significance.

Interestingly, we also identify two further intervening systems at $z = 0.966$ ('Intervening System 2', orange dotted-dashed lines) and $z = 0.907$ ('Intervening System 3', violet dotted lines), for which several absorption features agree with low-significance features in the Véron et al. (1990) spectrum. Again, the GMOS spectrum shows the Mg II doublet, Mg I and several iron lines at the redshift of the second intervening system, and Mg II and Fe II lines at the redshift of the third. Higher resolution cutouts around the detected lines for all three systems are provided in Figure 5.4. The wavelength range of our spectrum (3550 − 6780 Å) does not extend to cover typical strong, nebular emission lines such as [OII], [OIII] or Hβ at the redshift of any of our intervening systems, so we cannot search for these either against the bright quasar spectrum, or at the location along the slit corresponding to Galaxy A.

A re-examination of the FLASH spectrum within $\pm 500\,\mathrm{km\,s^{-1}}$ of $z = 0.907$ and $z = 0.966$ reveals no H I detection at the $3\,\sigma$ level. However, combining the root mean square noise in optical depth of the FLASH spectrum locally ($\tau_{\mathrm{rms}} = 0.004$) with the average full width zero intensity of intervening H I from Curran (2021) ($108\,\mathrm{km\,s^{-1}}$), we can place an upper limit on the amount of cold, neutral gas at the redshift of these second and third intervening systems. We estimate that they must each have $N_{\mathrm{HI}} \leq 1.3 \times 10^{20}\,\mathrm{cm^{-2}}$ ($\mathrm{T_s} = 100\,\mathrm{K}$, $f = 1$), which would make these sub-DLA systems. We note here that the H I system would be considered iron rich using the classification scheme of Dutta et al. (2017) ($EW_{\mathrm{rest}} > 1.0$ Å), who found that such systems were four times more likely to exhibit H I absorption than their iron poor counterparts (like intervening systems 2 and 3) at $0.5 < z < 1.5$. However, to say more on the abundances of metals in this intervening gas requires higher spectral resolution and, for system 3 at least, higher signal to noise also.

Ultimately, our question as to the origin of the detected H I remains unanswered. Since we were unable to extract any identifiable spectral features at the location of Galaxy A and PKS 0405−385, we were unable to either confirm or disprove this as the host of the intervening H I. Furthermore, many absorption features in the GMOS spectrum are likely saturated making it impossible to deduce abundances, though such analysis would theoretically be possible with the combined detection of (unsaturated) metal lines and neutral H I with well-constrained velocity dispersion. Integral field spectroscopy spanning PKS 0405−385 and galaxies A–E, along with deeper optical imaging, will be key to securely identifying both intervening systems seen in absorption against PKS 0405−385, and further analysing their metallicity.

			PKS 0405−385 z = 1.284		H I system z = 0.882		System #2 z = 0.966		System #3 z = 0.907	
Type	Transition	λ_0 (Å)	λ_{obs} (Å)	EW_{rest} (Å)	λ_{obs} (Å)	EW_{rest} (Å)	λ_{obs} (Å)	EW_{rest} (Å)	λ_{obs} (Å)	EW_{rest} (Å)
Emission	Mg II	2803.53	6396.2	8.6 ± 0.4						
	Mg II	2796.35								
	C III]	1908.73	4353.4	4.0 ± 0.9						
Absorption	Mg I	2852.13			5367.5	0.93 ± 0.07	5605.1	0.25 ± 0.08		
	Mg II	2803.53			5261.1	2.12 ± 0.07	5508.0	0.69 ± 0.07	5344.0	0.42 ± 0.07
	Mg II	2796.35			5275.1	2.25 ± 0.07	5494.0	0.73 ± 0.07	5330.6	0.28 ± 0.07
	Mn II	2606.51			4903.0	0.40 ± 0.12				
	Fe II	2599.40			4891.9	1.89 ± 0.12	5109.0	0.52 ± 0.08	4956.3	0.20 ± 0.11
	Mn II	2593.72			4880.6	0.51 ± 0.12				
	Fe II	2585.87			4865.6	1.53 ± 0.12	5081.6	0.30 ± 0.08	4930.5	0.24 ± 0.11
	Mn II	2576.10			4846.8	0.58 ± 0.12				
	Fe II	2382.04			4482.4	1.84 ± 0.14	4680.6	0.38 ± 0.12		
	Fe II	2373.74			4466.7	1.22 ± 0.14				
	Fe II	2344.28			4410.0	1.37 ± 0.15	4606.0	0.28 ± 0.12	4541.7	0.16 ± 0.13
	Fe II	2260.78			4252.0	0.42 ± 0.11				
	Fe II	2249.87			4232.0	0.21 ± 0.11				

Table 5.2: Lines identified in the GMOS spectrum assigned to each system. We note that the Mg II doublet seen in emission at the redshift of PKS 0405−385 is not resolved. All λ_{obs} values have a measurement uncertainty of ±0.05 Å, and the redshifts should likewise be considered to have a measurement uncertainty of ±0.0005.

5.5.3 Radio monitoring

As discussed in Section 5.2, PKS 0405−385 is an interesting source itself, exhibiting both powerful, episodic IDV and intermittent γ-ray flares. As a result, it has been the subject of long-running radio monitoring, both targeted and incidental, which we compile and present here for the first time in Figure 5.5 alongside the original IDV observations (lower-left inset). PKS 0405−385 has been regularly monitored with the Australia Telescope Compact Array over the frequency range 5 to 40 GHz under observing programmes C007 and C1730 (Stevens et al., 2012) since 2010 (filled circles in Figure 5.5). PKS 0405−385 was also one of the sources included in a search for intra-day variability at 2, 5, and 7 GHz under observing program C2898 between July 2014 and June 2015 (larger, semi-transparent circles). No evidence for IDV was seen (lower-centre inset plot), and the longer term monitoring indicates the source was at its most quiescent over that year. However, just as that program ended, PKS 0405−385 underwent a rapid brightening, reaching historical high flux densities in mid-2016 in the 15 mm and 7 mm bands. PKS 0405−385 is also a calibrator source for the Atacama Large Millimetre Array (ALMA), and 90–240 GHz observations show a peak in flux density coincident with this 2016 flare.[2] We also re-imaged a 10-hour observation at 0.95 GHz from the Evolutionary Map of the Universe survey (EMU Norris et al., 2011) taken with ASKAP in June 2025 using `dstools` (Pritchard, 2025). This is shown in the bottom right inset plot, where there is only minor variability at the level of 1 per cent over the course of several hours. However, we note that even in the original Kedziora-Chudczer et al. (1997) data, the variability was weakest below 2 GHz, so this does not place a strong constraint on the recent level of IDV in this source; higher frequency observations are needed. Kedziora-Chudczer et al. (2001) postulate that the IDV observed in 1996 ceased as a result of the increasing size of the scintillating component, although an alternative explanation could invoke changes in the properties of the scattering screen, as seen towards PKS 1257-326 (Koay et al., 2011).

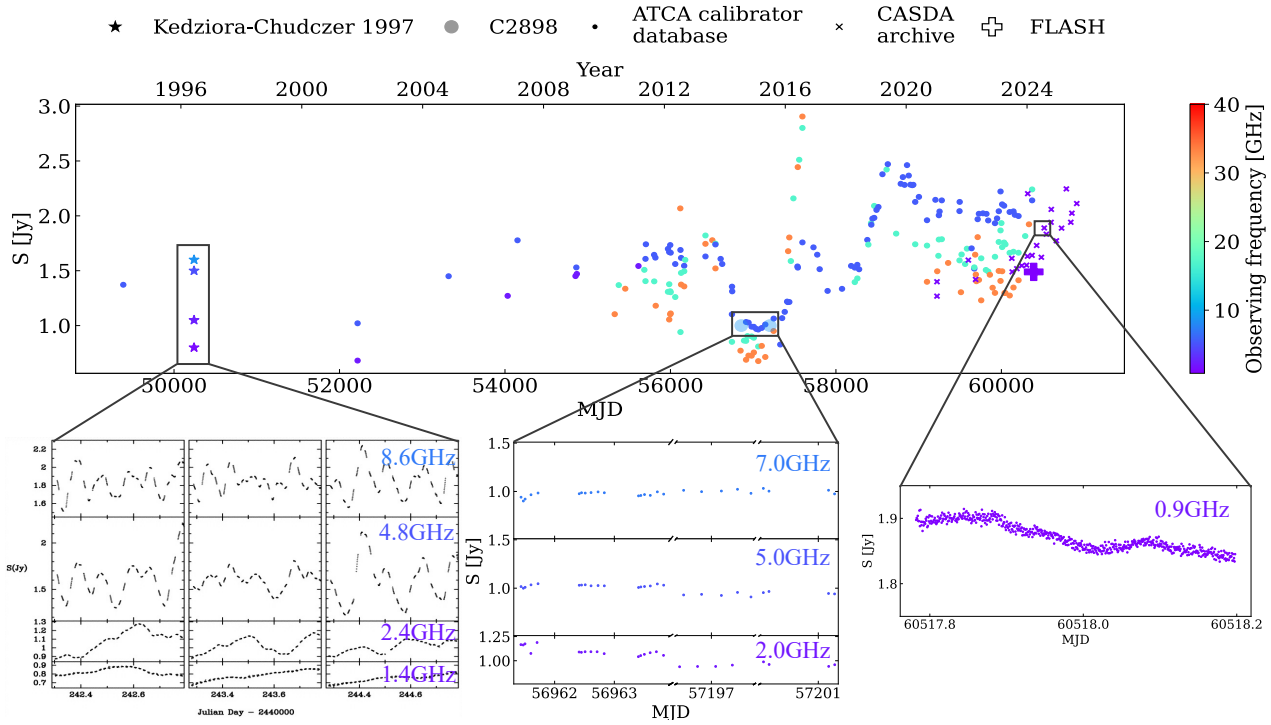


Figure 5.5: The radio lightcurve of PKS 0405−385 compiled from targeted monitoring programmes C007, and C1730 (filled circles) with the ATCA, labelled as 'ATCA calibrator database'. We additionally show the original, broadband fluxes from Kedziora-Chudczer et al. (1997) (stars), with an inset showing the IDV detected during those observations (bottom, left), as well as a later ATCA monitoring programme C2898 during which IDV was not observed (larger, semi-transparent circles, middle inset). Further, coincidental observations of PKS 0405−385 are taken from the CASDA archive (crosses), including the FLASH observations (filled vertical cross), and a 10-hour pointing observed as part of the Evolutionary Map of the Universe survey (EMU Norris et al., 2011).

[2] https://almascience.eso.org/alma-data/calibrator-catalogue

It is well established that radio flares in blazars are often accompanied by the ejection of a new parsec-scale jet component which can dominate the total flux density, and which can initially be sufficiently compact to produce IDV if a suitable scattering screen is present along the line of sight. Unfortunately, there was no IDV monitoring program in operation during the 2016 outburst to test this hypothesis, so broadband monitoring during and immediately after future flares would prove extremely useful in this regard.

At higher resolutions, Kedziora-Chudczer et al. (2001) derive an upper limit of 0.15 mas on the size of the core based on 8.4 GHz VLBA observations, which places an upper limit on the linear size of the core of 1.3 pc. Jet components are also visible in their 2.3 and 8.4 GHz images, extending up to 20 mas from the core, but the limited number of observations preclude a reliable estimate of apparent jet component speeds.

The linear polarisation of PKS 0405−385 was also studied from the period of IDV observed in 1996, and Rickett et al. (2002) determined the variations could be best explained by three, compact components forming an oblique source of $14 \times 20\,\mu$as at 4.8 GHz, corresponding to a linear size of approximately 0.3 pc at the source redshift. Current best estimates then put the PKS 0405−385 core at $0.3 - 1.3$ pc. At the redshift of the H I, the core emission has a linear extent of only ~ 1.5 pc then, easily subtended by a typical H I cloud which is thought to span ~a few parsecs in local ($z \lesssim 0.1$), extragalactic systems (Srianand et al., 2013; Gupta et al., 2018), or perhaps as much as ~ 10 pc in the analogous $z \sim 0.3$ quasar-DLA pair PKS 1127-2145 already discussed in Section 5.4 and presented in Kanekar & Chengalur (2001). Contemporary VLBI observations, ideally at or close to the frequency of the H I detection as in Aditya et al. (2025), will be crucial to understanding the structure of PKS 0405−385 , and may also provide insight into why it illuminates two H I structures with discrete velocities.

5.6 PKS 0405–385 revisited : propagation effects

In light of our new observations presented in Section 5.5, we consider here a few possibilities as to how the intervening matter from the three systems might affect the propagation of light from PKS 0405−385.

5.6.1 Could PKS 0405–385 be gravitationally lensed?

We now know there are three intervening galaxies along the line of sight to PKS 0405−385 close enough in angular separation to produce absorption lines in its spectrum. We might therefore consider whether the mass along this sightline is sufficient to gravitationally lens the background emissions of the blazar. For a source at $z_s \approx 1.284$ and lens $z \approx 0.882$, the critical column density of matter required for strong gravitational lensing – i.e., multiple imaging and/or significant magnification – to occur is $0.89\,\mathrm{g\,cm^{-2}}$, with our assumed cosmology.

The surface density of neutral hydrogen seen in the H I system is only $1 \times 10^{-3}\,\mathrm{g\,cm^{-2}}$, almost three orders of magnitude below that required for lensing. Since we cannot constrain the abundances of the gas in the intervening H I system from our current data, we cannot determine the complete gas surface density, let alone the total matter density along the line of sight. As a first approximation then, we can consider the total matter surface density in our own solar neighbourhood. McKee et al. (2015) put the local Galactic H I density at $\sim 2 \times 10^{-3}\,\mathrm{g\,cm^{-2}}$, remarkably close to the H I surface density detected in FLASH. They put the total matter surface density at $\sim 1 \times 10^{-2}\,\mathrm{g\,cm^{-2}}$, still two orders of magnitude below the gravitational lensing threshold if this were the density intersected at $z = 0.882$. Of course the matter distribution in the H I system might be entirely unlike our own Milky Way, and is almost certainly sampled by the PKS 0405−385 sightline at a different galactic radius, but in the absence of additional data it is impossible to say more. Furthermore, physical association between the three systems seen in absorption is unlikely, as they are separated along the line of sight by tens of Mpc (interpreting observed redshifts as cosmological) and in velocity by tens of thousands of $\mathrm{km\,s^{-1}}$ (assuming instead that redshift differences are due to peculiar motion). Therefore, there is currently no evidence for any additional mass contribution at group or cluster scales. In short it is unlikely that the light from PKS 0405−385 is lensed by the foreground systems, but a better understanding of their baryonic matter content will help to better constrain this problem.

5.6.2 What effect does Galactic scattering have?

Macquart (2005) showed that a multi-component H I profile could appear variable due to Galactic scattering and propagation effects. We cannot say anything of the variability in our H I system from one H I observation,

though followup during an episode of IDV would be particularly interesting to search for spectral line variability. Nevertheless, it is worth considering whether scintillation can offer further insights into the structure of PKS 0405−385, and the intervening gas.

In the original Kedziora-Chudczer et al. (1997) paper, the angular size of the PKS 0405−385 core was constrained to $< 5\,\mu$as, the Fresnel scale at which scintillation becomes significant for a scattering screen of Galactic plasma at a distance of 500 pc. However a screen at 30 pc is perfectly reasonable, and would require an angular diameter less than $20\,\mu$as; indeed screens at or below 10 pc have since been observed (Reardon et al., 2025; Wang et al., 2021b), further relaxing the source size constraints to $38\,\mu$as, corresponding to a linear size of ~ 0.3 pc at $z = 1.284$, similar to the size derived from VLBI in Section 5.5.3.

Of course, we now know from Section 5.5.2 that there are at least three intervening systems along the line of sight towards PKS 0405−385, each with their own ISM. Therefore Galactic plasma is not the only possible source of scattering (or angular broadening); we must also consider how the medium of the three intervening systems might contribute to the angular source size and variability.

5.6.3 Could there be scattering from intervening systems?

The theoretical breakthrough made by Macquart (2005) as mentioned in Section 5.6.2 was motivated by Kanekar & Chengalur (2001), who originally considered whether Interstellar Scintillation (ISS) caused by the ISM of an *intervening* galaxy might cause intraday fluctuations in both the background radio continuum, and intervening H I line profiles of a similar quasar-galaxy pair. We consider again whether the plasma in such intervening systems could contribute meaningfully to angular scattering, which would in turn minimise any observed ISS, and artificially increase the angular diameter of PKS 0405−385.

H I absorption traces a different phase of the ISM to that responsible for angular scattering (cold, neutral as opposed to ionised), but we can still use it to make a first order approximation on the expected level of scattering, provided that we assume some relationship between the two phases. We can derive one such relationship by looking at the pulse broadening of pulsars as a function of Galactic H I column density. Using the scattering time measurements from the Australia Telescope National Facility (ATNF) pulsar catalogue (Manchester et al., 2005) and a model of Galactic H I from Kalberla & Kerp (2009), a H I column of $\sim 4.2 \times 10^{20}$ cm^{-2} as seen towards PKS 0405−385 might produce a scattering angle $\alpha \sim 0.1 - 1$ mas at 1 GHz in our Galaxy, or something a factor of approximately 4 lower at $z = 0.882$, where it would correspond to intrinsically higher frequency and hence weaker scattering. At the frequencies at which IDV was observed and at the redshift of the FLASH detection, this drops to $\alpha \sim 1 - 10\,\mu$as, just below the threshold required to quench the IDV produced by the Galactic screen discussed in Section 5.6.2, or significantly affect the angular size of PKS 0405−385. Since the H I column density towards intervening systems 2 and 3 is even lower than the H I system, this framework would suggest they contribute an even smaller scattering angle to the light coming from PKS 0405−385. We must reiterate that the above is only a first order approximation of the effect of these intervening screens.

Once more, higher resolution optical spectroscopy would provide crucial insights into the mechanics of intergalactic scattering by allowing us to better constrain the multiphase gas along the line of sight. Coupled with further radio monitoring to detect new occurrences of IDV – potentially with a time-domain study of the H I profile – this extra data may offer new insights into plasma physics from cosmological distances to our own, Galactic neighbourhood.

5.7 Summary

A reconsideration of PKS 0405−385 shows it is in possession of a compact component $0.3 - 1.3$ pc based on both VLBI imaging and a better understanding of Galactic scattering. The linear scales probed by the ISS intra-day variability are 0.3 pc or less at the H I absorption frequency, so the IDV is likely to change across the HI absorption profile. Furthermore, long-term radio monitoring reveals several periods of rapid brightening indicative of episodic blazar activity from a compact core with structure seen on a scale of 1.3 pc with VLBI at 8.4 GHz. So, the HI absorption and its variability could be useful for ongoing studies of jet lifecycles.

Coincidentally, in an untargeted search for H I absorption conducted as part of the ASKAP-FLASH survey, intervening H I was detected towards PKS 0405−385 at $z = 0.882$, corresponding to the redshift of absorption

lines reportedly identified in the original optical spectrum of the source. Previous VLBI images and flares in ATCA monitoring suggest that the structure of the parsec-scale jet in PKS 0405−385 may have multiple components; a core and a bright jet component at the epoch of our FLASH observation could explain the two H I features separated by $\sim 45\,\mathrm{km\,s^{-1}}$ if each continuum component samples a different region of an intervening disk, or an extragalactic, clumpy medium. However, more recent VLBI observations would be required to confirm the presence of such structure today. We obtained Gemini GMOS spectroscopy towards PKS 0405−385 and the potential host of the H I gas to confirm the presence of the intervening system, and we identified the (likely) original, optical absorption lines at this redshift as the Mg II doublet, with additional Fe II and Mn II absorption features revealed in the new spectrum. However, we could not confirm Galaxy A as the host of the H I due to a lack of spectral resolution and signal-to-noise. Nevertheless, a number of other metal lines are also identifiable in the spectrum at this redshift, and we further identify the presence of two further, iron-rich intervening systems at $z = 0.907$ and $z = 0.966$, which are not currently detected in H I. The gaseous systems detected in intervening absorption are not likely to contribute to either IDV or scatter broadening of the background quasar. Scattering, and even interstellar scintillation in intervening galaxies *does* have a noticeable effect on Fast Radio Bursts, which are extragalactic sources with an extremely small diameter. Comparison to H I absorption in cases such as this one will be interesting, but is beyond the scope of this paper.

The evolution of metallicity at the redshifts probed by the intervening systems towards PKS 0405−385 is not well studied, largely due to a lack of optically-selected DLAs at these distances. This leaves the period just after cosmic noon critically under-explored, although we know star formation rates begin to decline here and gas distributions must therefore change (Madau & Dickinson, 2014). The case of PKS 0405−385 clearly demonstrates that radio selection of DLAs via intervening 21-cm absorption is a viable pathway to understanding metallicity evolution in this period. This technique will only grow in power with the progress of large-area, untargeted searches for H I in absorption such as FLASH. In all such cases though, the most interesting science can only be extracted from these systems with sufficient multiwavelength data. Repeat radio spectral observations will allow us to search for variability in the H I absorption features which has seldom been detected (Kanekar & Chengalur, 2001; Srianand et al., 2022; Allison et al., 2017), while optical spectroscopy with an Integral Field Unit is needed to properly constrain the redshifts of the handful of galaxies identified in our optical images to high precision. This will allow us to not only constrain the metal abundances for all intervening systems, but also determine the kinematic properties of the H I with higher precision (inflow, outflow, rotation), and spatially disentangle all the multiphase components of the gas haloes intersecting PKS 0405−385 (e.g. Péroux et al., 2019; Weng et al., 2022). In short, higher spatial and spectral resolution optical data is crucial to further disentangle the light of PKS 0405−385 and the intervening systems, in order to better understand the properties of the gas probed by this not-so-compact radio quasar.

5.8 Acknowledgements

The authors wish to thank Prof. Max Pettini for enlightening discussion and advice on the analysis of DLA absorption systems, Prof. Tom Oosterloo for pointing out data in the the ALMA calibrator catalogue, Dr. Mark Walker for helpful comments on the history surrounding PKS 0405−385, and Dr. Kimberly Emig for her helpful comments on a mature version of this manuscript. The authors also wish to thank the anonymous referee for their helpful comments, which improved the overall clarity of this work.

EFK is supported by an Australian Government Research Training Program (RTP) Scholarship.[3] HY is supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2025-00516062). MG is supported through UK STFC Grant ST/Y001117/1. MG acknowledges support from the Inter-University Institute for Data Intensive Astronomy (IDIA). IDIA is a partnership of the University of Cape Town, the University of Pretoria and the University of the Western Cape. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission

This scientific work uses data obtained from Inyarrimanha Ilgari Bundara, the CSIRO Murchison Radio-astronomy Observatory. We acknowledge the Wajarri Yamaji People as the Traditional Owners and native title holders of the Observatory site. CSIRO's ASKAP radio telescope is part of the Australia Telescope

[3] doi.org/10.82133/C42F-K220

National Facility (https://ror.org/05qajvd42). Operation of ASKAP is funded by the Australian Government with support from the National Collaborative Research Infrastructure Strategy. ASKAP uses the resources of the Pawsey Supercomputing Research Centre. Establishment of ASKAP, Inyarrimanha Ilgari Bundara, the CSIRO Murchison Radio-astronomy Observatory and the Pawsey Supercomputing Research Centre are initiatives of the Australian Government, with support from the Government of Western Australia and the Science and Industry Endowment Fund.

This paper includes archived data obtained through the CSIRO ASKAP Science Data Archive, CASDA.

Analysis in this paper is based on observations obtained under project GS-2024B-FT-215 (P.I. Yoon) at the international Gemini Observatory, a program of NSF NOIRLab, which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the U.S. National Science Foundation on behalf of the Gemini Observatory partnership: the U.S. National Science Foundation (United States), National Research Council (Canada), Agencia Nacional de Investigación y Desarrollo (Chile), Ministerio de Ciencia, Tecnología e Innovación (Argentina), Ministério da Ciência, Tecnologia, Inovações e Comunicações (Brazil), and Korea Astronomy and Space Science Institute (Republic of Korea).

This research uses services or data provided by the Astro Data Lab, which is part of the Community Science and Data Center (CSDC) Program of NSF NOIRLab. NOIRLab is operated by the Association of Universities for Research in Astronomy (AURA), Inc. under a cooperative agreement with the U.S. National Science Foundation. (Fitzpatrick et al., 2014; Nikutta et al., 2020; Juneau et al., 2021).

The DESI Legacy Imaging Surveys consist of three individual and complementary projects: the Dark Energy Camera Legacy Survey (DECaLS), the Beijing-Arizona Sky Survey (BASS), and the Mayall z-band Legacy Survey (MzLS). DECaLS, BASS and MzLS together include data obtained, respectively, at the Blanco telescope, Cerro Tololo Inter-American Observatory, NSF's NOIRLab; the Bok telescope, Steward Observatory, University of Arizona; and the Mayall telescope, Kitt Peak National Observatory, NOIRLab. NOIRLab is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation. Pipeline processing and analyses of the data were supported by NOIRLab and the Lawrence Berkeley National Laboratory (LBNL). Legacy Surveys also uses data products from the Near-Earth Object Wide-field Infrared Survey Explorer (NEOWISE), a project of the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration. Legacy Surveys was supported by: the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy; the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility; the U.S. National Science Foundation, Division of Astronomical Sciences; the National Astronomical Observatories of China, the Chinese Academy of Sciences and the Chinese National Natural Science Foundation. LBNL is managed by the Regents of the University of California under contract to the U.S. Department of Energy. The complete acknowledgments can be found at https://www.legacysurvey.org/acknowledgment/.

The Photometric Redshifts for the Legacy Surveys (PRLS) catalogue used in this paper was produced thanks to funding from the U.S. Department of Energy Office of Science, Office of High Energy Physics via grant DE-SC0007914.

6

Conclusions

Radio observations of young active galaxies – or peaked spectrum sources – open a unique window onto the lifecycle of AGN, their evolution across cosmic history, and their local jet-ISM interactions. Until very recently studies of these young active radio galaxies were limited to modest samples of no more that a few dozen, but the current generation of radio surveys is changing the way we think about population-scale studies of AGN. Continuum surveys are providing multi-band, multi-epoch coverage to characterise the SEDs of thousands of extragalactic sources, and spectral line surveys like ASKAP-FLASH make it possible to study their multi-phase environment with unprecedented coverage. In this way too, the 21 cm line, a familiar probe of neutral gas in the local Universe, is now also becoming an important probe of galaxy evolution at cosmic noon and beyond.

6.1 A summary of this thesis

This thesis covers new territory in the study of young active galaxies by using archival datasets, first to identify and then to understand them, all with a view to preparing for large-scale multi-wavelength studies made possible by current and future spectroscopic surveys. The main results of this work can be summarised as follows:

RadioSED: A new method for identifying PS sources from archival datasets

A major outcome of this research is the development of RadioSED, a new Bayesian inference tool for compiling and fitting radio SEDs of compact sources using large-area archival survey data. Though fitting of this kind has been performed for some time, we present the first open source code to both perform source association across different radio surveys and select the optimum model based on the Bayesian Evidence. In Chapter 2, we outline the performance of RadioSED, the default models implemented, and the range of surveys currently searched. We show that the results are reproducible, and that the models, though simplistic, are capable of matching a wide variety of radio spectral shapes. By making this tool available to the astronomical community, we intend to improve both the usage of and access to radio SEDs.

Discovering the PS sources within Stripe 82

In Chapter 3 we apply RadioSED to the equatorial Stripe 82 field to construct a new, uniform sample of PS sources. Owing to both the release of GLEAM-X DRII, and the ability of RadioSED to exploit survey detection limits as part of spectral fitting, we achieve a source density of PS sources of $\sim$1 deg^{-2}, the highest achieved to date for sources of this kind, and a direct result of our uniform search strategy. We show that using RadioSED we can increase the number of known PS sources by an order of magnitude, and again we stress that this is without need to obtain any data beyond that already available in the archives.

A multiwavelength study of this new sample of PS sources reveals that many are distant ($z \geq 1$), powerful active galaxies, with some residing as far as at least $z \sim 3$. We find that the most compact jets are located preferentially in quasar-type hosts, with galaxy-type hosts home to slightly more extended radio structures. We

obtain measures of their jet size via several different methods, including in broad terms the turnover-linear size relation (Jeyakumar, 2016). We also exploit a matched angular size between interplanetary scintillation (IPS) at ~200 MHz (Morgan et al., 2022), and visibility ratios at 20 GHz (Chhetri et al., 2013) to measure the fraction of their flux contained to ~0.3 arcsec scales at both frequencies. We show that while all PS sources detected in the 20 GHz catalogue are compact at high frequencies, some 30 per cent of PS sources detected in the 162 MHz IPS catalogue are only moderately compact, which we interpret as the presence of old, extended emission.

When this work was completed, the ASKAP-FLASH survey had not yet fully observed the Stripe 82 field, and so we did not search our new PS sample for H I absorption. However, we collated redshift constraints for over 45 per cent of the sample from either spectroscopy or photometric estimators, and we have also studied their WISE colours. This sample is therefore ideal for exploring the detection rate of H I absorption towards PS sources, where it is thought to be higher than amongst the general population of radio AGN (e.g. Geréb et al., 2014), whilst also considering the effects of things like dust reddening (Glowacki et al., 2019).

Studying the PS-H I connection through mock observations

While the ASKAP-FLASH survey progressed, we turned in Chapter 4 to cosmological simulations, to aid in our interpretation of result from the Pilot Surveys. Focusing on the high fraction (~70 per cent) of H I detections which were made towards PS sources (Yoon et al., 2025), we use a new package for mock observations to test whether our detection rates could be explained purely by the compact geometry of PS sources alone (as suggested by Curran et al. 2013), or whether jet-gas interactions would be additionally required. Our mock observation package, called Simulating Absorption of Neutral Gas for Radio Astronomy (SANGRiA), was developed specifically for this project, allowing us to exactly match our mock parameters to those of the ASKAP-FLASH survey. It is the first such package to produce mocks of 21 cm absorption.

In this chapter we study mocks of associated absorption, where the continuum backlight is produced by the jets of an active galactic nucleus associated with the same galaxy hosting the H I. We find that, although simple geometric arguments can explain some of the preferential detection towards compact jets, they are not sufficient to explain the high detection rates seen in the ASKAP-FLASH Pilot Surveys. We interpret this discrepancy between the pure, geometrically-motivated detection rate and the observed detection rate towards compact sources to be evidence that the most compact AGN jets reside in uniquely gas-rich environments. More broadly, the techniques established here for mock observations of 21 cm absorption have the potential to unlock many as yet unexplored synergies between radio observations and cosmological simulations.

Making the most of ASKAP-FLASH with multiwavelength datasets

Chapter 5 is a departure from the main line of PS-H I investigations which are the focus of this thesis. It arose from a discovery of H I absorption towards well-known radio quasar PKS 0405−385, first noticed during the regular course of monitoring incoming ASKAP-FLASH spectra. Unlike in the previous chapter, where our focus was H I absorption towards neutral gas co-located with the radio emission, this target offers us a window onto intervening gas that is not gravitationally bound to the radio emitting quasar.

An investigation of the archival radio data surrounding PKS 0405−385 revealed insights into the radio quasar itself. This quasar once displayed the most rapid known intraday variability at radio frequencies, from which it earned the title of 'the smallest radio quasar'. Although its size was revised upwards soon after based on updated scattering theory, PKS 0405−385 remains an important probe of Galactic plasma. We revisit the radio variability of PKS 0405−385 in light of advances in scintillation theory, and consider whether the original period of IDV could have been caused by the intervening gas now detected in H I. Although this is unlikely, the possibility remains interesting for the potential connection with studies of H I variability. We also make the first radio lightcurve of PKS 0405−385 using extended monitoring with the Australia Telescope Compact Array and the ASKAP, and find a revised linear size ≥ 0.3 pc, along with evidence of several radio flares, but no new evidence of repeating intraday variability.

By contrast, the ASKAP-FLASH detection and follow up optical observations provide information about the intervening systems along the sightline towards the quasar. We present new long-slit spectroscopy from GMOS on Gemini South, spanning both PKS 0405−385 and a candidate host of the intervening H I gas. We identify Mg II and Fe II absorption lines in this spectrum consistent with the redshift of the intervening H I

($z = 0.882$), as well as two additional, independent metal-line systems at $z = 0.907$ and $z = 0.966$, but we cannot accurately pinpoint the host(s) of this intervening gas in current data. Nevertheless, we expect ASKAP-FLASH to detect hundreds of intervening targets of this kind, and the Gemini follow up programme begun with PKS 0405−385 is now an integral part of the auxiliary data being compiled to better understand our ASKAP-FLASH detections.

6.2 Future work

Our new method for identification of PS sources has revealed that many young radio AGN out to $z \sim 3$ are detectable with current datasets, far more than those currently studied, offering a new window onto a unique phase in the AGN lifecycle. At the same time, our early work with mock radio observations of cosmological simulations has demonstrated that these young radio AGN are over-detected in H I absorption not just because of their compact geometry, but because their environments are likely gas-rich, hinting at potential links between the jets and the nuclear ISM. The final chapter of this thesis demonstrated the power of multiwavlength analysis in identifying not one but three intervening systems towards a radio bright quasar detected for the first time in H I absorption with ASKAP-FLASH. We consider next a few interesting avenues for future work, building on the findings presented here.

6.2.1 The cosmological evolution of PS sources

When first introducing PS sources back in Section 1.2, I mentioned that Slob et al. (2022) presented evidence for a population of PS sources which never evolve into extended radio galaxies based on their 144 MHz luminosity function. This is a promising step towards understanding the evolution of the PS population across cosmic time, since the luminosity function is a powerful and familiar tool for probing population-scale changes across redshift space (e.g. Mauch & Sadler, 2007; Best & Heckman, 2012). With only a modest amount of effort, the sample of 359 PS sources identified in Chapter 3 could be analysed in much the same way, but with several key improvements. First of all, where the Slob et al. (2022) luminosity function was restricted to only 138 PS sources due to incomplete redshift information, 97 sources in our Stripe 82 sample have spectroscopic redshifts, and the remainder all have photometric redshift estimates with informative probabilistic uncertainties measured using the Gaussian mixture model code PICZL (Roster et al., 2024; Roster et al., 2026). By more than doubling the number of sources we immediately improve the constraining power of the resulting luminosity function. Furthermore, since the Stripe 82 sample is fitted using RadioSED, it is simple to extract the radio luminosity of each source at any arbitrary frequency. This allows us to create a frequency-dependent luminosity function without any assumptions about spectral index, which we can then directly map to luminosity functions measured at different frequencies, from 144 MHz (Hardcastle et al., 2025), through the famous 1.4 GHz dataset (Best & Heckman, 2012) all the way to 20 GHz (Sadler et al., 2014). Such a multi-frequency analysis has not (yet) been attempted for PS sources, and it will be interesting to see what additional insights it offers.

6.2.2 PS sources in the early Universe

As the instantaneous sensitivity of instruments improves, so too does our ability to observe the early Universe. Already the James Webb Space Telescope is revealing massive galaxies out to higher redshifts than ever before (Labbé et al., 2023), and the MWA has successfully identified high redshift radio galaxies with the help of spectroscopic follow up (Capetti et al., 2025). However, many searches for high redshift radio galaxies pre-select 'ultra steep spectrum' (USS) sources to follow up with spectroscopy based on the so-called $\alpha - z$ correlation between redshift and spectral index (e.g. De Breuck et al., 2001; Broderick et al., 2007; Saxena et al., 2018). Some recent searches using GLEAM have relied on spectral curvature (specifically, spectral flattening) below $\sim$ 200 MHz as an alternative to a simple, two-point spectral index, though the underlying motivation is driven by similar empirical reasoning (Drouart et al., 2020), and an assumption that the steepening could trace changed environmental conditions at high redshift. As shown in Morabito & Harwood (2018) though, the observed $\alpha - z$ correlation is explainable with a combination of inverse Compton scattering off CMB photons (causing steepening) and a Malmquist bias where more distant sources in flux-limited samples are inherently more luminous (and more luminous AGN produce, on the whole, steeper integrated spectra; Chambers et al. 1990; Blundell & Rawlings 1999; Kutkin et al. 2025). This means that high−z searches pre-selecting USS

sources may have a higher than average success rate, but they are not well motivated by any underlying physical difference in the early Universe.

In Chapter 3 we showed that our Stripe 82 sample of PS sources contains young radio AGN at least as far as $z \sim 2$, and likely out to $z \sim 3$ or further based on their WISE photometry. Extending out to even higher redshifts, Zhai et al. (2025) demonstrate that the fraction of megahertz peaked quasars is constant out to $z \sim 4.8$ over a range of bolometric luminosities and BH masses. This suggests that PS sources could be another interesting sample to search for at high redshifts. Furthermore, if a suitable sample can be identified, it is a simple process to transform their radio SEDs into the source restframe and study the subsequent shape, which *can* inform us about the density of their local environment (discussed in Section 1.2.2), and thus the conditions of the early Universe.

Building towards such a sample would likely require an analysis akin to the one performed in Morabito & Harwood (2018) to understand the potential impact of selection effects on the detectability of PS sources at high$-z$. In addition to radio selection effects, any population-scale analysis at high$-z$ will also have to be mindful of optical selection effects. In our untargeted sample from Chapter 3, quasars dominated our PS sample beyond $z = 1.25$, and were the only host type detected at redshifts $z \geq 2$. They were also the only sources to have restframe peaks $\nu_{peak} \geq 3$ GHz indicative of extremely compact jets, but whether these two properties are related is not yet clear. In a similar vein, the Zhai et al. (2025) sample also focuses on the uniformity of quasar-type PS hosts out to $z \approx 4.8$. The lack of galaxy-type PS sources beyond $z \sim 1$ was first identified by Stanghellini et al. (1997) and there was some discussion several decades ago about whether PS quasars might represent a physically different class of objects to PS-type galaxies (see Pihlström et al., 2003, and references therein). A new search for high redshift PS sources making use of 8 m class optical telescopes for high sensitivity observations when selection effects are well understood, will be useful for revisiting this question, and for understanding the nature of young radio jets at high redshift more broadly.

6.2.3 ASKAP-FLASH observations and the PS-HI connection

At the time of writing ASKAP-FLASH is 89 per cent complete and is scheduled for completion in May 2026, with major data releases soon after that. There are many science cases to be addressed with the full ASKAP-FLASH dataset, but one of particular interest to me is the connection between H I absorption and broadband SED shape. A larger catalogue of ASKAP-FLASH detections, along with auxiliary data on their SED shape (whether from RadioSED or elsewhere), the optical redshift of any nearby galaxies (to determine whether the H I is associated with the radio continuum or intervening) and estimates of the gas covering fraction (from VLBI, as demonstrated in Aditya et al. *in prep.*) would allow us to address the following questions:

1. *What is the detection rate towards PS sources in an untargeted survey?* In the ASKAP-FLASH Pilot Surveys the detection rate towards PS sources was extremely high ($\sim$ 70 per cent), which motivated our work with mock observations in Chapter 4. If it remains this high in the larger H I samples of the full survey that would support the idea that these PS sources live in host galaxies which are particularly rich in cold, neutral gas.
2. *Are H I detections towards PS sources more likely to be associated, or intervening?* If, using spectroscopic redshift information, it turns out that many H I detections towards PS sources are instances of intervening H I, that may instead suggest that the compact nature of the jets simply makes them useful, point-like sightlines, easily subtended by intervening gas clouds.
3. *Is there any relationship between the broadband absorption mechanism, and the detection of H I?* Particularly for instances of associated H I, detailed modelling of the broadband SED below the turnover could reveal interesting insights into multiphase absorption. As discussed in Section 1.2.2, some PS sources show evidence of synchrotron self-absorption, while others are best explained by free-free absorption in a local, ionised medium. Studying these two absorption mechanisms - the broadband and the spectral line - would allow us to test for correlations between the neutral and ionised ISM probed by compact radio jets, and may offer interesting insights into the subset of the PS population embedded in cold, neutral gas.

6.2.4 High redshift HI

Putting together the ideas covered so far in this section, the ultimate test of our current instruments in this field is, I believe, a search for H I at even higher redshifts than ASKAP-FLASH, using PS sources for a targeted search. LOFAR is already capable of spectral line observations (e.g. Emig et al., 2019), and the spectral bandpass of the MWA has been improved by recent upgrades, making spectral line searches more tractable, though no work using this mode has yet been published. If suitable high redshift PS sources could be identified, they seem to be ideal targets to search for H I absorption at $z > 5$, given the apparent high detection rates towards them as far out as cosmic noon. Understanding more clearly the drivers of this high detection rate in H I absorption, the redshift evolution of PS sources themselves, and any observational biases in our experiment design will be key to conducting a search of this kind.

6.3 Final Remarks

This thesis comes at a time that may, with hindsight, be considered a period of paradigm shift in its own right. In 2015 there was the first, direct detection of gravitational waves from a binary black hole merger, with dozens following since (Abbott et al., 2017), and pulsar timing arrays are being used to search for a nanohertz gravitational wave background, thought to be produced by SMBH mergers, with tantalising hints of such a signature in the latest data (Agazie et al., 2024); we are now firmly in the era of not just multiwavelength, but multi-messenger astronomy. Theoretical and computational physicists are pushing cosmological models to new limits with machine learning (Villaescusa-Navarro et al., 2021), and artificial intelligence is fundamentally reshaping how we do science (Kusumegi et al., 2025). At the same time, new electromagnetic telescopes are directly imaging nearby SMBHs (Akiyama et al., 2019), and detecting massive galaxies out to higher redshifts than ever before (Labbé et al., 2023), revolutionising our understanding of galaxy formation across cosmic time. At radio wavelengths, the SKA is projected to begin science observations before 2030, promising to illuminate the Universe all the way out to the epoch of reionisation, but already its precursor instruments are revolutionising our understanding of young AGN, of H I at cosmic noon, and of how we do science in an ever growing sea of datasets. In a period of ever increasing complexity in our data, our models, and our understanding of the Universe, it is my hope that this thesis presents a useful path forward in radio studies of active galactic nuclei.

A

Is HI absorption more likely towards PS sources? An ASKAP-FLASH case study

Emily F. Kerrison[1,2,3], Elaine M. Sadler[1,2,3], Vanessa A. Moss[1,3], Elizabeth K. Mahony[3,2], and the FLASH team

[1]Sydney Institute for Astronomy, School of Physics A28, University of Sydney, NSW, 2006, Australia
[2]ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia
[3]ATNF, CSIRO Space and Astronomy, PO Box 76, Epping, NSW 1719, Australia

This chapter was accepted for publication in the *Proceedings of the International Astronomical Union*, Vol. S392 in October 2024 under the title *Why so young? A curious connection between the broadband flux and neutral gas content of AGN* with co-authors Kerrison, E. F., Sadler, E. M., Moss, V. A., and Mahony, E. K. doi:10.48550/arXiv.2412.08002

Neutral hydrogen may make up a "piffling" amount of the total matter in the universe, but its physical simplicity and the information it conveys about the gas content, distribution, kinematics and dynamics opens a powerful window on the evolution of galaxies, of the hierarchical structures in which they reside and of the Universe itself.

Haynes (2008), on the future of H I studies

This chapter in context

As discussed in Section 1.2.3, peaked spectrum sources are a favourite class of object to search for H I absorption because they are, by definition, compact, and many known examples are radio bright. This means that they are technically easy to observe, integration times can be relatively short whilst still reaching appreciable sensitivity, and cleaning and deconvolution is simpler than for sources with extended radio structures. Adding to their appeal, previous targeted searches for H I absorption have found a high detection rate towards PS sources; ~45 per cent of the PS sample in Gupta et al. (2006) are H I detected, this rises to ~55 per cent of the PS sources in Geréb et al. (2014). However, targeted searches are unavoidably prone to selection effects, and it is conceivable that this high detection rate is due to some underlying bias in the way that the PS sources (or indeed, the entire sample) was constructed.

The ASKAP-FLASH survey is the first of its kind: an untargeted search for H I absorption at $z = 0.4 - 1$ (Allison et al., 2022). This presents a unique opportunity to explore detection fractions in a truly unbiased sample of H I detections. At the time when this Chapter was written, the ASKAP-FLASH Pilot Surveys had been completed, from which the team identified 33 H I detections across ~3,000 deg^2 of sky (Yoon et al., 2025). This chapter examines the SEDs of those sources, with a particular focus on the PS AGN, which comprise 63 per cent of the sample. The SEDs discussed in this chapter were also presented individually as an appendix to Yoon et al. (2025).

Statement of contribution

This chapter faithfully reproduces material from Kerrison et al. (2024a), which was accepted for publication in the Proceedings of the International Astronomical Union, Vol. S392 in October 2024. The volume is still awaiting publication. I am the first author, with Elaine Sadler, Vanessa Moss, and Elizabeth Mahony as my co-authors. The research problem was conceived of by my supervisors, Elaine Sadler and Vanessa Moss, as one of the key motivations for my thesis. The analysis of the ASKAP-FLASH detections, and subsequent interpretation, is my own work. I drafted the proceeding, and all co-authors provided helpful comments and suggestions as to minor revisions.

Neutral Hydrogen in an around galaxies in the SKA era
Proceedings IAU Symposium No. 392,
D.J. Pisano, Moses Mogotsi, Julia Healy, Sarah Blyth, eds.
doi:xxx/xxxxx

Why so young? A curious connection between the broadband flux and neutral gas content of AGN

Emily F. Kerrison[1,2,3], Elaine M. Sadler[1,2,3], Vanessa A. Moss[3,1,2], Elizabeth K. Mahony[3,2] & FLASH team

[1]Sydney Institute for Astronomy, School of Physics A28, University of Sydney, NSW 2006, Australia
emily.kerrison@sydney.edu.au
[2]ARC Centre of Excellence for All-Sky Astrophysics in 3 Dimensions (ASTRO 3D)
[3]ATNF, CSIRO, Space and Astronomy, PO Box 76, Epping, NSW 1710, Australia

Abstract. We present here a study of the broadband spectral properties of 33 sources detected in HI absorption as part of the ASKAP-FLASH Pilot Surveys. We outline our approach to spectral classification and discuss the correlation seen between spectral shape and the detection of HI absorption. We further consider the implications of the observed correlation on the spatial distribution of the neutral gas, and on the jet-gas interactions. Our results are evaluated in the context of the forthcoming, full ASKAP-FLASH survey and other large, untargeted searches of the radio sky.

Keywords. radio continuum: galaxies, radio lines: galaxies, galaxies: nuclei

Although neutral hydrogen is both a key ingredient in cosmic star formation, and an excellent tracer of galaxy-scale interactions, very little is known about its distribution at intermediate redshifts. At $z \leq 1.7$ the 1215.7 Å Lyα line falls in the UV part of the spectrum and cannot be detected from the ground, and at $z \geq 0.4$ the 21-cm line is too faint to detect in emission with current instruments. One obvious way to fill this niche is through 21-cm absorption line studies, which allow us to detect column densities down to as low as $N_{HI} \sim 10^{19}\,cm^{-2}$ towards bright, background radio sources (Gupta et al. 2016).

The First Large Absorption Survey in HI (FLASH; Allison et al. 2022) is seeking to fill this niche, and is currently observing 24,000 deg^2 of the southern sky with the Australian SKA Pathfinder (ASKAP) to search for neutral hydrogen in absorption at intermediate redshifts ($0.4 < z < 1$). This is just after cosmic noon, where we notice a marked decline in star formation rate (Madau and Dickinson 2014), making this period critical to deepening our understanding of our Universe. This search is untargeted, and will provide us with thousands of new probes into the gaseous environments of galaxies and AGN in this under-explored redshift range. We present here a curious finding based on the broadband, radio properties of HI-detected sources in the ASKAP-FLASH Pilot Surveys.

1. A new, untargeted sample of HI absorbers

Our sample is composed of 33 radio sources detected in HI absorption as part of ASKAP-FLASH Pilot Surveys 1 and 2. These surveys, described in Yoon et al. (2024), covered 3,000 deg^2 of sky in a series of 2-hour pointings between 2019-2022, detecting over 10^4 sources in the continuum with $S_{855.5\,MHz} > 40$ mJy (the HI detection threshold of both pilot surveys). The sources in our sample were picked out from amongst these continuum sources using the Bayesian linefinder presented in Allison et al. (2012), which identified highly significant absorption lines in each 288 MHz-wide spectrum. These lines all have a Bayes factor greater than 30, corresponding approximately to a signal-to-noise ratio greater than 10 when integrating across the full linewidth. Our 33 radio galaxies therefore represent the strongest HI

absorption lines in an untargeted search for neutral hydrogen at intermediate redshifts, unbiased by optical selection effects such as dust reddening and obscuration, and we expect further examination of the FLASH pilot data to reveal additional lines at lower significance levels.

From amongst these 33 detections, three were previously known from ASKAP spectral line commissioning data, and detailed studies of these objects have already been published. These are PKS 1610-77 (an intervening line; Sadler et al. 2020), PKS 1740-517 (an associated line; Allison et al. 2015) and PKS 1830-210 (another intervening line first hypothesised to be from an intervening gravitational lens by Subrahmanyan et al. (1990), and then detected in the radio by Chengalur et al. (1999), with ASKAP data first presented in Allison et al. 2017). The line profiles of our sample span widths 20-200 km/s, and in the absence of optical spectroscopy, preliminary results from the machine learning methodology of Curran et al. (2016) suggest the sample contains a mixture of associated and intervening lines, as is to be expected from an untargeted search.

2. Broadband radio properties

By combining many independent flux density measurements of a radio AGN on a single plot, we can create its broadband spectral energy distribution (SED). In the IR-optical-UV regime, these SEDs are commonly used as a diagnostic tool for inferring the physical parameters of a host galaxy such as stellar mass and star formation rate (Pacifici et al. 2023). However here, we focus on the radio frequency portion of the SED, where the AGN component typically dominates and is unimpeded by dust obscuration. Through careful modelling of this radio SED component, we can determine if a source is young, typically within 10^3 years of triggering, or variable (i.e. a blazar). The youngest sources show a characteristic turnover in flux density over several decades in frequency, from which they get the name 'peaked spectrum' (PS) sources (O'Dea and Saikia 2021), while blazars will exhibit stochastic changes in flux density when considering multiple epochs of observations.

For each source in our sample, we have compiled flux densities from 80 MHz - 100 GHz using all large-area radio surveys currently available through the VizieR catalogue access service (Ochsenbein et al. 2000). The resulting SED of each source was then run through RADIOSED, a Bayesian fitting framework tailored to identifying young, peaked spectrum sources, as described in Kerrison et al. (2024).

There is a clear over-representation of PS sources from amongst this sample, with 21/33 sources statistically favouring a peaked model. Literature estimates suggest these comprise only 10-20% of all radio sources (O'Dea 1998; Callingham et al. 2017), so a fraction this high is surprising in a sample of absorbers from an untargeted search. In fact, assuming the upper limit from the literature, that 20% of all radio galaxies are PS sources (O'Dea 1998), the binomial probability of 21 or more HI-absorbed sources being peaked spectrum is only $\Pr(N_{\rm PS} \geq 21) = 3 \times 10^{-8}$, and this number falls if we assume the lower population fraction from Callingham et al. (2017). It is therefore clear that even though these HI absorbers are drawn from an untargeted survey, the galaxies against which we tend to detect neutral hydrogen are a specific subset that do not resemble the parent population in terms of their broadband radio characteristics.

3. Connecting the dots

Can we delve any further into the properties of this HI-absorbed radio population from the modest sample we have here? As a first step, we consider the apparent skew towards PS sources. To do this, we normalise all 21 PS SEDs by the location of their broadband peak and examine where the HI line falls relative to this, as shown on the left of Figure 1. Not only is there a preference for peaked spectrum sources amongst these HI detections, but there is also a tendency for the rest frame peak in these sources to fall just below 1.4 GHz, the frequency of

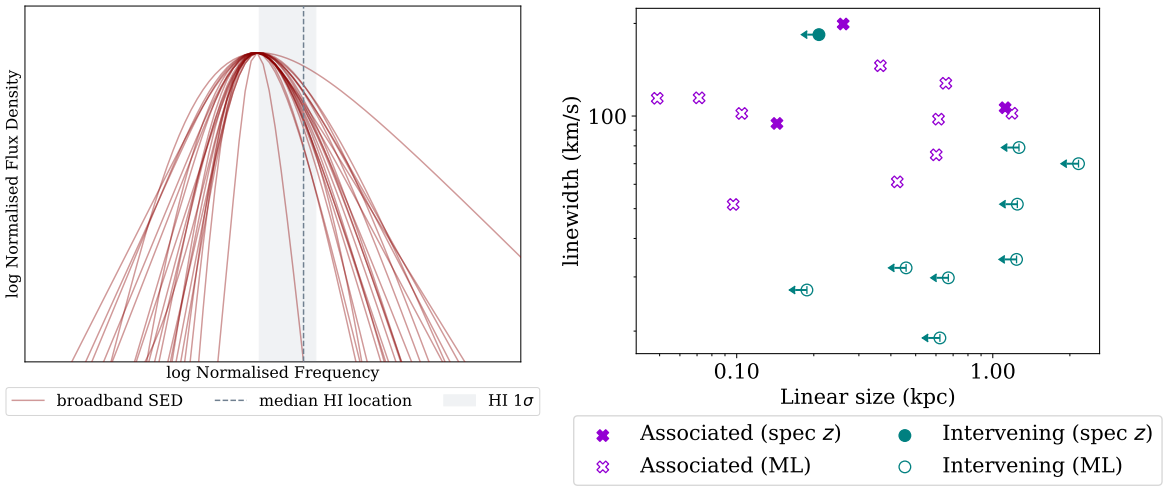


Figure 1. Left: the normalised SEDs of all 21 sources in our sample classified as peaked spectrum using RADIOSED. The vertical dashed line indicates the median position of all HI lines relative to the broadband SEDs, with the shaded band indicating the spread in this position using the interquartile range. Right: linewidth of the HI line profiles plotted against estimates of the linear size for each source, derived from its restframe peak frequency. Points are coloured by the classification of the HI detection as either associated with the radio source (purple) or intervening along the line of sight (teal). The marker style further differentiates the sources into those classified by a secure, spectroscopic redshift (filled), and those classified using machine learning (empty).

the HI line, assuming all cases of absorption are associated with the radio source. In this way the presence of the HI line allows us to place a lower limit on the rest frame peak frequency of the PS sources, as if the absorption line is not associated but from intervening gas, the SED of the radio galaxy will move further to the right in the rest frame of the source.

Going one step further, we can also use the relative positions of the HI line and the broadband spectral turnover to place limits on the linear sizes of the sources in our sample. It is well known that the peak of a PS source correlates closely with its linear size (e.g. O'Dea and Baum 1997; Jeyakumar 2016), so we here combine our estimates of rest frame peak frequencies with the empirical relation from Jeyakumar (2016) (especially Figure 3 of that work) to determine an approximate linear size for the PS sources in our sample. These estimates are plotted against the linewidth of the HI absorption profiles in the right of Figure 1, where we have further divided the PS sources of our sample into those identified as associated (purple) and intervening (teal), where the latter will only provide upper limits on the linear size of the background radio source, as well as into those securely identified as associated/intervening using optical spectroscopy (filled markers), and those identified in a preliminary application of machine learning (empty markers).

The width of the absorption line profile is not strongly correlated with the linear size of the radio source in this sample. However, we note the preliminary nature of the machine learning classifications, and we stress that larger samples will be required to more fully explore any connections between source structure and profile shape. A simpler observation with regards to this sample is that these sources all have estimated linear sizes between 0.05 – 2.2 kpc, with a mean size of 0.6 ± 0.5 kpc. This aligns with the high HI detection fraction seen at these sizes in inhomogeneous samples and supports the idea of a 'resonance' between the size of the absorbing screen and the radio source, as discussed in Curran et al. (2013).

4. Future work

Our investigation into the broadband radio spectral properties of the 33 HI detections from the untargeted ASKAP-FLASH Pilot Surveys is summarised here. We have observed a statistically significant preference for HI absorption to be detected towards young, peaked spectrum radio sources. Not only that, but the location of the spectral peaks in these sources suggest that they all have linear sizes less than ~ 2 kpc, matching the preference for compact sources seen in inhomogeneous, targeted samples of HI detections from the literature. It will be particularly interesting to see how these trends develop as we move into full survey mode, and the untargeted ASKAP-FLASH sample grows to hundreds, then thousands of absorbers. Alongside this growing sample, high-resolution VLBI imaging, combined with optical spectroscopy of the host galaxies will be key to better understanding the nature of the neutral gas distribution at these intermediate redshifts.

References

Subrahmanyan, R., Narasimha, D., Pramesh Rao, A., & Swarup, G. 1990, *Monthly Notices of the Royal Astronomical Society*, 246, 263–272.

O'Dea, C. P. & Baum, S. A. 1997, *The Astronomical Journal*, 113, 148.

Chengalur, J. N., de Bruyn, A. G., & Narasimha, D. 1999, *Astronomy and Astrophysics*, 343, L79–L82.

Ochsenbein, F., Bauer, P., Marcout, J., *et al.* 2000, *A&AS*, 143(1), 23–32.

Allison, J. R., Sadler, E. M., & Whiting, M. T. 2012, *PASA*, 29(3), 221–228.

Curran, S. J., Allison, J. R., Glowacki, M., *et al.* 2013, *Monthly Notices of the Royal Astronomical Society*, 431(4), 3408–3413.

Madau, P. & Dickinson, M. 2014, *ARA&A*, 52, 415–486.

Allison, J. R., Sadler, E. M., Moss, V. A., *et al.* 2015, *MNRAS*, 453(2), 1249–1267.

Curran, S. J., Duchesne, S. W., Divoli, A., & Allison, J. R. 2016, *MNRAS*, 462(4), 4197–4207.

Gupta, N., Srianand, R., Baan, W., *et al.* The MeerKAT Absorption Line Survey (MALS). In *Proceedings of MeerKAT Science: On the Pathway to the SKA — PoS(MeerKAT2016)* 2016,, Trieste, Italy. Sissa Medialab.

Callingham, J. R., Ekers, R. D., Gaensler, B. M., *et al.* 2017, *ApJ*, 836(2), 174.

Allison, J. R., Moss, V. A., Macquart, J.-P., *et al.* 2017, *Monthly Notices of the Royal Astronomical Society*, 465(4), 4450–4467.

Sadler, E. M., Moss, V. A., Allison, J. R., *et al.* 2020, *MNRAS*, 499(3), 4293–4311.

O'Dea, C. P. & Saikia, D. J. 2021, *A&ARv*, 29(1), 1–109.

Allison, J. R., Sadler, E. M., Amaral, A. D., *et al.* 2022, *PASA*, 39(1), 010–1.

Pacifici, C., Iyer, K. G., Mobasher, B., *et al.* 2023, *The Astrophysical Journal*, 944(2), 141.

Kerrison, E. F., Allison, J. R., Moss, V. A., *et al.* 2024, *Monthly Notices of the Royal Astronomical Society*,.

Yoon, H., Sadler, E. M., Mahony, E. K., *et al.* 2024, .

Jeyakumar, S. 2016, *Monthly Notices of the Royal Astronomical Society*, 458(4), 3786–3794.

O'Dea, C. P. 1998, *PASA*, 110(747), 493–532.

B

A Brief History of Astronomical Scintillation

Emily F. Kerrison[1,2,3], Ron D. Ekers[2], John Morgan[4], and Rajan Chhetri[4]

[1]Sydney Institute for Astronomy, School of Physics A28, University of Sydney, NSW, 2006, Australia
[2]ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia
[3]CSIRO Space and Astronomy, PO Box 76, Epping, NSW 1719, Australia
[3]CSIRO Space and Astronomy, PO Box 1130, Bentley, WA 6102, Australia

This chapter was accepted for publication in the *Proceedings of the International Astronomical Union*, Vol. S394 in December 2024 under the title *From terrestrial weather to space weather through the history of scintillation* with co-authors Kerrison, E. F., Ekers, R. D., Morgan, J., and Chhetri, R. doi:10.48550/arXiv.2412.19816

Even if the scientific spirit of observation and deduction (astronomy) has sometimes led to erroneous systems for predicting terrestrial events (astrology), we owe to the old astronomer and astrologer alike the deepest gratitude for their diligence in recording astronomical events.

Forbes (1909), on the history of astronomy

This chapter in context

Astronomy is an ancient science, indeed it predates the modern concept of 'science' by at least a millennium, though records of early astronomical measurements are fragmentary at best. In the course of my studies I have had the privilege to work with both classicists who read some of these ancient fragments in their original tongue, and astronomers who are interested in the ancient (and not so ancient) roots of this data-driven discipline under which we labour today. This chapter is a product of that environment - a *sphragis*, as a reader of the classics might call it - in that it captures something of my interest in the intersection between the history of science (both modern and ancient), and the new scientific productions of the present day.

Despite the unusual subject matter, this chapter also arose during the larger work of my thesis, much of which has focused on compact radio AGN. This work has, in general, been greatly strengthened by an ongoing collaboration between the ASKAP-FLASH team, and the Western Australian space weather group, who use both the MWA and ASKAP to measure the solar wind using the technique of interplanetary scintillation (IPS; Clarke, 1964). The connection comes as a by-product of their IPS measurements, from which they have produced catalogues of angular size measurements for large samples of radio sources within our field of view (Morgan et al., 2022; Chhetri et al., 2023), some of which were used in Chapter 3. But this chapter is focused on the history of IPS, and its connection to the broader history of scintillation studies, which we suggest extends all the way back to classical antiquity. This work originated in a serendipitous discussion with R. Ekers, from which a larger story of scientific progress grew with the help of co-authors. I hope the reader finds this small addition to the thesis as enjoyable to read as it was to put together.

Statement of contribution

This chapter faithfully reproduces material from Kerrison et al. (2024b), which was accepted for publication in the Proceedings of the International Astronomical Union, Vol. S394 in December 2024. The volume is still awaiting publication. I am the first author, with Ron Ekers, John Morgan, and Rajan Chhetri as my co-authors. The research problem was conceived of by myself and R. Ekers after a brief exchange over an excerpt from Aristotle's *De Caelo* (discussed in text). R. Ekers provided much of the early modern context surrounding scintillation studies, J. Morgan and R. Chhetri provided additional insights into the MWA data. I wrote the paper, co-authors provided comments on the text.

A multipoint view of the sun
Proceedings IAU Symposium No. 390,
doi:xxx/xxxxx

From terrestrial weather to space weather through the history of scintillation

Emily F. Kerrison[1,2,3], Ron D. Ekers[3], John Morgan[4] & Rajan Chhetri[4]

[1]Sydney Institute for Astronomy, School of Physics A28, University of Sydney, NSW 2006, Australia
emily.kerrison@sydney.edu.au
[2]ARC Centre of Excellence for All-Sky Astrophysics in 3 Dimensions (ASTRO 3D)
[3]CSIRO, Space and Astronomy, PO Box 76, Epping, NSW 1710, Australia
[4]CSIRO Space and Astronomy, P.O. Box 1130, Bentley, WA 6102, Australia

Abstract. Recent observations of interplanetary scintillation (IPS) at radio frequencies have proved to be a powerful tool for probing the solar environment from the ground. But how far back does this tradition really extend? Our survey of the literature to date has revealed a long history of scintillating observations, beginning with the oral traditions of Indigenous peoples from around the globe, encompassing the works of the Ancient Greeks and Renaissance scholars, and continuing right through into modern optics, astronomy and space science. We outline here the major steps that humanity has taken along this journey, using scintillation as a tool for predicting first terrestrial, and then space weather without ever having to leave the ground.

Keywords. history and philosophy of astronomy, (Sun:) solar wind, interplanetary medium

---

The scintillation of astronomical sources, or the stochastic variation in their phase and amplitude, is so well recognised it has its own nursery rhyme. Yet beyond the 'twinkle twinkle' of little stars, this phenomenon presents a useful probe of the scattering medium responsible for such variation. In the case of stars visible to the naked eye, this medium is our own atmosphere, while for radio frequency emitting pulsars and the active galactic nuclei (AGN) beyond our own Milky Way, it is typically the interstellar medium or the solar wind (where the effects are referred to as interstellar and interplanetary scintillation, respectively). It is this last instance, of scattering by the solar wind, that leads directly to predictions for space weather, but to fully appreciate the process of discovery required to reach this level of understanding, it is worthwhile looking back at how both individuals and cultures have used scintillation of various kinds as a type of remote sensing in its simplest form.

1. Proto-scientific scintillation

Some of earliest evidence for scintillation-based prediction has survived millenia through the oral traditions of Indigenous peoples in Alaska, South America and Australia. Despite the great length of time from then to now, and the lack of written document, we know that both the Yup'ik of Alaska and the Mocoví people of South America were aware of the twinkling of the stars (described in both cultures as "dancing" stars), while the Kamilaroi people of south eastern Australia describe them as 'laughing', and the Wardaman in the far north of the country say they are 'talking' (Hamacher 2019, and references therein). It is interesting that these disparate cultures, separated by centuries and languages, all draw on the language of human interaction to describe the phenomenon of stellar scintillation. Most interesting for our purposes though, is the fact that, in each of these cultures, these anthropomorphised stars are used as a tool for weather prediction, with stronger scintillation indicative of an impending storm. This has a sound basis in meteorological science, as Sofieva (2013) have shown that stellar scintillation can indeed trace atmospheric turbulence, including (presumably) the strong winds foreshadowing a storm.

Moving from the oral to the written, we see again references to stellar scintillation, and a possible connection with weather prediction, in the writings of the Ancient Greeks. Aristotle wrote of 'the apparent twinkling of the fixed stars and the absence of twinkling in the planets' in his *De Caelo* some time around 350 BCE, which is perhaps the first written record of the familiar adage "stars twinkle, planets don't" (though in truth this saying is not always accurate, see Fuller 2014). However, given the cosmological nature of this work, he does not tie this statement to weather prediction, instead (incorrectly) rationalising this observation on the basis that starlight has further to travel than planet-light, and so becomes "weak and wavering" by virtue of the distance traversed. To find predictive scintillation, we must look at a very different work by a poet Aratus, writing some 80 years later a psuedo-didactic poem called the *Phaenomena*. In this, he talks about the 'darkening disk' that falls upon stars before rain, and warns his readers to expect a storm 'when the bright light of the stars is dimmed ... and suddenly becomes wavering' (Kidd 1997). This looks very much like the weather prediction seen in the oral traditions of Indigenous peoples above, and suggests that the Greeks too recognised the connection between the strength of stellar scintillation, and the proximity of bad weather.

There are almost certainly other references to stellar scintillation and its use in weather forecasting which are strewn throughout the literature of our collective past. However, even the sample presented here clearly illustrates the longstanding use of twinkling stars as a tool for weather prediction on Earth.

2. Renaissance and modern science

Moving from the Ancient world into the Renaissance and finally into what we might reasonably call modern science, scintillation continues to appear sporadically in writings spanning both optics and astronomy. However at this point it is divorced from its connections to weather prediction. Instead, several famous thinkers attempted to explain the physics behind scintillation, but without much more success than Aristotle centuries before them. In the world of optics, da Vinci thought that scintillation was an optical illusion in the eye (Veltman 1986), an idea which had a curious resurgence in a 1949 *Nature* article (Hartridge 1949). Shortly after da Vinci, Tycho Brahe and Johannes Kepler, both studying the scintillation of supernovae, proclaimed it an intrinsic effect at the source. It was not until Hooke's *Micrographia* (1665), soon followed by Newton's *Opticks* in 1704, that scientists returned to the notion that stellar scintillation might be an extrinsic effect caused by the atmosphere millenia after it was used for weather prediction (Monaco 1990). Yet Newton's focus was on optics, and his explanation dealt only with the effect of aperture size (with larger apertures smoothing out the effects of scintillation). Further discussion of these more recent developments in scintillation can be found in Campbell (1991), but to return once more to the predictive powers of scintillation we need to switch both wavelengths, and sources.

3. Scintillation with radio eyes

From stars to AGN, and optical to radio wavelengths, we see scintillation used once more for remote sensing. Radio frequency scintillation probes three different media, each producing variations on different timescales and discovered in different ways. Ionospheric scintillation probes the upper layers of the atmosphere and manifests as variation on ~ 10 second timescales, interplanetary scintillation (IPS) probes the solar wind with variation on second timescales, and interstellar scintillation probes the ISM with variation over hours or days (a good summary of these processes can be found in Narayan 1992). It is the second of these which concerns us here, although all three have their own interesting tales of discovery.

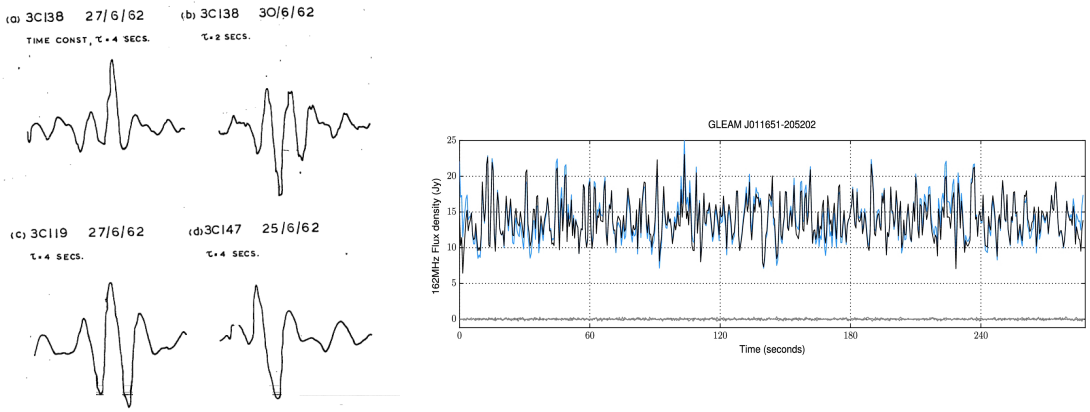


Figure 1. Left: The earliest recorded data exhibiting IPS (Clarke 1964), Right: A typical, modern IPS dataset taken with the Murchison Widefield Array (MWA) telescope with a 0.5 second sampling interval, from Chhetri (2018).

IPS was first discovered serendipitously by Margaret Clarke, a PhD student studying compact sources. It was published as an appendix to her thesis, in which she correctly identified not only the angular dependence of the phenomenon (that is, it is most visible in sources with angular scales $\theta \lesssim 1$ arcsecond), but also the solar corona as the scattering screen (Clarke 1964). Although her understanding was correct, it took some months for Tony Hewish, the local Cambridge expert on ionospheric scintillation, to come to terms with this explanation. Once he did, a *Nature* paper soon followed outlining, for the first time, the utility of IPS for making astrophysical measurements of "quasi-stellar" sources (Hewish 1964). Even so, a few months from discovery to comprehension is a remarkable feat in comparison to the slow growth of knowledge surrounding optical scintillation. Indeed our expertise in this field has grown exponentially, from Clarke's earliest observations (Figure 1, left) to the sophisticated multi-baseline interferometic measurements taken today (Figure 1, right).

Today, IPS is a powerful tool for tracking space weather, and there is an entire network of ground-based observatories spanning the globe, dedicated to making IPS measurements. The data are being used as input into heliospheric models to constrain solar wind parameters (e.g. Jackson 2023). They are used directly to constrain solar wind speeds (Mejia-Ambriz 2015), and they are used to detect and model coronal mass ejections (Morgan 2023) and stream interaction regions (Waszewski 2023). Critically, these measurements can be made from the ground, in some cases even using instruments which are shared with astronomical science (e.g. the Murchison Widefield Array (Tingay 2013) and Australian SKA Pathfinder, (Hotan 2021)), improving the efficiency of space weather tracking and of our scientific instruments more broadly. Recently, space weather IPS measurements have even returned to their roots, being used to study the compact sources behind the scintillating medium (Chhetri 2018).

4. Scintillation across epochs and frequencies

This whirlwind tour through the literature captures something of the many ways and many languages in which scintillation has been used, from the oral traditions of Indigenous peoples to the scientific papers of today. There are interesting parallels in the growth of our understanding as to the cause of scintillation, and in its predictive power that span both centuries and continents. It is through this historical appreciation of scintillation that we gain a greater appreciation of scintillation as a tool for space weather, and an example of global science at its best.

References

Campbell, L., 1991, *Stellar scintillation and its use in atmospheric measurements*, PhD Thesis, University of Adelaide, http://hdl.handle.net/2440/19571
Chhetri, R., Morgan, J.,*et al.*, 2018, *MNRAS*, 474(4), 3937-4955
Clarke, M., 1964, *Measuring radio source positions*, PhD Thesis, Cambridge University, https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.597742
Fuller, R.S., Norris, R. P., and Trudgett, M., 2014, *Australian Aboriginal Studies*, 2, 3-27
Hamacher, D. W., Barsa, J., *et al.*, 2019, *Proceedings of the Royal Society of Victoria*, 131(1), 24-33
Hartridge, H. and Weale, R., 1949, *Nat.*, 164, 999-1000
Hewish, A., Scott, P. F., and Wills, D., 1964, *Nat.*, 203, 1214-1217
Hotan, A.W., Bunton, J.D., *et al.*, 2021, *PASA*, 38, e009
Jackson, B.V., Tokumaru, M., *et al.*, 2023, *Solar Physics*, 298, 74
Kidd, D. (ed), 1997, *Aratus: Phaenomena*, Cambridge University Press
Mejia-Ambriz, J.C., Jackson, B.V., *et al.*, 2015, *Solar Physics*, 290, 2539-2552
Monaco, G., 1990, *Memorie della Società Astronomia Italiana*, 61, 819-827
Morgan, J., McCauley, P.I., *et al.*, 2023, *Space Weather*, 21, e2022SW003396
Narayan, R., 1992, *Phil. Trans. R. Soc. London, Ser. A*, 241, 151-165
Sofieva, V.F., Dalaudier, F. and Vernin, J., 2013, *Phil. Trans. R. Soc. London, Ser. A*, 371, 20120174
Tingay, S.J., Goeke, R.. *et al.*, 2013, *PASA*, 30, e007
Veltman, K.H., 1986, *Leonardo Studies II: Continuity and Discovery in Optics and Astronomy*
Waszewski, A., Morgan, J.S., Chhetri, R. *et al.*, 2023, *Space Weather*, 21, e2023SW003570

C

SED parameters for a selection of radio sources from the literature

We present here a table of the SED parameters for a selection of radio sources drawn from the literature, chosen for their representative SED shapes. Source names are from the RACS-DR1 catalogue, and literature values for model parameters (peak frequency: $\nu_{\mathrm{p,\,lit}}$, and spectral indices: $\alpha_{\mathrm{low,\,lit}}$, $\alpha_{\mathrm{high,\,lit}}$) are drawn from the citations listed. Note that these typically do not correspond exactly to optically thick and thin spectral indices α_{thick}, and α_{thin} as produced by RadioSED. We have not included a column for peak flux density from the literature since none of the works used as references provide these values. For the exact frequencies at which the literature spectral indices are calculated the reader is directed to the paper in column 12 (Refs), except for sources with parameters derived from the AT20G High Angular Resolution Catalogue (Chhetri et al., 2013), as that work reports several spectral indices at various intervals. In the case of those sources, we have chosen α_{low} to be the spectral index reported between $1 - 4.8$ GHz, and α_{high} to be the spectral index between $19 - 20$ GHz. All other model parameters are derived from RadioSED. Citations are not exhaustive for each source, but represent those works in which the source SED shape was first identified, and from which the literature parameters were derived. The citations in Column 11 are as follows: O98: O'Dea (1998), C13: Chhetri et al. (2013), C17: Callingham et al. (2017), S21: Stein et al. (2021), and A24: Aditya et al. (2024). Where a model parameter is unknown or not applicable to a particular source, that column is marked with a dash. Sources are ordered by RadioSED peak frequency where available, otherwise by α_{thin}. Those marked with an asterisk are classified as blazars, and are discussed in more detail in Chapter 2.

Table C.1: SED parameters for sources from the literature shown in Figure 2.14

Name	$\nu_{\rm p,lit}$ (GHz)	$\alpha_{\rm thick,lit}$	$\alpha_{\rm thin,\ lit}$	$\nu_{\rm p}$ (GHz)	$S_{\rm p}$ (Jy)	$\alpha_{\rm thick}$	$\alpha_{\rm thin}$	$\alpha_{\rm retrig}$	Model	Refs.
J225350.2-272434	-	-	-0.60	-	-	-	$-0.69^{+0.01}_{-0.01}$	-	2.1	S21
J231028.8-105426	-	-	-0.72	-	-	-	$-0.70^{+0.01}_{-0.01}$	-	2.1	S21
J224131.5-005144	-	-	-1.08	-	-	-	$-0.76^{+0.06}_{-0.05}$	-	2.1	S21
J014808.4-005546	-	-	-1.11	-	-	-	$-0.77^{+0.01}_{-0.01}$	-	2.1	S21
J093046.1-220708	-	-	-1.12	-	-	-	$-0.83^{+0.01}_{-0.01}$	-	2.1	S21
J231158.6-194733	-	-	-0.75	-	-	-	$-0.89^{+0.07}_{-0.07}$	-	2.1	S21
J232850.6-142146	-	-	-0.89	-	-	-	$-0.91^{+0.06}_{-0.03}$	-	2.1	S21
J205658.1-123732	-	-	-0.96	-	-	-	$-1.07^{+0.06}_{-0.04}$	-	2.1	S21
J040106.2+030057	0.325	1.71 ± 0.31	−0.61 ± 0.08	$0.33^{+0.02}_{-0.02}$	$0.34^{+0.01}_{-0.01}$	$2.37^{+0.28}_{-0.30}$	$-0.52^{+0.05}_{-0.04}$	-	2.2	S21
J203713.4-001057	0.325	0.73 ± 0.32	−0.74 ± 0.08	$0.41^{+0.03}_{-0.04}$	$0.86^{+0.03}_{-0.02}$	$1.01^{+0.12}_{-0.14}$	$-0.90^{+0.04}_{-0.04}$	-	2.2	S21
J231951.4+282829	0.325	0.4 ± 0.18	−1.06 ± 0.07	$0.46^{+0.07}_{-0.08}$	$2.57^{+0.22}_{-0.21}$	$0.44^{+0.18}_{-0.22}$	$-1.25^{+0.10}_{-0.10}$	-	2.2	S21
J044535.3-001901	>0.843	2.04	0.48	$0.55^{+0.04}_{-0.04}$	$0.58^{+0.01}_{-0.01}$	$1.63^{+0.14}_{-0.15}$	$-0.83^{+0.05}_{-0.04}$	-	2.2	C17
J212339.1-011235	0.325	0.5 ± 0.31	−0.61 ± 0.11	$0.65^{+0.02}_{-0.03}$	$1.85^{+0.04}_{-0.04}$	$0.53^{+0.03}_{-0.03}$	$-1.10^{+0.02}_{-0.02}$	-	2.2	S21
J034039.1+000534	0.325	0.41 ± 0.32	−0.53 ± 0.11	$0.71^{+0.15}_{-0.18}$	$0.22^{+0.03}_{-0.03}$	$0.28^{+0.11}_{-0.14}$	$-1.01^{+0.12}_{-0.12}$	-	2.2	S21
J002225.4+001456	1.4	0.31 ± 0.08	−0.82 ± 0.10	$0.81^{+0.02}_{-0.02}$	$4.22^{+0.04}_{-0.04}$	$1.06^{+0.03}_{-0.03}$	$-1.15^{+0.01}_{-0.01}$	-	2.2	S21
J161511.5-243059	1.4	1.29 ± 0.12	−0.80 ± 0.23	$1.07^{+0.03}_{-0.03}$	$0.94^{+0.01}_{-0.02}$	$1.82^{+0.06}_{-0.06}$	$-1.11^{+0.02}_{-0.02}$	-	2.2	S21
J002604.7-475617	0.213	-0.99	1.20	$0.24^{+0.01}_{-0.01}$	$0.99^{+0.01}_{-0.01}$	$1.45^{+0.06}_{-0.06}$	$-1.62^{+0.03}_{-0.03}$	-	2.3	C17
*J230659.4-652132	-	0.01	-1.11	$1.55^{+0.04}_{-0.04}$	$0.43^{+0.01}_{-0.01}$	$1.23^{+0.04}_{-0.04}$	$-1.90^{+0.04}_{-0.03}$	-	2.3	C13
J234708.6-185618	-	-0.13	-0.99	$1.82^{+0.07}_{-0.06}$	$0.60^{+0.01}_{-0.01}$	$1.25^{+0.07}_{-0.07}$	$-1.76^{+0.07}_{-0.07}$	-	2.3	C13
J002442.8-420202	-	0.06	-1.48	$2.03^{+0.01}_{-0.01}$	$2.62^{+0.01}_{-0.01}$	$2.55^{+0.04}_{-0.04}$	$-3.43^{+0.05}_{-0.05}$	-	2.3	C13
*J233159.4-381146	-	0.07	-0.43	$2.85^{+0.15}_{-0.15}$	$0.50^{+0.01}_{-0.01}$	$0.72^{+0.04}_{-0.04}$	$-0.92^{+0.04}_{-0.04}$	-	2.3	C13
*J231359.7-370445	-	0.30	-0.51	$3.69^{+0.13}_{-0.13}$	$0.29^{+0.01}_{-0.01}$	$0.81^{+0.03}_{-0.03}$	$-0.75^{+0.04}_{-0.03}$	-	2.3	C13
*J230738.6-224753	-	0.60	-0.61	$5.80^{+0.12}_{-0.11}$	$0.71^{+0.02}_{-0.02}$	$1.21^{+0.03}_{-0.03}$	$-1.82^{+0.06}_{-0.06}$	-	2.3	C13
*J012528.8-000555	>0.843	0.46	0.27	$12.69^{+0.81}_{-0.84}$	$1.59^{+0.02}_{-0.03}$	$0.26^{+0.01}_{-0.01}$	$-0.77^{+0.02}_{-0.02}$	-	2.3	C17
J053354.5-234429	>0.843	0.71	0.29	$0.63^{+0.01}_{-0.01}$	$1.82^{+0.03}_{-0.03}$	$0.84^{+0.03}_{-0.02}$	$-0.77^{+0.02}_{-0.02}$	$-0.37^{+0.05}_{-0.04}$	2.4	C17
J010551.0+155342	1.4	0.64 ± 0.1	−0.75 ± 0.14	$0.70^{+0.02}_{-0.02}$	$1.40^{+0.02}_{-0.02}$	$1.38^{+0.03}_{-0.03}$	$-0.52^{+0.09}_{-0.08}$	$-0.66^{+0.47}_{-0.30}$	2.4	S21
*J015310.1-331025	-	-0.43	0.38	$0.95^{+0.05}_{-0.03}$	$1.20^{+0.01}_{-0.01}$	$0.99^{+0.10}_{-0.06}$	$-0.32^{+0.01}_{-0.01}$	$-0.50^{+0.09}_{-0.06}$	2.4	C17
J044133.7-334004	-	-0.48	0.75	$1.49^{+0.02}_{-0.02}$	$1.14^{+0.01}_{-0.01}$	$0.71^{+0.04}_{-0.03}$	$-0.86^{+0.01}_{-0.01}$	$-2.29^{+0.18}_{-0.28}$	2.4	C17
*J135706.0-174401	-	-0.27	0.25	$1.93^{+0.06}_{-0.06}$	$1.50^{+0.03}_{-0.03}$	$0.28^{+0.02}_{-0.02}$	$-0.87^{+0.02}_{-0.02}$	$-1.15^{+0.30}_{-0.18}$	2.4	C17
*J212912.1-153840	-	-0.51	0.34	$6.93^{+0.22}_{-0.23}$	$1.52^{+0.06}_{-0.07}$	$0.67^{+0.01}_{-0.01}$	$-0.77^{+0.11}_{-0.10}$	$-0.70^{+0.08}_{-0.08}$	2.4	C17
*J213638.5+004154	4.3	-	-	$7.73^{+0.33}_{-0.26}$	$9.76^{+0.28}_{-0.31}$	$0.75^{+0.01}_{-0.01}$	$-0.85^{+0.02}_{-0.02}$	$-1.36^{+0.08}_{-0.08}$	2.4	O98
*J225805.9-275821	-	-	-	$61.82^{+0.01}_{-0.01}$	$5.26^{+0.03}_{-0.03}$	$0.39^{+0.01}_{-0.01}$	$-0.57^{+0.01}_{-0.01}$	$-0.94^{+0.02}_{-0.02}$	2.4	A24

D

Radio SED parameters for PS sources in Stripe 82

We present here the fit parameters from RadioSED for our new sample of PS sources presented in Chapter 3. The table here contains the first 14 entries, but the full catalogue is available online.

The radio SED parameters for a small subset of the PS sources discussed in Chapter 3. Column 1 is the IAU designation for the source, Columns 2 and 3 are its coordinates in decimal degrees. Column 4 is the classification of the source based on its radio SED, either "PS" for Peaked Spectrum, or "SPS" for Soft Peaked Spectrum sources (a historical distinction that is discussed in Section 3.3.1). Columns 5 and 6 show the Best Model amongst those used by RadioSED and as outlined in Paper I, and the Bayes factor comparing this best model, and the next most likely one. Column 7 is the flux density of the broadband peak in Jy, and column 8 the corresponding frequency in MHz. Column 9 is the spectral index below this peak frequency, column 10 the spectral index above. Column 11 is the spectral index of the low frequency upturn (where this is detected in a source). Columns 12 and 13 relate to what we call the "trough", the point in the SED below which a low frequency upturn occurs, with Column 12 giving the flux density of this point (again in Jy), and Column 13 the frequency (in MHz).

Table D.1: Radio SED parameters for a subset of PS sources discussed in Chapter 3. The full version of this table can be found in the online supplementary material for the relevant publication.

(1) IAU Designation	(2) RA (deg)	(3) Dec. (deg)	(4) SED Class	(5) Best Model	(6) Best Model $\log_{10} Z$	(7) $S_{\rm peak}$ (Jy)	(8) $\nu_{\rm peak}$ (MHz)	(9) $\alpha_{\rm thick}$	(10) $\alpha_{\rm thin}$	(11) $\alpha_{\rm retrig}$	(12) $S_{\rm trough}$ (Jy)	(13) $\nu_{\rm trough}$ (MHz)
J000001.6-002209	0.0068	-0.3692	SPS	Snellen	4.36	$0.28^{+0.001}_{-0.001}$	79^{+1}_{-1}	$3.96^{+0.07}_{-0.03}$	$-0.23^{+0.0}_{-0.0}$			
J000247.3+003111	0.6972	0.5198	PS	Orienti	32.84	$0.03^{+0.001}_{-0.001}$	217^{+27}_{-30}	$0.86^{+0.26}_{-0.24}$	$-1.53^{+0.16}_{-0.17}$			
J000332.1-011026	0.8839	-1.174	PS	Orienti	29.13	$0.02^{+0.001}_{-0.001}$	821^{+63}_{-51}	$2.72^{+0.37}_{-0.34}$	$-0.68^{+0.16}_{-0.16}$			
J000452.3-003547	1.218	-0.5965	PS	Snellen	26.78	$0.17^{+0.004}_{-0.004}$	90^{+3}_{-3}	$2.36^{+0.45}_{-0.45}$	$-0.64^{+0.02}_{-0.02}$			
J000541.7+004918	1.4239	0.8217	SPS	Snellen	32.03	$0.03^{+0.002}_{-0.003}$	735^{+180}_{-121}	$0.32^{+0.06}_{-0.11}$	$-1.31^{+0.32}_{-0.38}$			
J000544.7+003558	1.4364	0.5995	PS	Snellen	29.48	$0.03^{+0.002}_{-0.004}$	979^{+203}_{-157}	$0.74^{+0.1}_{-0.19}$	$-1.64^{+0.61}_{-0.39}$			
J000741.3+003100	1.9222	0.5167	SPS	Snellen	34.91	$0.03^{+0.002}_{-0.003}$	985^{+287}_{-186}	$0.33^{+0.06}_{-0.12}$	$-0.95^{+0.35}_{-0.36}$			
J001009.9+005440	2.5414	0.9112	SPS	Snellen	33.24	$0.05^{+0.003}_{-0.006}$	1394^{+182}_{-75}	$0.31^{+0.04}_{-0.08}$	$-2.0^{+1.11}_{-0.6}$			
J001038.1-004559	2.6589	-0.7665	SPS	Retriggered	32.43	$0.04^{+0.009}_{-0.005}$	491^{+66}_{-55}	$0.46^{+0.41}_{-0.2}$	$-0.93^{+0.16}_{-0.22}$	$-0.83^{+0.3}_{-0.51}$	$0.03^{+0.001}_{-0.001}$	$186.0^{+22.0}_{-17.0}$
J001052.4-002006	2.718647	-0.335062	SPS	Snellen	0.94	$0.09^{+0.008}_{-0.007}$	1093^{+270}_{-167}	$1.78^{+1.08}_{-0.5}$	$-0.49^{+0.12}_{-0.1}$			
J001055.6+005951	2.7318	0.9976	PS	Snellen	25.39	$0.12^{+0.003}_{-0.004}$	84^{+3}_{-3}	$2.82^{+0.63}_{-0.6}$	$-0.62^{+0.02}_{-0.02}$			
J001130.3+005752	2.8764	0.9645	SPS	Retriggered	34.18	$0.18^{+0.01}_{-0.008}$	2537^{+679}_{-337}	$0.53^{+0.03}_{-0.03}$	$-0.24^{+0.13}_{-0.16}$	$-0.85^{+0.14}_{-0.19}$	$0.07^{+0.002}_{-0.002}$	$240.0^{+20.0}_{-18.0}$
J001513.8-000337	3.8076	-0.0603	PS	Retriggered	30.85	$0.05^{+0.016}_{-0.006}$	471^{+45}_{-51}	$0.97^{+0.48}_{-0.27}$	$-2.29^{+0.41}_{-0.44}$	$-0.93^{+0.2}_{-0.35}$	$0.02^{+0.001}_{-0.001}$	$159.0^{+18.0}_{-19.0}$
J001611.0-001511	4.0459	-0.2531	SPS	Snellen	-218.43	$0.92^{+0.004}_{-0.004}$	410^{+10}_{-10}	$0.72^{+0.01}_{-0.01}$	$-0.41^{+0.0}_{-0.0}$			

E

Additional radio parameters for PS sources in Stripe 82

We present here the additional radio parameters for the PS sample in Chapter 3, including the debiased VI obtained from VAST (Table E.1), and the NSI obtained from MWA-IPS measurements (Table E.2). The tables here contain the first 10 entries, but the full catalogue is available online.

In Table E.1, the columns are as follows: Column 1 is the IAU name of the source, Column 2 the VAST ID of the source against which each PS object is matched. Columns 3 and 4 are the minimum and maximum signal to noise (SNR) for this source obtained across the VAST Pilot images which contain it. Column 5 is the average flux of the source across all VAST Pilot data. Column 6 is the separation between the position of each object as given in Table D.1 and given by the VAST pipeline. Columns 7 and 8 are the debiased variability index (VI) obtained when using the integrated radio flux density of the source, and the peak radio flux density, respectively.

Table E.1: The variability parameters for our sources as obtained from the VAST Pilot surveys. The full version of this table can be found in the online supplementary material for the relevant publication.

(1) IAU Designation	(2) VAST ID	(3) VAST SNR min.	(4) VAST SNR max.	(5) VAST avg. flux (mJy)	(6) VAST separation (arcsec)	(7) Debiased VI (int. flux)	(8) Debiased VI (peak flux)
J035045.7-010848	3937280	58	120	20.72	0.62	0.3008	0.2859
J001130.3+005752	3587332	73	631	201.28	0.80	0.2538	0.2389
J022213.8-010148	3866373	39	134	37.53	0.76	0.2174	0.2056
J035219.8+011421	3868128	87	320	68.32	0.71	0.1958	0.1910
J031505.0+011001	3748162	24	128	27.17	0.57	0.1614	0.1866
J031357.0+003507	4074157	36	132	32.04	0.23	0.1818	0.1720
J224423.5-001712	3389059	41	72	17.21	0.19	0.1444	0.1669
J211206.6+001732	3870471	106	202	37.35	0.41	0.1518	0.1653
J222744.5+003450	4078633	118	221	44.07	1.22	0.1250	0.1484
J031600.9+005854	3528910	10	31	8.70	0.26	0.1641	0.1442

In Table E.2, the columns are as follows: again, Column 1 is the IAU designation of each object as it appears in Table D.1, and Column 2 is the solar elongation of the object during the observations (further from the Sun generally means scintillation will be weaker, even for the most compact sources). Column 3 is the IPS class of an object, whether it is a secure detection ('detected'), a marginal detection, or only an upper limit on scintillation can be calculated. Column 4 is the actual measure of scintillation, NSI, and column 5 its associated error. Columns 6 and 7 are the upper limit on NSI and associated error, where these are calculated instead of a secure NSI. Column 8 is the likelihood that the NSI is non-zero, column 9 that the true NSI is not unity. Column 10 is the number of measurements used in calculating the NSI, and column 11 the effective number of observations taking into account the weighting of each observation (this is further explained in Morgan et al. 2022). Column 12 is the number of 5-σ detections of each object in the variability image.

Table E.2: The IPS parameters of our sources. Most columns are as appear in Morgan et al. (2022).

(1) IAU Designation	(2) solar elongation (deg)	(3) IPS class	(4) NSI_{fit}	(5) err	(6) UL_{fit}	(7) err	(8) $\mathcal{L}_0$	(9) $\mathcal{L}_1$	(10) N_{fit}	(11) N_{eff}	(12) N_{MOM}
J000001.6-002209	33.09652	detected	1.05	0.04			3552.89	25.17	50	41.5	18
J000452.3-003547	32.14316	detected	1.10	0.07			217.27	6.41	24	21.8	2
J001055.6+005951	29.859293	detected	1.15	0.09			166.49	9.09	16	14.6	2
J001611.0-001511	31.059143	detected	1.07	0.04			46439.87	378.38	67	51.9	46
J002225.4+001456	31.17189	detected	1.10	0.04			439445.20	5723.15	77	56.5	68
J004606.6-004341	32.35645	detected	1.07	0.03			322621.70	2278.25	87	63.5	73
J004631.3-002023	31.81639	marginal	0.63	0.15			5.75	11.91	6	5.4	0
J005205.5+003538	33.09957	marginal	0.47	0.07			14.22	163.63	35	30.4	0
J005716.9-002432	32.953613	upper_limit			0.27	0.19	0.28	39.55	9	8.6	0
J005905.4+000652	31.804417	detected	0.96	0.03			5145384.50	10120.92	83	59.0	83

F

Alternative population compositions for SANGRiA Forward Modelling

In Section 4.6.2, we show the relative fractions of the jet population in our entire, re-weighted 'Jet rotation' sample, as well as the relative fractions of AGN which are H I-detected. To test the effect of our parent population weights more rigorously we consider a few alternative models here.

Figure F.1 provides three alternative parent populations and the resulting H I detected population, again drawn from the 'Jet rotation' simulations. On the left, we assume a simple, two-component population comprising 13 per cent FR0s, and 87 per cent 10 kpc FRIIs. This represents a scenario where the compact, FR0 jets remain confined for some non-negligible amount of time, and then evolve rapidly upon initial breakout from the host galaxy before reaching a slowly evolving state at a length of approximately 10 kpc. As a result, the vast majority of the observed population are either in their compact phase or a fully extended phase. This model also assumes that more extended structures, like the 50 kpc jets, are at a low surface brightness and are not detected in the ASKAP-FLASH observations. The resulting detection rate towards compact sources (25%) is not appreciably different to the rate we find in Section 4.6.2 where the extended jets are more evenly distributed amongst different lengths. The same assumptions hold for the middle panel of Figure F.1, except we allow for some of the 50 kpc jets to be luminous enough for detection in the ASKAP-FLASH observations. This boosts the compact source fraction of H I detections to 36 per cent, but it is still significantly outside the 95 per cent confidence interval for the observed compact H I detection rate.

The third and final alternative considered here is that the rapid expansion phase takes place over even larger length scales, so that the observational sample can be described by only compact, FR0s, and 50 kpc FRIIs, the most extended jets in our simulations. This approximately mimics the size distribution of Horton et al. (2025). Under this assumption, it is possible to recover the high rate of H I detections towards compact jets (70 per cent) from geometric arguments alone. However, observational samples show that this is not a good characterisation of the length distribution of radio galaxies; the Horton et al. (2025) is biased towards large galaxies because they require an angular size (> 90 arcsec) for galaxies in their sample to be resolved by LOFAR. By comparison, DiPompeo et al. (2013) examine a number of samples and show that they contain exponentially more compact sources than extended (> 50 kpc) ones. Further, in the Norris et al. (2025) ASKAP sample, 88 per cent of sources have angular sizes < 90 arcsec, though at present their catalogue does not have accompanying redshift information with which to derive linear sizes, this is planned for a future work.

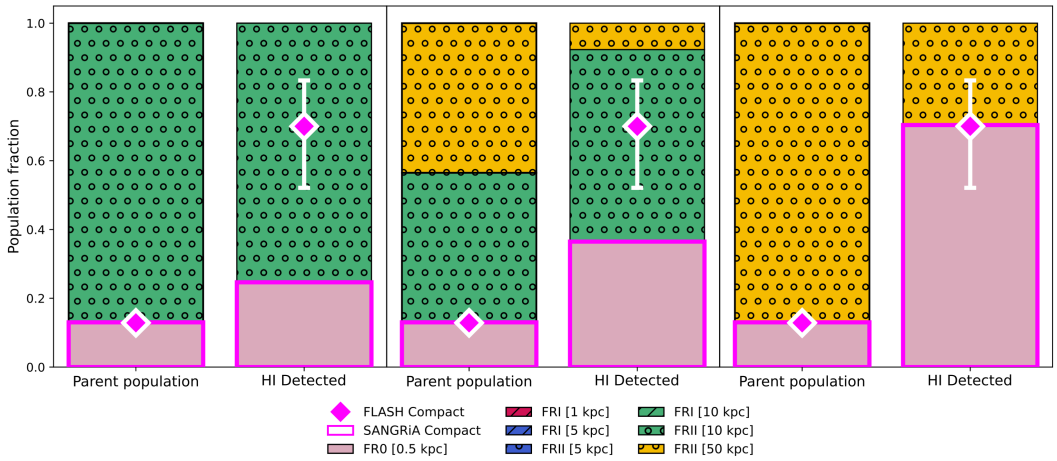


Figure F.1: Predicted detection fractions for simplified jet populations, assuming only FR0 (compact) and FRII (breakout) jets. The compact fraction is set to 13% (Kerrison et al., 2025). The assumed length of the breakout jets are varied across the three panels: *left*: all 10 kpc; *middle* 50:50 10 kpc and 50 kpc; *right*: all 50 kpc. Only in the extreme case of assuming all breakout jets are 50 kpc in radius does the predicted compact HI detection fraction match the observed value. This demonstrates that our conclusion that such high compact fractions cannot be achieved by inclination alone holds for all but the most extreme assumptions about the parent jet population.

G

Rotating jet and galaxy results from SANGRiA Forward Modelling

For completeness, we include here a version of Figures 4.5 and 4.13 for our 'Galaxy rotation' suite of simulations, where the jet and galaxy are fixed perpendicular to one another, then rotated through 180 degrees. These are shown in Figure G.1. Overall, the same trends seen in the 'Jet rotation' simulations discussed in the main text are seen here again; compact sources contribute the vast majority of H I detections at low ($L_{1.4\mathrm{GHz}}$<10^{26} W Hz^{-1}) radio luminosities, but the overall fraction of compact sources amongst H I detections does not reach that seen in ASKAP-FLASH Pilot Survey observations.

The biggest difference is for the 50 kpc FRII jets – there are now some geometries where the HI is never detected, no matter how bright the luminosity. This is reflected in the dotted yellow line in the left-hand panel never reaching 1. This has a small effect on the detected fractions (right panel), pushing up the detected compact fraction in the rightmost bar chart. However, this effect only increases the compact fraction from 27 per cent to 34 per cent which is still significantly less than the observed value of 73 per cent. Thus our conclusion that geometry alone cannot explain this observed compact fraction still holds even in this alternative rotation scenario.

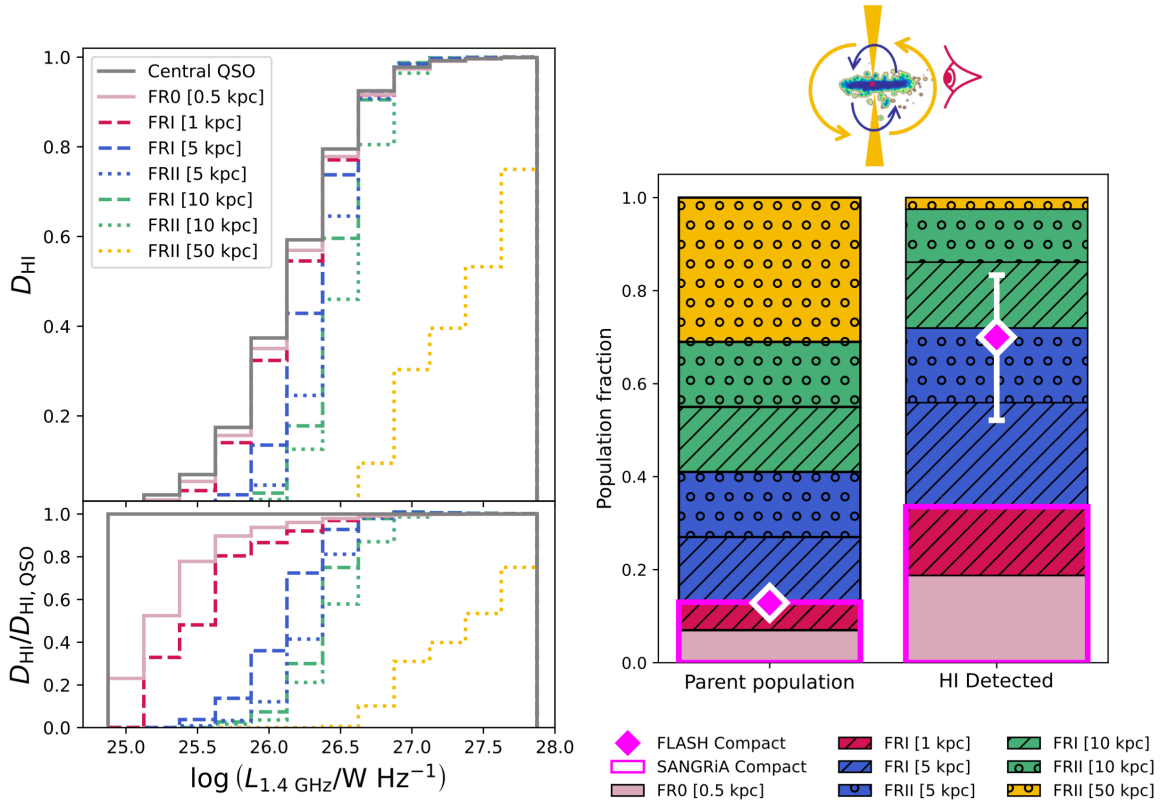


Figure G.1: *Left*: The HI detection fraction as a function of luminosity (see Figure 4.13) and *Right*: The ratio of jet types detected in HI compared to the overall parent population (see Figure 4.15). These are now calculated for 'Galaxy rotation' simulations, and the results are broadly similar to 'Jet rotation' case presented in the main body of the paper. The most significant differences are seen in the 50 kpc FRII detection rate at high luminosities, and the overall detection rate of compact sources.

Bibliography

Abazajian K. N., et al., 2009, ApJS, 182, 543

Abbott B. P., et al., 2017, Phys. Rev. Lett., 119, 161101

Abdo A. A., et al., 2010, ApJ, 716, 30

Abdollahi S., et al., 2023, ApJS, 265, 31

Ade P. A. R., et al., 2014, A&A, 571, A28

Adelberger K. L., Shapley A. E., Steidel C. C., Pettini M., Erb D. K., Reddy N. A., 2005, ApJ, 629, 636

Aditya J. N., 2019, MNRAS, 482, 5597

Aditya J. N. H. S., Jorgenson R., Joshi V., Singh V., An T., Chandola Y., 2021, MNRAS, 500, 998

Aditya J. N. H. S., et al., 2024, MNRAS, 527, 8511

Aditya J. N. H. S., et al., 2025, arXiv e-prints, p. arXiv:2507.19957

Agazie G., et al., 2024, ApJ, 966, 105

Ahumada R., et al., 2020, ApJS, 249, 3

Ajello M., 2007, PhD thesis, Munich University of Technology, Germany

Akiyama K., et al., 2019, ApJL, 875, L1

Akritas M. G., Bershady M. A., 1996, ApJ, 470, 706

Allison J. R., Sadler E. M., Whiting M. T., 2012, PASA, 29, 221

Allison J. R., et al., 2017, MNRAS, 465, 4450

Allison J. R., et al., 2022, PASA, 39, 010

Ambikasaran S., Foreman-Mackey D., Greengard L., Hogg D. W., O'Neil M., 2016, IEEE Transactions on Pattern Analysis and Machine Intelligence, 38, 252

An T., Baan W. A., 2012, ApJ, 760, 77

Antonucci R., 1993, ARA&A, 31, 473

Aoyama S., et al., 2019, MNRAS, 484, 1852

Ashton G., et al., 2019, ApJS, 241, 27

Astropy Collaboration et al., 2013, A&A, 558, A33

Astropy Collaboration et al., 2018, ApJ, 156, 123

Astropy Collaboration et al., 2022, ApJ, 935, 167

Astropy-Specutils Development Team 2019, Specutils: Spectroscopic analysis and reduction, Astrophysics Source Code Library, record ascl:1902.012

Baldi R. D., Capetti A., Giovannini G., 2015, A&A, 576, A38

Baldi R. D., Giovannini G., Capetti A., Lico R., 2025, arXiv e-prints, p. arXiv:2501.03787

Ballieux F. J., Callingham J. R., Röttgering H. J. A., Slob M. M., 2024, A&A, 689, A264

Barvainis R., Lehar J., Birkinshaw M., Falcke H., Blundell K. M., 2005, ApJ, 618, 108

Baum S., O'Dea C., Murphy D., de Bruyn A., 1990, A&A, 232, 19

Beaklini P. P. B., Dominici T. P., Abraham Z., 2017, A&A, 606, A87

Beck R., Dodds S. C., Szapudi I., 2022, MNRAS, 515, 4711

Behiri M., et al., 2025, A&A, 702, A10

Best P. N., Heckman T. M., 2012, MNRAS, 421, 1569

Best P. N., Kauffmann G., Heckman T. M., Ivezić v., 2005, MNRAS, 362, 9

Best P. N., Ker L. M., Simpson C., Rigby E. E., Sabater J., 2014, MNRAS, 445, 955

Bhattacharjee A., Seo J., Ryu D., Kang H., 2024, ApJ, 976, 91

Bicknell G. V., 1995, ApJS, 101, 29

Bicknell G. V., Dopita M. A., O'Dea C. P. O., 1997, ApJ, 485, 112

Bicknell G. V., Mukherjee D., Wagner A. Y., Sutherland R. S., Nesvadba N. P. H., 2018, MNRAS, 475, 3493

Bignall H. E., et al., 2003, ApJ, 585, 653

Bîrzan L., Rafferty D. A., McNamara B. R., Wise M. W., Nulsen P. E. J., 2004, ApJ, 607, 800

Blake G. M., 1970, Astrophys. Lett., 6, 201

Blake C., Mauch T., Sadler E. M., 2003, MNRAS, 347, 787

Blandford R. D., Königl A., 1979, ApJ, 232, 34

Blandford R. D., Rees M. J., 1978, in Wolfe A. M., ed., BL Lac Objects. pp 328–341

Blandford R., Meier D., Readhead A., 2019, ARA&A, 57, 467

Blundell K. M., Rawlings S., 1999, Nature, 399, 330

Bogdán Á., et al., 2024, Nature Astronomy, 8, 126

Bolton J. G., Stanley G. J., Slee O. B., 1949, Nat, 164, 101

Bolton J. G., Gardner F. F., Mackey M. B., 1963, Nature, 199, 682

Bonato M., et al., 2019, MNRAS, 485, 1188

Bondi M., Ciliegi P., Schinnerer E., Smolčić V., Jahnke K., Carilli C., Zamorani G., 2008, ApJ, 681, 1129

Bonning E. W., Cheng L., Shields G. A., Salviander S., Gebhardt K., 2007, ApJ, 659, 211

Bonolis L., Furlan S., 2025, The European Physical Journal H, 50, 14

Booth C. M., Theuns T., 2007, MNRAS, 381, L89

Boquien M., Burgarella D., Roehlly Y., Buat V., Ciesla L., Corre D., Inoue A. K., Salas H., 2019, A&A, 622, A103

Bordoloi R., et al., 2011, ApJ, 743, 10

Borodina O., et al., 2025, ApJ, 981, 149

Bose D., Chitnis V. R., Majumdar P., Shukla A., 2022, The European Physical Journal Special Topics, 231, 27

Braun R., Bourke T., Green J. A., Keane E., Wagg J., 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 174, doi:10.22323/1.215.0174

Broderick J. W., Bryant J. J., Hunstead R. W., Sadler E. M., Murphy T., 2007, MNRAS, 381, 341

Burgarella D., Buat V., Iglesias-Páramo J., 2005, MNRAS, 360, 1413

Butsky I. S., Hummels C. B., Hopkins P. F., Quinn T. R., Werk J. K., 2024, MNRAS, 535, 1672

Callingham J. R., et al., 2015, ApJ, 809, 168

Callingham J. R., et al., 2017, ApJ, 836, 174

Callingham J. R., et al., 2025, Nature, 647, 603

Capetti A., Balmaverde B., Coloma Puga M., Vizzone B., Jimenez-Gallardo A., García-Pérez A., Venturi G., 2025, A&A, 697, A238

Carilli C., Rawlings S., 2004, New Astron. Rev., 48, 979

Carilli C. L., Menten K. M., Reid M. J., Rupen M. P., Yun M. S., 1998, ApJ, 494, 175

Carr B., Kohri K., Sendouda Y., Yokoyama J., 2021, Reports on Progress in Physics, 84, 116902

Chambers K. C., Miley G. K., van Breugel W. J. M., 1990, ApJ, 363, 21

Charlot P., et al., 2020, A&A, 644, A159

Chen L., 2014, ApJ, 788, 179

Chen Y., Gu Q., Fan J., Yu X., Ding N., Xiong D., Guo X., 2023, ApJ, 944, 157

Chhetri R., Ekers R. D., Jones P. A., Ricci R., 2013, MNRAS, 434, 956

Chhetri R., Morgan J., Ekers R. D., Macquart J.-P., Sadler E. M., Giroletti M., Callingham J. R., Tingay S. J., 2018, MNRAS, 474, 4937

Chhetri R., et al., 2023, Advances in Space Research, 72, 5361

Chiaberge M., Gilli R., Lotz J. M., Norman C., 2015, ApJ, 806, 147

Ching J. H., et al., 2017, MNRAS, 464, 1306

Chisholm J., Tremonti C. A., Leitherer C., Chen Y., 2017, MNRAS, 469, 4831

Ciesla L., et al., 2015, A&A, 576, A10

Clarke M., 1964, PhD thesis, Cambridge University

Cohen A. S., Lane W. M., Cotton W. D., Kassim N. E., Lazio T. J. W., Perley R. A., Condon J. J., Erickson W. C., 2007, AJ, 134, 1245

Condon J. J., 2002, in Pramesh Rao A., Swarup G., Gopal-Krishna eds, IAU Symposium Vol. 199, The Universe at Low Radio Frequencies. p. 3

Condon J. J., Broderick J. J., Seielstad G. A., Douglas K., Gregory P. C., 1994, AJ, 107, 1829

Condon J. J., Cotton W. D., Greisen E. W., Yin Q. F., Perley R. A., Taylor G. B., Broderick J. J., 1998, AJ, 115, 1693

Conroy C., 2013, ARA&A, 51, 393

Conway J., 1997, in Snellen I. A. G., Schilizzi R. T., Roettgering H. J. A., Bremer M. N., eds, Gigahertz Peaked Spectrum and Compact Steep Spectrum Radio Sources. pp 198–207

Crain R. A., van de Voort F., 2023, ARA&A, 61, 473

Crain R. A., et al., 2017, MNRAS, 464, 4204

Croton D. J., et al., 2006, MNRAS, 365, 11

Cui L., Liu X., Liu J., Song H.-G., Ding Z., 2010, A&A, 518, A23

Cui W., Knebe A., Yepes G., Yang X., Borgani S., Kang X., Power C., Staveley-Smith L., 2018, MNRAS, 473, 68

Curran S. J., 2021, MNRAS, 506, 1548

Curran S. J., Allison J. R., Glowacki M., Whiting M. T., Sadler E. M., 2013, MNRAS, 431, 3408

Cutri R. M., et al., 2021, VizieR Online Data Catalog: AllWISE Data Release (Cutri+ 2013), VizieR On-line Data Catalog: II/328. Originally published in: IPAC/Caltech (2013)

Czerny B., Siemiginowska A., Janiuk A., Nikiel-Wroczyński B., Stawarz L., 2009, ApJ, 698, 840

DESI Collaboration et al., 2025, arXiv e-prints, p. arXiv:2503.14745

Dallacasa D., Stanghellini C., Centonza M., Fanti R., 2000, A&A, 363, 887

Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827

Davé R., Crain R. A., Stevens A. R. H., Narayanan D., Saintonge A., Catinella B., Cortese L., 2020, MNRAS, 497, 146

De Breuck C., et al., 2001, AJ, 121, 1241

De Breuck C., Tang Y., de Bruyn A. G., Rottgering H., van Breugel W., 2002, A&A, 394, 59

Deka P. P., et al., 2024, ApJS, 270, 33

Dey A., et al., 2019, AJ, 157, 168

Di Matteo T., Springel V., Hernquist L., 2005, Nature, 433, 604

DiPompeo M. A., Runnoe J. C., Myers A. D., Boroson T. A., 2013, ApJ, 774, 24

Dicke R. H., Peebles P. J. E., Roll P. G., Wilkinson D. T., 1965, ApJ, 142, 414

Dickey J. M., et al., 2013, Publ. Astron. Soc. Australia, 30, e003

Djorgovski S. G., Williams R., 2005, in Kassim N., Perez M., Junor W., Henning P., eds, Astronomical Society of the Pacific Conference Series Vol. 345, From Clark Lake to the Long Wavelength Array: Bill Erickson's Radio Science. p. 517 (arXiv:astro-ph/0504006), doi:10.48550/arXiv.astro-ph/0504006

Douglas J. N., et al., 1996, AJ, 111, 1945

Draine B. T., 2011, Physics of the Interstellar and Intergalactic Medium. Princeton University Press, https://press.princeton.edu/books/paperback/9780691122144/physics-of-the-interstellar-and-intergalactic-medium

Drake F. D., 1982, Physics Today, 35, 9

Driessen L. N., et al., 2024, Publ. Astron. Soc. Australia, 41, e084

Driver S. P., et al., 2009, A&G, 50, 12

Drouart G., et al., 2020, Publ. Astron. Soc. Australia, 37, e026

Dubois Y., et al., 2014, MNRAS, 444, 1453

Duchesne S. W., et al., 2023, Publ. Astron. Soc. Australia, 40, e034

Duffy P., Blundell K. M., 2011, MNRAS, 421, 108

Dugan Z., Gaibler V., Silk J., 2017, ApJ, 844, 37

Duncan K. J., 2022, MNRAS, 512, 3662

Dutta R., 2019, J. Astrophys. Astron., 40, 41

Dutta R., Srianand R., Gupta N., Joshi R., Petitjean P., Noterdaeme P., Ge J., Krogager J.-K., 2017, MNRAS, 465, 4249

Dutta R., Srianand R., Gupta N., 2018, MNRAS, 480, 947

Dutta R., Raghunathan S., Gupta N., Joshi R., 2020, MNRAS, 491, 838

D'Abrusco R., et al., 2019, ApJS, 242, 4

Edwards P. G., Tingay S. J., 2004, A&A, 424, 91

Ekers J. A., 1969, Australian Journal of Physics Astrophysical Supplement, 7, 3

Ekers R. D., 2012, in Resolving The Sky - Radio Interferometry: Past, Present and Future. p. 7, doi:10.22323/1.163.0007

El-Badry K., et al., 2018, MNRAS, 477, 1536

Ellison S. L., Viswanathan A., Patton D. R., Bottrell C., McConnachie A. W., Gwyn S., Cuillandre J.-C., 2019, MNRAS, 487, 2491

Elson E. C., Blyth S. L., Baker A. J., 2016, MNRAS, 460, 4366

Emig K. L., Salas P., de Gasperin F., Oonk J. B. R., Toribio M. C., Röttgering H. J. A., Tielens A. G. G. M., 2019, A&A, 622, A7

Emonts B. H. C., Morganti R., Tadhunter C. N., Oosterloo T. A., Holt J., van der Hulst J. M., 2005, MNRAS, 362, 931

Escott E. L., et al., 2025, MNRAS, 536, 1166

Ezeugo J. C., Ubachukwu A. A., 2010, MNRAS, 408, 2256

Fabian A. C., 2012, ARA&A, 50, 455

Fan L., et al., 2016, ApJ, 822, L32

Fanaroff B. L., Riley J. M., 1974, MNRAS, 167, 31P

Fanti C., Fanti R., Parma P., Schilizzi R. T., van Breugel W. J. M., 1985, A&A, 143, 292

Fanti R., et al., 1990, A&A, 231, 333

Fawcett V. A., et al., 2023, MNRAS, 525, 5575

Fitzpatrick M. J., et al., 2014, in Peck A. B., Benn C. R., Seaman R. L., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 9149, Observatory Operations: Strategies, Processes, and Systems V. p. 91491T, doi:10.1117/12.2057445

Forbes G., 1909, History of astronomy. G. P. Putnam

Förster Schreiber N. M., et al., 2019, ApJ, 875, 21

Furlanetto S. R., Oh S. P., Briggs F. H., 2006, Phys. Rep., 433, 181

Gaibler V., Khochfar S., Krause M., Silk J., 2012, MNRAS, 425, 438

Gaspari M., Ruszkowski M., Oh S. P., 2013, MNRAS, 432, 3401

Gebek A., Trčka A., Baes M., Martorano M., Pillepich A., Kapoor A. U., Nersesian A., van der Wel A., 2024, MNRAS, 531, 3839

Geil P. M., Wyithe J. S. B., 2008, MNRAS, 386, 1683

Geréb K., Morganti R., Oosterloo T. A., 2014, A&A, 569, 35

Geréb K., Maccagni F. M., Morganti R., Oosterloo T. A., 2015, A&A, 575, A44

Ginsburg A., et al., 2019, AJ, 157, 98

Girdhar A., et al., 2024, MNRAS, 527, 9322

Gloudemans A. J., et al., 2022, A&A, 668, A27

Glowacki M., et al., 2019, MNRAS, 489, 4926

Glowacki M., Deg N., Blyth S.-L., Hank N., Davé R., Elson E., Spekkens K., 2022, MNRAS, 517, 1282

Gommers R., et al., 2025, scipy/scipy: SciPy 1.16.3, doi:10.5281/zenodo.17467817, https://doi.org/10.5281/zenodo.17467817

Gong Y., Zhou L., Yuan M., Zhang H., Yi T., Fang J., 2022, ApJ, 931, 168

Goodfellow I., Bengio Y., Courville A., 2016, Deep Learning. MIT Press

Gordon Y. A., et al., 2021, ApJS, 255, 30

Goss W. M., Hooker C., Ekers R. D., Pawsey J. L., 2023, Joe Pawsey and the founding of Australian radio astronomy: early discoveries, from the sun to the cosmos. Springer Cham, doi:10.1007/978-3-031-07916-0

Grasha K., Darling J., Leroy A. K., Bolatto A. D., 2020, MNRAS, 498, 883

Griffith M. R., Wright A. E., 1993, AJ, 105, 1666

Griffith M. R., Wright A. E., Griffith M. R., Wright A. E., 1993, AJ, 105, 1666

Gupta N., et al., 2006, MNRAS, 373, 972

Gupta N., et al., 2016, in Proceedings of MeerKAT Science: On the Pathway to the SKA — PoS(MeerKAT2016). Sissa Medialab, Trieste, Italy, doi:10.22323/1.277.0014

Gupta N., et al., 2018, MNRAS, 476, 2432

Hale C. L., et al., 2021, PASA, 38, e058

Hallinan G., et al., 2019, in Bulletin of the American Astronomical Society. p. 255 (arXiv:1907.07648), doi:10.48550/arXiv.1907.07648

Hancock P. J., Tingay S. J., Sadler E. M., Phillips C., Deller A. T., 2009, MNRAS, 397, 2030

Hancock P. J., Sadler E. M., Mahony E. K., Ricci R., 2010, MNRAS, 408, 1187

Hancock P. J., Charlton E. G., Macquart J.-P., Hurley-Walker N., 2019, arXiv

Hardcastle M. J., 2018, MNRAS, 475, 2768

Hardcastle M., Croston J., 2020, New Astron. Rev., 88, 101539

Hardcastle M. J., et al., 2019, MNRAS, 488, 3416

Hardcastle M. J., et al., 2025, MNRAS, 539, 1856

Harris C. R., et al., 2020, Nature, 585, 357

Harrison C. M., Ramos Almeida C., 2024, Galaxies, 12, 17

Harrison C. M., Alexander D. M., Mullaney J. R., Swinbank A. M., 2014, MNRAS, 441, 3306

Haworth T. J., Glover S. C. O., Koepferl C. M., Bisbas T. G., Dale J. E., 2018, New Astron. Rev., 82, 1

Haynes M. P., 2008, in Minchin R., Momjian E., eds, American Institute of Physics Conference Series Vol. 1035, The Evolution of Galaxies Through the Neutral Hydrogen Window. AIP, pp 337–344 (arXiv:0806.1673), doi:10.1063/1.2973610

Hazard C., Mackey M. B., Shimmins A. J., 1963, Nature, 197, 1037

Healey S. E., Romani R. W., Taylor G. B., Sadler E. M., Ricci R., Murphy T., Ulvestad J. S., Winn J. N., 2007, ApJS, 171, 61

Healey S. E., et al., 2008, ApJS, 175, 97

Heckman T. M., Best P. N., 2014, ARA&A, 52, 589

Helfand D. J., White R. L., Becker R. H., 2015, ApJ, 801, 26

Hewish A., Scott P. F., Wills D., 1964, Nature, 203, 1214

Hewish A., Bell S. J., Pilkington J. D. H., Scott P. F., Collins R. A., 1968, Nature, 217, 709

Hey J. S., Parsons S. J., Phillips J. W., 1946, Nature, 158, 234

Heywood I., et al., 2020, MNRAS, 494, 5018

Hickox R. C., Alexander D. M., 2018, ARA&A, 56, 625

Higson E., Handley W., Hobson M., Lasenby A., 2018, Bayesian Analysis, 13

Higson E., Handley W., Hobson M., Lasenby A., 2019a, Statistics and Computing, 29, 891

Higson E., Handley W., Hobson M., Lasenby A., 2019b, MNRAS, 483, 2044

Hinton S., Davis T. M., Lidman C., Glazebrook K., Lewis G., 2016, Astronomy and Computing, 15, 61

Hlavacek-Larrondo J., Li Y., Churazov E., 2022, in Bambi C., Sangangelo A., eds, , Handbook of X-ray and Gamma-ray Astrophysics. Springer Singapore, p. 5, doi:10.1007/978-981-16-4544-0_122-1

Hogan M. T., et al., 2015, MNRAS, 453, 1223

Hogg D. W., 2026, arXiv e-prints, p. arXiv:2602.10181

Holt J., Tadhunter C., Morganti R., Bellamy M., González Delgado R. M., Tzioumis A., Inskip K. J., 2006, MNRAS, 370, 1633

Holt J., Tadhunter C. N., Morganti R., 2008, MNRAS, 387, 639

Hopkins P. F., 2015, MNRAS, 450, 53

Hopkins P. F., Hernquist L., 2006, ApJS, 166, 1

Hopkins P. F., Hernquist L., Cox T. J., Kereš D., 2008, ApJS, 175, 356

Hopkins P. F., et al., 2018, MNRAS, 480, 800

Horton M. A., Hardcastle M. J., Miley G. K., Tasse C., Shimwell T., 2025, A&A, 699, A338

Hotan A. W., et al., 2021, Publ. Astron. Soc. Australia, 38, e009

Hovatta T., Tornikoski M., Lainela M., Lehto H. J., Valtaoja E., Torniainen I., Aller M. F., Aller H. D., 2007, A&A, 469, 899

Hovatta T., Nieppola E., Tornikoski M., Valtaoja E., Aller M. F., Aller H. D., 2008, A&A, 485, 51

Hunter J. D., 2007, Computing in Science & Engineering, 9, 90

Hurley-Walker N., et al., 2017, MNRAS, 464, 1146

Hurley-Walker N., et al., 2022a, Publ. Astron. Soc. Australia, 39, e035

Hurley-Walker N., et al., 2022b, Nature, 601, 526

Hurley-Walker N., et al., 2023, Nature, 619, 487

IceCube Collaboration et al., 2018, Science, 361, 147

Ighina L., et al., 2024, A&A, 692, A241

Intema H. T., Jagannathan P., Mooley K. P., Frail D. A., 2017, A&A, 598, A78

Ishibashi W., Auger M. W., Zhang D., Fabian A. C., 2014, MNRAS, 443, 1339

Ishibashi W., Fabian A. C., Maiolino R., 2018, MNRAS, 476, 512

Jaiswal S., An T., Wang A., Tingay S., 2022, MNRAS, 509, 2122

Jansky K., 1935, Proceedings of the Institute of Radio Engineers, 23, 1158

Jarvis M., et al., 2018, in Proceedings of MeerKAT Science: On the Pathway to the SKA — PoS(MeerKAT2016). Sissa Medialab, Trieste, Italy, p. 006, doi:10.22323/1.277.0006

Jarvis M. E., et al., 2019, MNRAS, 485, 2710

Jauncey D. L., et al., 2003, Publ. Astron. Soc. Australia, 20, 151

Jeffreys H., 1946, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 186, 453

Jeyakumar S., 2016, MNRAS, 458, 3786

Jeyakumar S., Saikia D. J., Pramesh Rao A., Balasubramanian V., 2000, A&A, 362, 27

Jhee H., et al., 2026, MNRAS,

Jog C. J., Combes F., 2009, Phys. Rep., 471, 75

Juneau S., Olsen K., Nikutta R., Jacques A., Bailey S., 2021, Computing in Science & Engineering, 23, 15

Jung S. L., Whittam I. H., Jarvis M. J., Hale C. L., Tudorache M. N., Yasin T., 2025, MNRAS, 539, 2362

Kalberla P. M. W., Kerp J., 2009, ARA&A, 47, 27

Kanekar N., Briggs F. H., 2003, A&A, 412, L29

Kanekar N., Briggs F. H., 2004, New Astron. Rev., 48, 1259

Kanekar N., Chengalur J. N., 2001, MNRAS, 325, 631

Kanekar N., et al., 2014, MNRAS, 438, 2131

Kass R. E., Raftery A. E., 1995, Journal of the American Statistical Association, 90, 773

Kedziora-Chudczer L., 2006, MNRAS, 369, 449

Kedziora-Chudczer L., Jauncey D. L., Wieringa M. H., Walker M. A., Nicolson G. D., Reynolds J. E., Tzioumis A. K., 1997, ApJ, 490, L9

Kedziora-Chudczer L., et al., 2001, in Laing R. A., Blundell K. M., eds, Astronomical Society of the Pacific Conference Series Vol. 250, Particles and Fields in Radio Galaxies Conference. p. 128

Keim M. A., Callingham J. R., Röttgering H. J., 2019, A&A, 628, 56

Keller P. M., Thyagarajan N., Kumar A., Kanekar N., Bernardi G., 2024, MNRAS, 528, 5692

Kellermann K. I., 1965, AJ, 70, 681

Kellermann K. I., 1972, AJ, 77, 531

Kellermann K. I., Bouton E. N., 2023, Star Noise: Discovering the Radio Universe. Cambridge University Press

Kellermann K. I., Long R. J., Allen L. R., Moran M., 1962, Nature, 195, 692

Kellermann K. I., Condon J. J., Kimball A. E., Perley R. A., Ivezic Z., 2016, ApJ, 831, 168

Kerrison E. F., Sadler E. M., Moss V. A., Mahony E. K., 2024a, arXiv e-prints, p. arXiv:2412.08002

Kerrison E. F., Ekers R. D., Morgan J., Chhetri R., 2024b, arXiv e-prints, p. arXiv:2412.19816

Kerrison E. F., Allison J. R., Moss V. A., Sadler E. M., Rees G. A., 2024c, MNRAS, 533, 4248

Kerrison E. F., et al., 2025, MNRAS, 543, 3895

Kerrison E. F., et al., 2026, Publ. Astron. Soc. Australia, 43, e058

Kiehlmann S., et al., 2024a, ApJ, 961, 240

Kiehlmann S., et al., 2024b, ApJ, 961, 241

Kim C.-G., et al., 2020, ApJ, 900, 61

Kimball A. E., Ivezić Ž., 2008, AJ, 136, 684

King A., Pounds K., 2015, ARA&A, 53, 115

King O. G., et al., 2010, SPIE, 7741, 77411I

Klein J. P., Moeschberger M. L., 2003, in , Survival Analysis. Statistics for Biology and Health. Springer, New York, pp 63–90, doi:10.1007/0-387-21645-6_3

Koay J. Y., Bignall H. E., Macquart J.-P., Jauncey D. L., Rickett B. J., Lovell J. E. J., 2011, A&A, 534, L1

Koay J. Y., et al., 2018, MNRAS, 474, 4396

Kocevski D. D., et al., 2015, ApJ, 814, 104

Kochanek C. S., et al., 2017, PASP, 129, 104502

Kondapally R., et al., 2022, MNRAS, 513, 3742

Kondapally R., et al., 2023, MNRAS, 523, 5292

Koposov S., et al., 2023, joshspeagle/dynesty: v2.1.2 (v2.1.2)

Kormendy J., Ho L. C., 2013, ARA&A, 51, 511

Kosmaczewski E., et al., 2020, ApJ, 897, 164

Kovalev Y. A., Kovalev Y. Y., 2006, in Gaskell C. M., McHardy I. M., Peterson B. M., Sergeev S. G., eds, Astronomical Society of the Pacific Conference Series Vol. 360, AGN Variability from X-Rays to Radio Waves. p. 137

Kravchenko E. V., Pashchenko I. N., Homan D. C., Kovalev Y. Y., Lister M. L., Pushkarev A. B., Ros E., Savolainen T., 2025, MNRAS, 538, 2008

Kravtsov A. V., Klypin A. A., Khokhlov A. M., 1997, ApJS, 111, 73

Krumholz M. R., Gnedin N. Y., 2011, ApJ, 729, 36

Kuehr H., Witzel A., Pauliny-Toth I. I. K., Nauber U., Kuehr H., Witzel A., Pauliny-Toth I. I. K., Nauber U., 1981, A&AS, 45, 367

Kusumegi K., Yang X., Ginsparg P., de Vaan M., Stuart T., Yin Y., 2025, Science, 390, 1240

Kutkin A. M., Morganti R., Oosterloo T. A., Adams E. A. K., Dénes H., van Leeuwen J., Norden M. J., Orru E., 2025, A&A, 700, A63

LaMassa S. M., et al., 2013, MNRAS, 436, 3581

Labbé I., et al., 2023, Nature, 616, 266

Labiano A., Barthel P. D., O'Dea C. P., de Vries W. H., Pérez I., Baum S. A., 2007, A&A, 463, 97

Labrie K., et al., 2023, Research Notes of the AAS, 7, 214

Lancaster L., Kim C.-G., Kim J.-G., Ostriker E. C., Bryan G. L., 2025, ApJ, 989, 43

Land-Strykowski M., Lewis G. F., Murphy T., 2025, MNRAS, 543, 3229

Lane W. M., Cotton W. D., van Velzen S., Clarke T. E., Kassim N. E., Helmboldt J. F., Lazio T. J., Cohen A. S., 2014, MNRAS, 440, 327

Langan I., Popping G., Ginolfi M., Gentile F., Valentino F., Kaasinen M., 2024, A&A, 689, A283

Large M. I., et al., 1981, MNRAS, 194, 693

Le Reste A., et al., 2025, A&A, 693, A253

Lister M. L., 2003, in Zensus J. A., Cohen M. H., Ros E., eds, Astronomical Society of the Pacific Conference Series Vol. 300, Radio Astronomy at the Fringe. p. 71 (arXiv:astro-ph/0301332), doi:10.48550/arXiv.astro-ph/0301332

Łokas E. L., 2022, A&A, 662, A53

Lyke B. W., et al., 2020, ApJS, 250, 8

Maccagni F. M., Morganti R., Oosterloo T. A., Mahony E. K., 2014, A&A, 571, 67

Maccagni F. M., Santoro F., Morganti R., Oosterloo T. A., Oonk J. B. R., Emonts B. H. C., 2016, A&A, 588, A46

Maccagni F. M., Morganti R., Oosterloo T. A., Geréb K., Maddox N., 2017, A&A, 604, 43

Macquart J.-P., 2005, A&A, 433, 827

Macquart J.-P., Tingay S., 2016, MNRAS, 460, 2322

Madau P., Dickinson M., 2014, ARA&A, 52, 415

Mahony E. K., Sadler E. M., Murphy T., Ekers R. D., Edwards P. G., Massardi M., 2010, ApJ, 718, 587

Mahony E. K., Morganti R., Emonts B. H. C., Oosterloo T. A., Tadhunter C., 2013, MNRAS, 435, L58

Manchester R. N., Hobbs G. B., Teoh A., Hobbs M., 2005, AJ, 129, 1993

Marocco F., et al., 2021, ApJS, 253, 8

Marscher A. P., Gear W. K., 1985, ApJ, 298, 114

Martínez-Aldama M. L., del Olmo A., Marziani P., Sulentic J. W., Negrete C. A., Dultzin D., D'Onofrio M., Perea J., 2018, A&A, 618, A179

Marziani P., Sulentic J. W., 2014, MNRAS, 442, 1211

Massardi M., et al., 2011, MNRAS, 412, 318

Massaro E., Giommi P., Leto C., Marchegiani P., Maselli A., Perri M., Piranomonte S., Sclavi S., 2009, A&A, 495, 691

Massaro E., Maselli A., Leto C., Marchegiani P., Perri M., Giommi P., Piranomonte S., 2015, APSS, 357, 75

Matthee J., et al., 2024, ApJ, 963, 129

Mauch T., Sadler E. M., 2007, MNRAS, 375, 931

Mauch T., Murphy T., Buttery H. J., Curran J., Hunstead R. W., Piestrzynski B., Robertson J. G., Sadler E. M., 2003, MNRAS, 342, 1117

Mauch T., et al., 2013, VizieR Online Data Catalog: Sydney University Molonglo Sky Survey (SUMSS V2.1) (Mauch+ 2008), https://ui.adsabs.harvard.edu/abs/2013yCat.8081....0M/abstract

McClure-Griffiths N. M., Stanimirović S., Rybarczyk D. R., 2023, ARA&A, 61, 19

McConnell D., Sadler E. M., Murphy T., Ekers R. D., 2012, MNRAS, 422, 1527

McConnell D., et al., 2020, PASA, 37, 18

McKee C. F., Parravano A., Hollenbach D. J., 2015, ApJ, 814, 13

Meenakshi M., Mukherjee D., Wagner A. Y., Nesvadba N. P. H., Morganti R., Janssen R. M. J., Bicknell G. V., 2022, MNRAS, 511, 1622

Merloni A., Heinz S., di Matteo T., 2003, MNRAS, 345, 1057

Merson A. I., et al., 2013, MNRAS, 429, 556

Meyers B. W., et al., 2017, PASA, 34, e013

Miley G., 1980, ARA&A, 18, 165

Miley G., De Breuck C., 2008, aapr, 15, 67

Millon M., Courbin F., Galan A., Sluse D., Ding X., Tewes M., Djorgovski S. G., 2023, Nature Astronomy, 7, 959

Mills B. Y., Slee O. B., Hill E. R., 1958, Australian Journal of Physics, 11, 360

Mingaliev M. G., Sotnikova Y. V., Torniainen I., Tornikoski M., Udovitskiy R. Y., 2012, A&A, 544, A25

Mingo B., et al., 2019, MNRAS, 488, 2701

Miranda Marques B. L., Rodríguez-Ardila A., Fonseca-Faria M. A., Panda S., 2025, ApJ, 978, 16

Mocz P., et al., 2019, Phys. Rev. Lett., 123, 141301

Morabito L. K., Harwood J. J., 2018, MNRAS, 480, 2726

Morabito L. K., et al., 2022a, MNRAS, 515, 5758

Morabito L. K., et al., 2022b, A&A, 658, A1

Morgan J. S., et al., 2018, MNRAS, 473, 2965

Morgan J. S., Chhetri R., Ekers R., 2022, Publ. Astron. Soc. Australia, 39, e063

Morganti R., Oosterloo T., 2018, A&ARv, 26, 4

Morganti R., Oosterloo T. A., Tadhunter C. N., van Moorsel G., Killeen N., Wills K. A., 2001, MNRAS, 323, 331

Morganti R., Tadhunter C. N., Oosterloo T. A., 2005, A&A, 444, L9

Morganti R., Sadler E. M., Curran S., 2015, Advancing Astrophysics with the Square Kilometre Array (AASKA14), p. 134

Morganti R., Murthy S., Guillard P., Oosterloo T., Garcia-Burillo S., 2023, Galaxies, 11, 24

Morganti R., Oosterloo T., Tadhunter C., Murthy S., 2026, arXiv e-prints, p. arXiv:2602.09478

Morselli L., Renzini A., Popesso P., Erfanianfar G., 2016, MNRAS, 462, 2355

Mountrichas G., et al., 2022, A&A, 667, A145

Mukherjee D., 2025, Galaxies, 13, 102

Mukherjee D., Wagner A. Y., Bicknell G. V., Morganti R., Oosterloo T., Nesvadba N., Sutherland R. S., 2018a, MNRAS, 476, 80

Mukherjee D., Bicknell G. V., Wagner A. Y., Sutherland R. S., Silk J., 2018b, MNRAS, 479, 5544

Murphy T., et al., 2010, MNRAS, 402, 2403

Murphy T., et al., 2013, Publ. Astron. Soc. Australia, 30, e006

Murphy E. J., et al., 2018, in Murphy E., ed., Astronomical Society of the Pacific Conference Series Vol. 517, Science with a Next Generation Very Large Array. p. 3, doi:10.48550/arXiv.1810.07524

Murphy T., et al., 2021, Publ. Astron. Soc. Australia, 38, e054

Murthy S., Morganti R., Wagner A. Y., Oosterloo T., Guillard P., Mukherjee D., Bicknell G., 2022, Nature Astronomy, 6, 488

Myers A. D., et al., 2015, ApJS, 221, 27

Netzer H., 1990, in Blandford R. D., Netzer H., Woltjer L., Courvoisier T. J. L., Mayor M., eds, Active Galactic Nuclei. pp 57–160

Netzer H., 2013, The Physics and Evolution of Active Galactic Nuclei. Cambridge University Press

Netzer H., 2015, ARA&A, 53, 365

Nguyen H., et al., 2024, MNRAS, 534, 3478

Nikutta R., Fitzpatrick M., Scott A., Weaver B. A., 2020, Astronomy and Computing, 33, 100411

Nims J., Quataert E., Faucher-Giguère C.-A., 2015, MNRAS, 447, 3612

Norris R. P., et al., 2011, Publ. Astron. Soc. Australia, 28, 215

Norris R. P., et al., 2025, Publ. Astron. Soc. Australia, 42, e124

Nyland K., et al., 2020, ApJ, 905, 74

O'Dea C. P., 1998, PASA, 110, 493

O'Dea C. P., Baum S. A., 1997, AJ, 113, 148

O'Dea C. P., Baum S. A., Stanghellini C., 1991, ApJ, 380, 66

O'Dea C. P., Stanghellini C., Baum S. A., Charlot S., 1996, ApJ, 470, 806

Oayda O. T., Mittal V., Lewis G. F., Murphy T., 2024, MNRAS, 531, 4545

Obreschkow D., Meyer M., 2014, arXiv e-prints, p. arXiv:1406.0966

Ochsenbein F., Bauer P., Marcout J., Ochsenbein F., Bauer P., Marcout J., 2000, A&AS, 143, 23

Oman K. A., 2019, MARTINI: Mock spatially resolved spectral line observations of simulated galaxies, Astrophysics Source Code Library, record ascl:1911.005 (ascl:1911.005)

Oman K. A., 2024, Journal of Open Source Software, 9, 6860

Oman K. A., Marasco A., Navarro J. F., Frenk C. S., Schaye J., Benítez-Llambay A., 2019, MNRAS, 482, 821

Oosterloo T., Morganti R., Tadhunter C., Raymond Oonk J. B., Bignall H. E., Tzioumis T., Reynolds C., 2019, A&A, 632, A66

Orienti M., Dallacasa D., 2014, MNRAS, 438, 463

Orienti M., Dallacasa D., Stanghellini C., Orienti M., Dallacasa D., Stanghellini C., 2007, A&A, 475, 813

Orienti M., Dallacasa D., Stanghellini C., 2010, MNRAS, 408, 1075

Owen F. N., Rudnick L., 1976, ApJ, 205, L1

O'Dea C. P., Saikia D. J., 2021, A&ARv, 29, 1

Pacifici C., et al., 2023, ApJ, 944, 141

Padovani P., 2011, MNRAS, 411, 1547

Padovani P., 2016, aapr, 24, 13

Padovani P., et al., 2017, A&ARv, 25, 2

Parijskij Y. N., 1992, Astronomical and Astrophysical Transactions, 1, 85

Patil P., et al., 2022, ApJ, 934, 26

Pence W. D., Chiappetti L., Page C. G., Shaw R. A., Stobie E., 2010, A&A, 524, A42

Penston M. V., Fosbury R. A. E., 1978, MNRAS, 183, 479

Penzias A. A., Wilson R. W., 1965, ApJ, 142, 419

Perley R., et al., 2009, IEEE Proceedings, 97, 1448

Péroux C., Howk J. C., 2020, ARA&A, 58, 363

Péroux C., et al., 2019, MNRAS, 485, 1595

Perucho M., Martí J.-M., Quilis V., 2022, MNRAS, 510, 2084

Peterson B. M., 2006, in Alloin D., ed., , Vol. 693, Physics of Active Galactic Nuclei at all Scales. Springer Berlin, Heidelberg, p. 77, doi:10.1007/3-540-34621-X_3

Petrov L. Y., Kovalev Y. Y., 2025, ApJS, 276, 38

Phillips R. B., Mutel R. L., 1980, ApJ, 236, 89

Pihlström Y. M., Conway J. E., Vermeulen R. C., 2003, A&A, 404, 871

Pillepich A., et al., 2018, MNRAS, 473, 4077

Pineau F.-X., Boch T., Derrière S., Schaaff A., 2020, in Ballester P., Ibsen J., Solar M., Shortridge K., eds, Astronomical Society of the Pacific Conference Series Vol. 522, Astronomical Data Analysis Software and Systems XXVII. p. 125

Planck Collaboration 2020, A&A, 641, A6

Planck Collaboration XIII 2016, A&A, 594, A13

Polisensky E., Clarke T. E., Giacintucci S., Peters W., 2024, Frontiers in Astronomy and Space Sciences, 11

Pontzen A., Rey M. P., Cadiou C., Agertz O., Teyssier R., Read J., Orkney M. D. A., 2021, MNRAS, 501, 1755

Pritchard J., 2025, askap-vast/dstools: v2.0.0, doi:10.5281/zenodo.15232974

Pskovskii Y. P., 1962, Soviet Ast., 6, 172

Pushkarev A. B., Kovalev Y. Y., 2012, A&A, 544, A34

Rahmati A., Pawlik A. H., Raičević M., Schaye J., 2013, MNRAS, 430, 2427

Randall K. E., Hopkins A. M., Norris R. P., Edwards P. G., 2011, MNRAS, 416, 1135

Randall K. E., Hopkins A. M., Norris R. P., Zinn P.-C., Middelberg E., Mao M. Y., Sharp R. G., 2012, MNRAS, 421, 1644

Rao S. M., Turnshek D. A., Sardane G. M., Monier E. M., 2017, MNRAS, 471, 3428

Readhead A. C. S., Taylor G. B., Xu W., Pearson T. J., Wilkinson P. N., Polatidis A. G., 1996, ApJ, 460, 612

Reardon D. J., et al., 2025, Nature Astronomy, 9, 1053

Reynolds J. E., 1994, ATNF Technical Document Series, 39.3

Ricci R., Prandoni I., Gruppioni G., Sault R. J., De Zotti G., 2006, A&A, 445, 465

Richards E. A., et al., 1999, ApJL, 526, L73

Rickett B. J. B., 1968, PhD thesis, University of Manchester, UK, https://ui.adsabs.harvard.edu/abs/1968PhDT.....87R

Rickett B., 2001, Astrophysics and Space Science, 278, 5

Rickett B., 2002, Publ. Astron. Soc. Australia, 19, 100

Rickett B., Coles W., Bourgois G., 1984, A&A, 134, 390

Rickett B., Kedziora-Chudczer L., Jauncey D. L., 2002, Publ. Astron. Soc. Australia, 19, 106–110

Risaliti G., Elvis M., Nicastro F., 2002, ApJ, 571, 234

Roper F. A., Oman K. A., Frenk C. S., Benítez-Llambay A., Navarro J. F., Santos-Santos I. M. E., 2023, MNRAS, 521, 1316

Ross N. P., et al., 2015, MNRAS, 453, 3933

Ross K., et al., 2021, MNRAS, 501, 6139

Ross K., et al., 2024, Publ. Astron. Soc. Australia, 41, e054

Roster W., et al., 2024, A&A, 692, A260

Roster W., et al., 2026, Publ. Astron. Soc. Australia, 43, e065

Ryle M., 1955, The Observatory, 75, 137

Ryle M., Hewish A., 1960, MNRAS, 120, 220

Ryon J. E., 2021, in , Vol. 20, ACS Instrument Handbook for Cycle 29 v. 20.0. STScI, p. 20

Sadler E. M., 2016, Astronomische Nachrichten, 337, 105

Sadler E. M., et al., 2006, MNRAS, 371, 898

Sadler E. M., et al., 2007, MNRAS, 381, 211

Sadler E. M., Ekers R. D., Mahony E. K., Mauch T., Murphy T., 2014, MNRAS, 438, 796

Sadler E. M., Chhetri R., Morgan J., Mahony E. K., Jarrett T. H., Tingay S., 2019, MNRAS, 483, 1354

Saikia D. J., Jamrozy M., 2009, Bulletin of the Astronomical Society of India, 37, 63

Sajina A., et al., 2011, ApJ, 732, 45

Sandage A., 1965, ApJ, 141, 1560

Santoro F., Tadhunter C., Baron D., Morganti R., Holt J., 2020, A&A, 644, A54

Sargent C. L., Alexander D. M., Greenwell C. L., Fawcett V. A., Morabito L. K., Harrison C. M., Meenakshi M., Hickox R. C., 2026, MNRAS, 548, stag597

Savage A., Wright A. E., 1981, MNRAS, 196, 927

Saxena A., et al., 2018, MNRAS, 475, 5041

Schaye J., et al., 2015, MNRAS, 446, 521

Schaye J., et al., 2026, MNRAS, 548, stag375

Schmidt M., 1963, Nature, 197, 1040

Schoenmakers A. P., de Bruyn A. G., Röttgering H. J. A., van der Laan H., Kaiser C. R., 2000, MNRAS, 315, 371

Seyfert C. K., 1943, ApJ, 97, 28

Shabala S. S., Ash S., Alexander P., Riley J. M., 2008, MNRAS, 388, 625

Shakeshaft J. R., Ryle M., Baldwin J. E., Elsmore B., Thomson J. H., 1955, Mem. RAS, 67, 106

Shang Z., et al., 2005, ApJ, 619, 41

Shao Y., Wagg J., Wang R., Carilli C. L., Riechers D. A., Intema H. T., Weiss A., Menten K. M., 2020, A&A, 641, A85

Sharma R. S., et al., 2024, MNRAS, 527, 9461

Shimizu T. T., Mushotzky R. F., Meléndez M., Koss M., Rosario D. J., 2015, MNRAS, 452, 1841

Shimwell T. W., et al., 2017, A&A, 598, A104

Shimwell T. W., et al., 2019, A&A, 622, A1

Shklovsky J., 1965, Nature, 206, 176

Sijacki D., Springel V., Di Matteo T., Hernquist L., 2007, MNRAS, 380, 877

Singh K. K., Meintjes P. J., 2020, Astronomische Nachrichten, 341, 713

Sinigaglia F., Elson E., Rodighiero G., Vaccari M., 2022, MNRAS, 514, 4205

Sivia D., Skilling J., 2006, Data Analysis: A Bayesian Tutorial, 2 edn. Oxford science publications, Oxford University Press, Incorporated, Oxford

Skilling J., 2004, in AIP Conference Proceedings. AIP, pp 395–405, doi:10.1063/1.1835238

Skilling J., 2006, Bayesian Analysis, 1

Slee O. B., Slee B. O., 1995, AuJPh, 48, 143

Slee O. B., Roy A. L., Murgia M., Andernach H., Ehle M., 2001, AJ, 122, 1172

Slob M. M., Callingham J. R., Röttgering H. J. A., Williams W. L., Duncan K. J., de Gasperin F., Hardcastle M. J., Miley G. K., 2022, A&A, 668, A186

Smith B. D., et al., 2017, MNRAS, 466, 2217

Snellen I. A. G., Schilizzi R. T., De Bruyn A. G., Miley G. K., Rengelink R. B., Röttgering H. J., Bremer M. N., 1998, A&AS, 131, 435

Snellen I. A. G., Schilizzi R. T., Miley G. K., de Bruyn A. G., Bremer M. N., Rottgering H. J. A., 2000, MNRAS, 319, 445

Somerville R. S., Davé R., 2015, ARA&A, 53, 51

Sommovigo L., et al., 2025, ApJ, 990, 114

Speagle J. S., 2020, MNRAS, 493, 3132

Spoelstra T. A. T., Patnaik A. R., Gopal-Krishna 1985, A&A, 152, 38

Springel V., Yoshida N., White S. D. M., 2001, New Astron., 6, 79

Srianand R., Gupta N., Rahmani H., Momjian E., Petitjean P., Noterdaeme P., 2013, MNRAS, 428, 2198

Srianand R., et al., 2022, MNRAS, 516, 1339

Stanghellini C., O'dea C. P., Baum S. A., Dallacasa D., Fanti R., Fanti C., 1997, A&A, 325, 943

Stein Y., et al., 2021, A&A, 655, A17

Stevens J., et al., 2012, arXiv e-prints, p. arXiv:1205.2403

Stevens A. R. H., et al., 2019, MNRAS, 483, 5334

Stewart G. S. C., Shabala S. S., Turner R. J., Yates-Jones P. M., Krause M. G. H., Wong O. I., Power C., Hardcastle M. J., 2025, Publ. Astron. Soc. Australia, 42, e152

Stewart G. S. C., et al., 2026, arXiv e-prints, p. arXiv:2605.10156

Storey-Fisher K., Hogg D. W., Rix H.-W., Eilers A.-C., Fabbian G., Blanton M. R., Alonso D., 2024, ApJ, 964, 69

Su R., et al., 2022, MNRAS, 516, 2947

Sun G.-C., et al., 2025, ApJ, 985, 189

Tanner R., Weaver K. A., 2022, AJ, 163, 134

Tchekhovskoy A., Bromberg O., 2016, MNRAS, 461, L46

Thomas N., Davé R., 2022, MNRAS, 515, 5539

Thomas N., Davé R., Jarvis M. J., Anglés-Alcázar D., 2021, MNRAS, 503, 3492

Thomas N., Whittam I. H., Hale C. L., Morabito L. K., Davé R., Jarvis M. J., Cook R. H. W., 2025, MNRAS, 536, 2873

Thompson A. R., Clark B. G., Wade C. M., Napier P. J., 1980, ApJS, 44, 151

Tingay S. J., Edwards P. G., 2015, MNRAS, 448, 252

Tingay S. J., de Kool M., 2003, AJ, 126, 723

Tingay S. J., et al., 2013, Publ. Astron. Soc. Australia, 30, e007

Tinti S., Dallacasa D., De Zotti G., Stanghellini C., Celotti A., 2005, in Romney J., Reid M., eds, Astronomical Society of the Pacific Conference Series Vol. 340, Future Directions in High Resolution Astronomy. p. 45

Tody D., 1993, in Hanisch R. J., Brissenden R. J. V., Barnes J., eds, Astronomical Society of the Pacific Conference Series Vol. 52, Astronomical Data Analysis Software and Systems II. p. 173

Torniainen I., Tornikoski M., Teräsranta H., Aller M. F., Aller H. D., 2005, A&A, 435, 839

Torniainen I., Tornikoski M., Lähteenmäki A., Aller M. F., Aller H. D., Mingaliev M. G., 2007, A&A, 469, 451

Torrey P., et al., 2015, MNRAS, 447, 2753

Tramacere A., Massaro E., Taylor A. M., 2011, ApJ, 739, 66

Trčka A., et al., 2022, MNRAS, 516, 3728

Tumlinson J., Peeples M. S., Werk J. K., 2017, ARA&A, 55, 389

Turner R. J., Shabala S. S., 2015, ApJ, 806, 59

Turner R. J., Rogers J. G., Shabala S. S., Krause M. G. H., 2018, MNRAS, 473, 4179

Turner R. J., Yates-Jones P. M., Shabala S. S., Quici B., Stewart G. S. C., 2023, MNRAS, 518, 945

Tzioumis A. K., et al., 2010, AJ, 140, 1506

Urry C. M., Padovani P., 1995, PASA, 107, 803

Van Rossum G., Drake F. L., 2009, Python 3 Reference Manual. CreateSpace, Scotts Valley, CA

Veilleux S., Maiolino R., Bolatto A. D., Aalto S., 2020, A&ARv, 28, 2

Vermeulen R. C., et al., 2003, A&A, 404, 861

Véron P., Véron-Cetty M. P., Djorgovski S., Magain P., Meylan G., Surdej J., 1990, A&AS, 86, 543

Villaescusa-Navarro F., et al., 2021, ApJ, 915, 71

Virtanen P., et al., 2020, Nature Methods, 17, 261

Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, Nature Reviews Physics, 2, 42

Vollmer B., Davoust E., Dubois P., Genova F., Ochsenbein F., van Driel W., 2005, A&A, 436, 757

Vries N. d., Snellen I., Schilizzi R., Mack K., 2009, Astronomische Nachrichten, 330, 214

Wada K., Kudoh Y., Nagao T., 2023, MNRAS, 526, 2717

Waddell S. G. H., et al., 2024, A&A, 690, A132

Wagner A. Y., Umemura M., Bicknell G. V., 2013, ApJ, 763, L18

Wall J. V., Shimmins A. J., Merkelijn J. K., 1971, Australian Journal of Physics Astrophysical Supplement, 19, 1

Wang Y., Tuntsov A., Murphy T., Lenc E., Walker M., Bannister K., Kaplan D. L., Mahony E. K., 2021a, MNRAS, 502, 3294

Wang Y., Tuntsov A., Murphy T., Lenc E., Walker M., Bannister K., Kaplan D. L., Mahony E. K., 2021b, MNRAS, 502, 3294

Wang Z., et al., 2025, Publ. Astron. Soc. Australia, 42, e005

Ward S. R., Costa T., Harrison C. M., Mainieri V., 2024, MNRAS, 533, 1733

Wayth R. B., et al., 2018, Publ. Astron. Soc. Australia, 35, e033

Weng S., et al., 2022, MNRAS, 512, 3638

Weng S., Péroux C., Ramesh R., Nelson D., Sadler E. M., Zwaan M., Bollo V., Casavecchia B., 2024, MNRAS, 527, 3494

Weng S., et al., 2025, MNRAS, 539, 1977

Wenger M., et al., 2000, A&AS, 143, 9

Windhorst R. A., Hopkins A., Richards E. A., Waddington I., 1999, in Bunker A. J., van Breugel W. J. M., eds, Astronomical Society of the Pacific Conference Series Vol. 193, The Hy-Redshift Universe: Galaxy Formation and Evolution at High Redshift. p. 55

Winkel N., et al., 2023, A&A, 670, A3

Wolfe A. M., Gawiser E., Prochaska J. X., 2005, ARA&A, 43, 861

Wołowska A., et al., 2021, ApJ, 914, 22

Wright A., Otrupcek R., 1990, PKS Catalog (1990), p. 0

Wright E. L., et al., 2010, AJ, 140, 1868

Yang S., Somerville R. S., Pullen A. R., Popping G., Breysse P. C., Maniyar A. S., 2021, ApJ, 911, 132

Yang G., et al., 2022, ApJ, 927, 192

Yates-Jones P. M., Shabala S. S., Power C., Krause M. G. H., Hardcastle M. J., Mohd Noh Velastín E. A. N., Stewart G. S. C., 2023, Publ. Astron. Soc. Australia, 40, e014

Yoon H., et al., 2025, Publ. Astron. Soc. Australia, 42, e088

York D. G., et al., 2000, AJ, 120, 1579

Young S. A., Turner R. J., Shabala S. S., Stewart G. S., Yates-Jones P. M., 2025, Publ. Astron. Soc. Australia, 42, e100

Yuan F., Narayan R., 2014, ARA&A, 52, 529

Zanni C., Murante G., Bodo G., Massaglia S., Rossi P., Ferrari A., 2005, A&A, 429, 399

Zhai S., Gloudemans A. J., Gürkan G., Ballieux F. J., Hardcastle M. J., De Gasperin F., Röttgering H. J. A., 2025, A&A, 695, A7

Zhao X., Marchesi S., Ajello M., Cole D., Hu Z., Silver R., Torres-Albà N., 2021, A&A, 650, A57

Zheng X., Zhang Y., Röttgering H., 2024, A&A, 686, A169

Zinn P.-C., Middelberg E., Norris R. P., Dettmar R.-J., 2013, ApJ, 774, 66

da Cunha E., Charlot S., Elbaz D., 2008, MNRAS, 388, 1595

de Vries W. H., Barthel P. D., O'Dea C. P., 1997, A&A, 321, 105

de Witt A., Charlot P., Gordon D., Jacobs C. S., 2022, Universe, 8, 374

van Breugel W., Miley G., Heckman T., 1984, AJ, 89, 5

van Haarlem M. P., et al., 2013, A&A, 556, A2

van Velzen S., Falcke H., Körding E., 2015, MNRAS, 446, 2985

von Hoerner S., 1973, ApJ, 186, 741

von Toussaint U., 2011, Rev. Mod. Phys., 83, 943