From star-formation to recombination: expanding our view of the radio recombination line universe

Emig, K.L.

From Star-Formation to Recombination

Expanding our View of the Radio Recombination Line Universe

From Star-Formation to Recombination
Expanding our View of the Radio Recombination Line Universe

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus Prof. dr. ir. Hester Bijl,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 29 April 2021 klokke 16:15 uur

door

Kimberly Lynn Emig

geboren te
Norristown, Pennsylvania
in 1989

Promotiecommissie

Promotores:	Prof. dr. A. G. G. M. Tielens
	Prof. dr. H. J. A. Röttgering
Overige leden:	Prof. dr. M. Haverkorn (Radboud Universiteit)
	Prof. dr. H. J. van Langevelde (Universiteit Leiden)
	Dr. F. J. Lockman (Green Bank Observatory)
	Prof. dr. R. Morganti (Rijksuniversiteit Groningen)
	Prof. dr. P. P. van der Werf (Universiteit Leiden)

ISBN 978-94-6419-187-5

A digital copy of this thesis can be found at: https://openaccess.leidenuniv.nl

Cover design: Veronica Emig, Skylar Tibbits, and Kimberly Emig.
The front cover shows the central (170 pc) starburst of the nearby galaxy NGC 4945 in ALMA 93 GHz continuum. The emission primarily arises from free electrons in ionized gas; at 2 pc resolution, super star clusters appear as point like sources. The back cover shows diffuse ionized gas, as observed with LOFAR at 142 MHz, in an area of roughly $2^\circ \times 4^\circ$ in the Cygnus X star-forming region of the Milky Way.

To women and girls everywhere

Contents

1 | Introduction

Galaxies are gravitationally bound systems of stars, gas, dust, and a central supermassive black hole, all of which reside in a dark matter potential. A major question guides current astrophysical research: what regulates the evolution of galaxies? The characteristics of the interstellar medium (ISM), the formation of stars and their feedback on the ISM, events driven by the central black hole, accretion of intergalactic gas, galaxy mergers and their environment, and the dark matter potential in which a galaxy resides may all influence the lifecycle of a galaxy.

The gaseous and dusty components amongst the stars of a galaxy are referred to as the interstellar medium. In the ISM of a galaxy, clouds with characteristic phases typically form (see Section 1.1). The basic lifecycle of interstellar clouds is illustrated in Figure 1.1. Cold, dense self-gravitating molecular clouds collapse to form new stars. As stars evolve, stellar ejecta influence the structure, physical properties, and chemical composition of the ISM. Gas and dust that has been heated, stirred up, and enriched, might cool and condense into clouds, setting the conditions for new star formation.

The origin and evolution of galaxies are closely tied to the cyclic feedback processes between stars and the ISM. Stellar feedback will disrupt molecular clouds thereby stopping star formation (negative feedback) while at the same time gas is compressed in shells, promoting gravitational instabilities (positive feedback) (Elmegreen & Lada 1977; Hopkins et al. 2014). Phases of the ISM reflect radiative and mechanical energy input by massive stars into their environment (Field et al. 1969; Weaver et al. 1977; McCray & Snow, T. P. 1979; Wolfire et al. 1995). Stellar radiation dissociates, ionizes, and heats gas, and together with the cooling law of atomic gas, sets the temperature and controls the cold and warm phase structures of the ISM. Mechanical energy — input by protostellar outflows, stellar winds, and supernovae — stirs up the medium, controls the pressure of the gas and transfers material from one phase to another. Numerical simulations reveal the active interplay of radiative and mechanical feedback by massive stars (Walch et al. 2012; Kim et al. 2013; Dale et al. 2014; Walch & Naab 2015; Walch et al. 2015; Kim et al. 2018; Haid et al. 2018).

The aim of this thesis is to explore the characteristics of the interstellar medium, on global (galactic) scales down to sub-cloud (pc) scales. We explore new methods to investigate the ISM, through radio recombination line observations, and develop the tools and strategies needed to process new low-frequency observations. We also infer the presence of massive stars and characterize their influence on the ISM, by determining gas physical conditions. In subsequent chapters, we specifically address

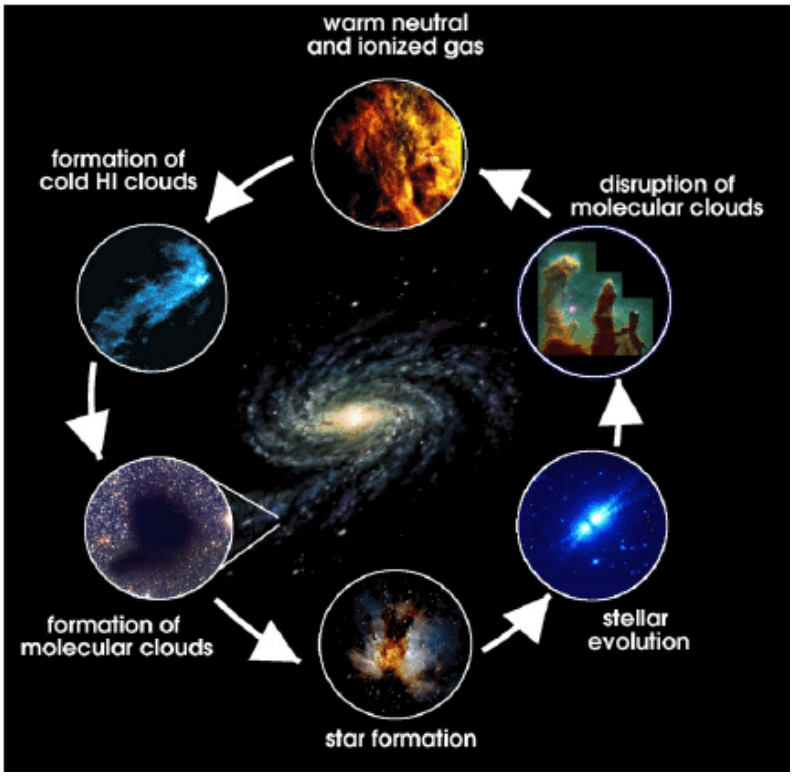


Figure 1.1: The lifecycle of interstellar clouds (Groppi et al. 2009). Starting from the central bottom panel, stars form in dense molecular clouds. As these stars evolve, they develop winds enriching the ISM with nucleosynthetic products. The radiative and mechanical feedback from stars and supernovae disrupt molecular clouds, heating and rarefying the medium. As massive stars die out, warm gas eventually cools. In diffuse clouds, dust grains and a complex chemical network create molecular species, leading to the formation of cold and dense molecular clouds.

the questions:

- How does low-density ionized gas affect the evolution of the massive, galactic star-forming region, Cygnus X? Are the same fingerprints present in surveys of low-density ionized gas in our Galaxy?
- What are the properties of star formation (star clusters) in the central starburst of the galaxy NGC 4945?
- Can the ISM be explored outside of the local universe through radio recombination line observations? What are the ISM properties of a dwarf-like galaxy at $z = 1.1$?
- What techniques are best suited to detect faint radio recombination lines (at an unknown redshift) in extragalactic sources?

In the sections of this chapter, we frame these concepts for the reader, guiding how each of these questions furthers our understanding of processes influencing the lifecycle of a galaxy. We describe phases of the ISM and its basic tracers (Section 1.1), star formation in star clusters (Section 1.2), radio recombination lines as a method to observe gas in the ISM, stars and stellar feedback (Section 1.3), observatories which have enabled the investigations reported in this thesis (Section 1.4), a brief outline of the contents of this thesis in Section 1.5, and a future outlook which builds on the advancements facilitated by this thesis (Section 1.6).

1.1 Phases of the Interstellar Medium

The ISM contains a number of phases — diffuse clouds, warm intercloud gas, and hot intercloud gas — characterized by different temperatures, densities, and ionization fractions and that exist in pressure equilibrium (Tielens 2005). We show a schematic of the ISM in a typical Galactic region in Figure 1.2. We first briefly describe the characteristics of these phases, and in additional sections, we elaborate on cold diffuse clouds, warm ionized media and extragalactic tracers of these phases.

Hydrogen in its atomic state, traced by the H I 21 cm line (van de Hulst 1945), constitutes a large reservoir of gas ($M \sim 5 \times 10^9$ M$_\odot$) in the ISM. It exists in two stable phases primarily as a result of cooling by the fine structure lines of [C II] and [O I] and heating by FUV photons ($6 - 13.6$ eV). Diffuse ($n \sim 50$ cm^{-3}) H I clouds — the cold neutral medium (CNM) — contain about 40% of the H I mass with temperatures of 70 K (Heiles & Troland 2003). The remainder of H I is in the warm neutral medium (WNM) — a phase with the largest filling factor (∼0.4) in the disk — with typical densities of $n \sim 0.5$ cm^{-3} and temperatures of $T \sim 8000$ K, although about half of the WNM lies in the thermally unstable regime with 500 — 5000 K (for a review see Dickey & Lockman 1990; Kalberla & Kerp 2009). When hydrogen is ionized, the gas heats up and temperatures stabilize near $T \sim 10^4$ K as a result of cooling from collisionally excited lines such as [O III] 4959/5007 Å and [N II] 6548/6553 Å, and low-density ionized gas ($n \sim 0.1$ cm^{-3}) — the warm ionized medium (WIM) — fills a considerable volume (filling factor of 0.25) of the disk (for

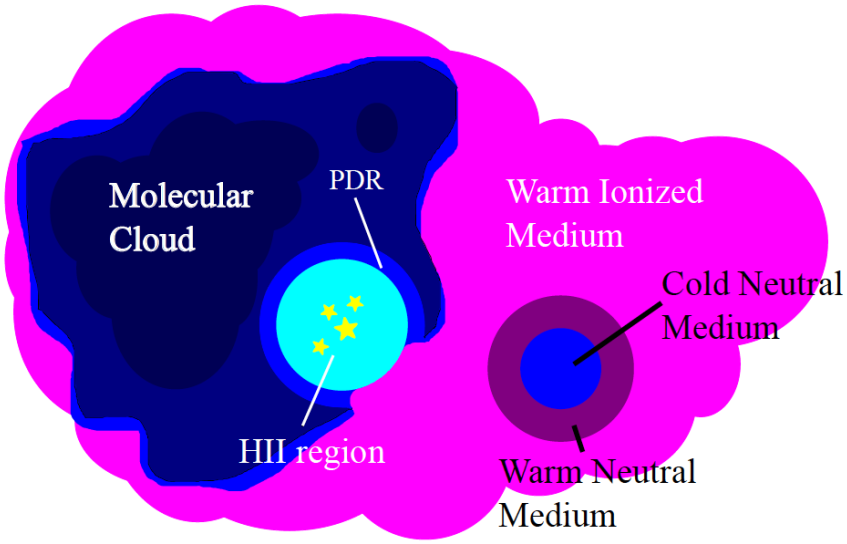


Figure 1.2: A schematic of the classic structure of the interstellar medium in a normal galaxy (Salgado Cambiazo 2015). Phases of the ISM, including the cold neutral medium, warm neutral medium, and warm ionized medium, are depicted. Also shown is a giant molecular cloud in which stars have formed and have created an ionized H II region. Notice the photodissociation region (PDR) at the ionized gas and molecular cloud interface, as well as the cold neutral material on the surfaces of molecular clouds.

a review see Haffner et al. 2009). The hot ($T \sim 10^6$ K) intercloud medium (HIM) of tenuous ($n \sim 10^{-3}$ cm^{-3}) gas has been revealed through X-ray and UV observations. Within the disk it originates through shocks driven by stellar winds and supernovae, and in some cases, it vents out of the disk. Hot gas fills the volume of the halo. Star formation activity is limited to the cold ($T \sim 10$ K) denser ($n \sim 200$ cm^{-3}) molecular (H_2) clouds traced in bulk by CO, i.e., the ^{12}CO(1-0) transition at 2.6 mm. While molecular clouds are in pressure equilibrium with the ISM, they are self gravitating, and thus deep inside the pressure can be much higher.

1.1.1 The Cold, Diffuse ISM

Since warm hydrogen dominates the emission of H I 21 cm observations, the CNM can only be probed when conditions lend themselves to self- and continuum absorption studies (e.g., Dickey & Lockman 1990). Temperatures of cold H I range from 40 – 100 K, weighted by column density the median temperature is 70 K (Heiles & Troland 2003; Dickey et al. 2009). Densities of $n \sim 60$ cm^{-3} (Gibson et al. 2000; Heiles & Troland 2003) have been inferred from a representative pressure of $P/k \approx 3800$ K cm^{-3} (Jenkins & Tripp 2001; Jenkins et al. 2011). The CNM is distributed within 150 pc of the plane (e.g., Kalberla & Kerp 2009). While the classic CNM cloud size is a few pc, observations have revealed a considerable fraction with sheet-like and filament morphology and some regions with globular substructure on sub-pc (< 0.1 pc) scales (Heiles 1997; Gibson et al. 2000; Heiles & Troland 2003). These sizes contrast markedly with the 600 pc cloud sizes of the WNM (e.g., Kalberla & Kerp 2009).

The 21 cm line does not constrain the physical conditions of the gas (e.g., temperature, density) well, making it hard to identify the full extent and physical conditions of the CNM. Observation are limited to select lines of sight where H I can be observed in absorption and/or at high (less confusing) galactic latitude. Optical and UV absorption line studies — which are good probes of the physical conditions — are limited to pinhole experiments towards bright, nearby stars with only low column densities of gas. While the CNM shines in [C II] 158 μm emission, so do denser regions of PDRs and ionized gas (Heiles 1994), making it difficult to elucidate the cold, diffuse component.

A key missing link in the galactic lifecycle is thus the interrelationship of neutral gas and molecular gas. The traditional three-phase model of the ISM (Mckee & Ostriker 1977) cannot reconcile the significant presence of luke-warm ($T \sim 500 - 5000$ K), thermally unstable H I gas in the ISM (Heiles & Troland 2003). In addition, the presence of "translucent" clouds has long been recognized (van Dishoeck & Black 1988) and its general importance has been driven home by γ-ray and far-IR studies, indicating that much of the molecular gas mass is in a form not probed by CO (so-called CO-dark molecular gas) (Grenier et al. 2005; Planck Collaboration et al. 2011). For H I diffuse clouds, hydrogen is atomic and carbon is ionized (C+). In CO-dark molecular gas, hydrogen is in H_2 but C+ rather than CO is the dominant form of carbon (Visser et al. 2009; Wolfire et al. 2010). The major mass constituents of our Galaxy of cold H I and CO-dark phases have so far eluded detailed characterization.

1.1.2 Warm Ionized Media

The warm ionized medium, with a typical density of 0.1 cm^{-3}, occupies a large volume (filling factor of 0.25) in the disk of our Galaxy, is characterized by large ($z \sim 1$ kpc) scale heights, and contains about $M_{\rm ion} \sim 10^9$ $M_\odot$(for a review see Haffner et al. 2009). Within the plane of the Galaxy, ionized gas is found in (i) compact ($R <$ a few pc) H II regions ($n > 10^3$ cm^{-3}) associated with O stars in the early stages of their interaction with the ISM when they are still embedded in (i.e., interacting with) their birth environment, (ii) the diffuser ionized gas due to stars drifting away from the association and/or when large (wind) bubbles have broken open and vented their ionized gas (and hot plasma) into the environment, and (iii) an ionized medium surrounding neutral clouds. Mezger (1978) first emphasized the importance of an ionized gas phase that fills a substantial portion of the ISM in the plane of the Milky Way with a density of $\sim$5 cm^{-3} and a temperature of 7000 K. Recent studies of the [N II] fine-structure lines indicate that an even denser phase of ionized gas ($\sim$50 cm^{-3}) is also prominent in the ISM of the Milky Way as well as other galaxies (Goldsmith et al. 2015; Pineda et al. 2019).

Massive stars of type O and earlier than B3 produce EUV photons ($E > 13.6$ eV) capable of photoionizing gas. Although the ionization of the WIM involves less than 1% of the produced ionizing photons, it does require that the ISM is structured in a way (also considering the motion of stars) that enables EUV photons to escape their immediate environment. Outstanding open questions in studies of diffuse ionized gas include (a) possible heating sources of the WIM by non-ionizing means (photoelectric heating from dust grains, dissipation of turbulence, shocks and cooling hot gas), (b) the distribution, morphology and physical properties of diffuse ionized gas in the midplane, and (c) how many ionizing photons escape their immediate environments as well as the disk and galaxy. In addition, the characteristics of the ionized gas phases, their distribution in the galaxy, and their relationship to each other and to compact HII regions and their ionizing stars is not well understood.

Optical lines (e.g., of Hα, [N II], [O II] and [S II]) and pulsar dispersion measures (DM) have been major workhorses in characterizing the WIM. Somewhat denser gas has been uncovered through thermal radio continuum and high-frequency radio recombination lines (for more details, see Section 1.3). However, they have not (yet) provided answers to remaining open questions. Particularly close to the Galactic plane, optical lines are heavily attenuated, and the DM only provides an average density along the line of sight, without much information about inhomogeneities. Through these means, physical properties cannot be determined (outside of classic H II regions) that would quantify the direct impact of mechanical and radiative processes from (massive) stars.

1.1.3 Extragalactic Tracers of the ISM

Looking outside the Milky Way, in nearby galaxies H I 21 cm observations probe the majority of the warm neutral component. FIR atomic fine structure lines of [O I] 63 μm and [C II] 158 μm trace neutral components, but [C II] also arises from cold neutral and warm ionized phases. Molecular gas content can be observed in bulk

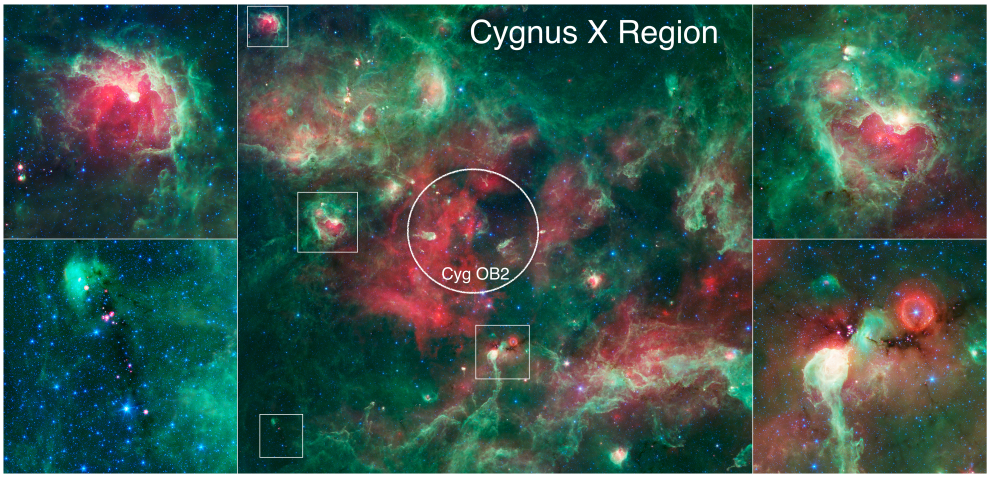


Figure 1.3: The Cygnus X star-forming region, a bubbling cauldron of star birth surrounding the Cyg OB2 association — observed with the *Spitzer* Space Telescope: 3.6 μm in blue, 4.5 μm in blue-green, 8.0 μm in green, and 24 μm in red. PDRs traced by UV-heated PAHs appear bright in green-white. Massive stars that have blown bubbles and cavities in the dust and gas are evident where green-white PDR emission surrounds a bubble of glowing red emission from warm dust (ionized gas). The brightest, yellow-white regions are warm centers of star formation. Tendrils of dust appear green, and stars generally appear as blue point sources. The pillar-like and elongated features directed away from Cyg OB2 have been shaped by the stellar radiation and winds from this massive ($M_\star \approx 2 \times 10^4$ M$_\odot$) association. The boxed zoom-in regions show massive star(s) (formation) in AFGL 2636 (upper left) and DR22 (upper right). The lower left and right (DR15) images show clouds that are so thick to be absorbed at the mid-infrared wavelengths of Spitzer. Young stars, visible as red points, are buried in the dark clouds. Image credit: NASA / JPL-Caltech / Harvard-Smithsonian.

by CO (and isotopes and transitions up the ladder) and several other molecules such as HCN, CS, HCO^+, SO, etc. As in the Milky Way, directly detecting the emission of cold diffuse clouds is difficult. Observations of the [C I] 370 and 609 μm lines help, and cold neutral gas can be inferred using [C II] or FIR dust continuum observations together with CO and H I (e.g., Bolatto et al. 2013a; Herrera-Camus et al. 2017). Ionized gas can be studied to deep emission measures through hydrogen recombination lines at $\lambda \lesssim 2$ μm (i.e., Lyα, Hα, Hβ, Paα) and other optical emission lines such as [N II], [O II] and [S II], and through FIR fine structure lines of [C II] 158 μm, [N II] 122 and 205 μm and [O III] 52 and 88 μm. However, the WIM (and/or diffuse ionized gas) is difficult to differentiate between the discrete (or unresolved) sources of stars in face-on galaxies (e.g., Oey et al. 2007). Studies of edge-on systems have been particularly illuminating for the WIM (e.g., Rossa & Dettmar 2003a,b; Levy et al. 2019).

When we look outside of the local universe, the neutral gas content is relatively unknown. Neutral H I has been detected in individual systems to $z \lesssim 0.2$. The neutral gas content of the universe is probed in large numbers through e.g., Lyα and Mg II absorption systems in quasar spectra, though studying individual systems responsible for absorption has been difficult and the origin of the absorption systems remains under investigation (Wolfe et al. 2005). H I 21 cm absorption from intervening and associated systems has so far been detected in just over 200 systems (e.g., Kanekar & Briggs 2004; Morganti & Oosterloo 2018). [C II] seems to predominantly trace the ionizing photon luminosity through PDRs emission in galaxies (though not well for galaxies with FIR luminosities of $> 10^{11}$ $L_\odot$); with ALMA, galaxies of $z > 4$ can be investigated in this manner (but it is difficult in intermittent redshifts) (e.g., Carilli & Walter 2013). Galaxies have traditionally been probed across all redshifts through some of the brightest emitting lines of ionized gas, though the contribution from diffuse gas and ionized gas is not often possible to distinguish. Before ALMA, CO molecular gas was only detected in ∼200 systems, but this has steadily been increasing revealing the molecular gas history of the universe (e.g., Decarli et al. 2016). In conclusion, it is challenging to trace the ISM of galaxies outside of the local universe.

1.2 (Clustered) Star Forming Regions

The vast majority of massive stars form in clusters (we refer the reader to reviews by Lada & Lada 2003; Portegies Zwart et al. 2010; Krumholz 2014; Krumholz et al. 2019). 70% of O stars reside in young clusters or associations (Gies 1987). Furthermore, most of the field population can be identified as runaways (de Wit et al. 2005). Massive young clusters tend to form with compact (∼1 pc) sizes or in somewhat extended regions (∼7 pc) which may be unbound (associations). Giant molecular clouds, the birth sites of stars and star clusters, have typical sizes of 50 pc and masses of $\sim 10^5$ $M_\odot$. Filaments and clumps form within them through the combined action of turbulence and self-gravity, and massive clusters tend to form at the intersection of filaments. Cloud collisions may also be important for their formation.

In the Galaxy, the nearest young cluster that has formed O stars is 414 ± 7 pc away (Menten et al. 2007) in Orion. The Orion Nebula Cluster (ONC), with a total stellar mass of ∼4000 $M_\odot$ and age of $1 - 2$ Myr (e.g., Zari et al. 2017, and references

therein), is one of the most comprehensively studied regions of massive star formation. Its massive stars, an O7 star and two early B types, have created a bubble filled with hot X-ray gas and warm ionized gas that is a few pc in diameter and sits at the edge of its parent molecular cloud (e.g., Pabst et al. 2019). In the association Cyg OB2 ($d \approx 1.5$ kpc), massive star formation ($M_\star \sim 2 \times 10^4$ M$_\odot$) has occurred in a region of relatively low mass surface density and the system is not gravitationally bound (see Figure 1.3). Cyg OB2 has an age spanning 3-5 Myr and thus is no longer actively forming stars. While Cyg OB2 has clearly been influential in dispersing its molecular cloud — a ∼100 pc region has been cleared and ionizing radiation continues to eat away at the nearby clouds — evidence for (subdominant) triggered star-formation is also present in small globules and pillars (e.g., Schneider et al. 2016). Moreover, the Cygnus X region is accompanied by new regions of active star formation spread across the entire complex (see Figure 1.3).

The most massive ($> 10^5$ M$_\odot$), young clusters — super star clusters (SSCs) — in the local universe are often found in starbursting regions and merging galaxies (e.g., Portegies Zwart et al. 2010; Whitmore et al. 2010; Linden et al. 2017). The efficiency with which matter is converted into stars is higher in SSCs — $\epsilon \sim 0.5$ compared with the typical value of ∼0.05. They provide key insights into galaxy formation as cluster formation is expected to be a main mode of star formation in high intensity, starburst environments (e.g., Kruijssen 2012) and thus possibly dominant during the peak epoch of star formation in the universe ($z \sim 1 - 3$). The presence of globular clusters also indicates that star formation in massive clusters was a hallmark of the era of galaxy formation.

In nearby starbursting galaxies ($d < 4$ Mpc), such as M 82 (e.g., Tsai et al. 2009), NGC 253 (Leroy et al. 2018) and NGC 4945 (Emig et al. 2020b), SSCs populate the nuclear starbusting regions. In M 82, 15 SSCs or more are found within the central starburst (see Figure 1.4; McDonald et al. 2002). The concerted efforts are capable of driving galaxy scale outflows, reflecting the profound impact they have on lifecycle of galaxies (see Figure 1.4).

1.3 Radio Recombination Lines

Radio recombination lines provide a powerful tool to study the characteristics of neutral and ionized phases of the ISM. As ions and electrons recombine to form atoms, electrons left in high principal quantum numbers and which cascade to lower energy levels (or may be excited to higher energy levels) are observable through the spectral signature of radio recombination lines (RRLs). RRL transitions in these Rydberg states have (some of the) properties of the Bohr model of the atom. RRLs are typically observed at the radio frequencies of 350 GHz ($\lambda \sim 0.86$ mm) down to 10 MHz ($\lambda \sim 30$ m), corresponding to low energy α-transitions ($\Delta\mathsf{n} = 1$) of principal quantum numbers $\mathsf{n} = 26$ – 869 [1]. At such high principal quantum numbers, for example at the largest detected (δ; $\Delta\mathsf{n} = 4$) transition in space of $\mathsf{n} = 1009$ (Stepkin

[1] In this notation, n represents the principal quantum number of the final state

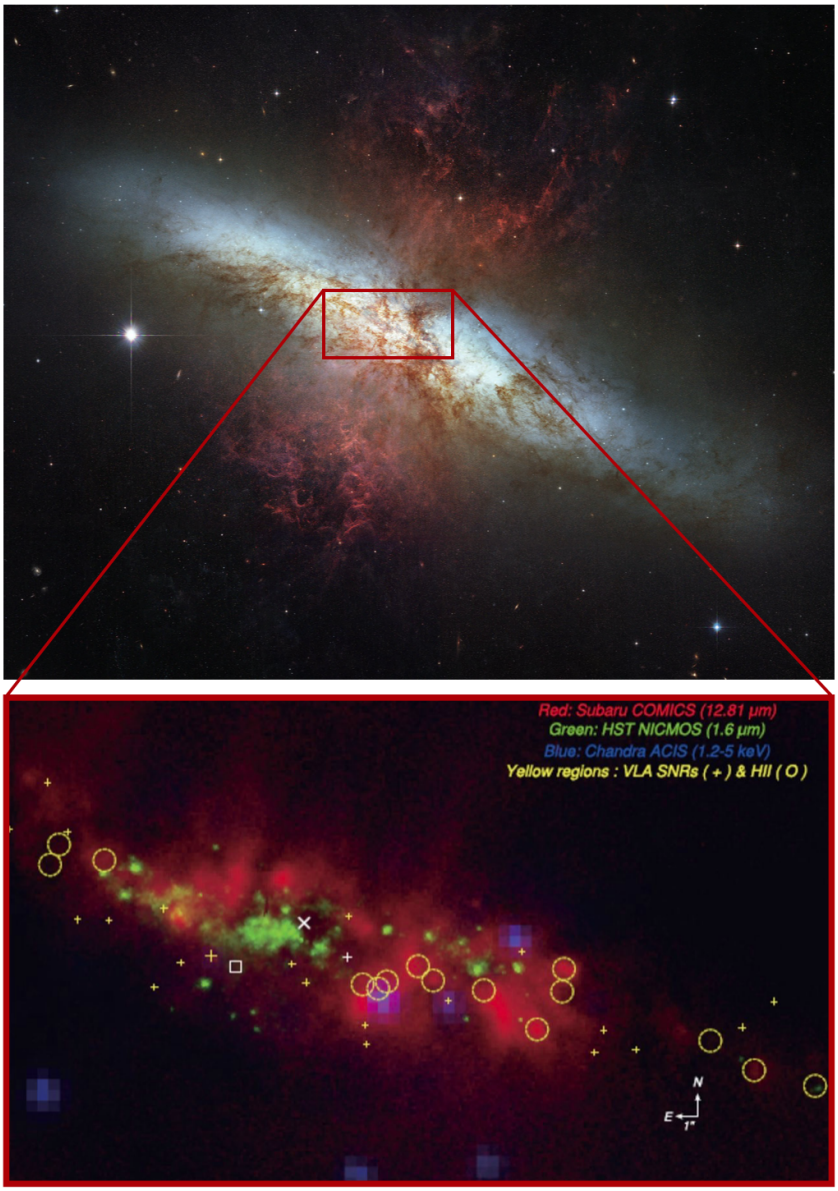


Figure 1.4: In the top panel, the disk of the nearby (3.5 Mpc) edge-on galaxy, M82, glows in white-blue and the prominent galactic winds from the nucleus are traced by the PAH emission in red. Super star clusters (yellow circles) in the nucleus (in the bottom panel) drive the outflow. Image credits: top panel NASA / ESA / The Hubble Heritage Team and bottom panel from Gandhi et al. (2011).

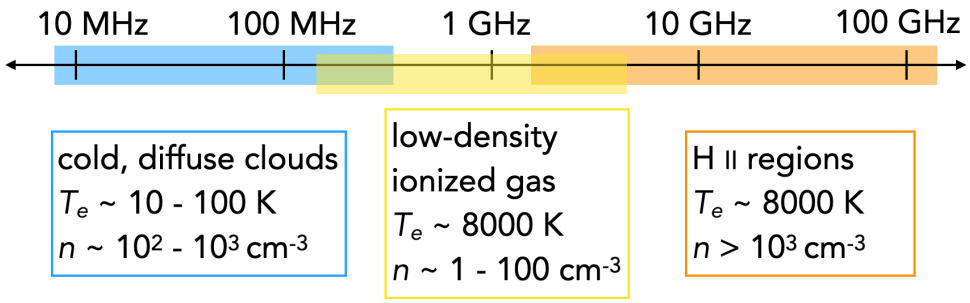


Figure 1.5: The gas components and their frequencies traced by radio recombination lines. Emission from cold diffuse clouds is probed by $\nu \lesssim 500$ MHz, low-density ionized gas by 300 MHz $\lesssim \nu \lesssim 5$ GHz, and classic H II regions by $\nu \gtrsim 1$ GHz.

et al. 2007), the Bohr diameter of an atom[2] reaches $(2n^2\hbar)/(Zk_e e^2 m_e) \approx 110$ μm — for scale, the average width of a human hair is roughly 75 μm!

Observations have now sampled the full frequency coverage of the RRL spectrum, primarily through hydrogen, carbon, and helium RRLs. Broadly speaking, three gas components have been found to dominate line emission, as shown in Figure 1.5. At high frequencies ($\gtrsim 10$ GHz), recombination lines arise from spontaneous transitions in dense ($n_e \sim 10^3 - 10^4$ cm^{-3}) ionized gas — conditions associated with compact H II regions and PDRs. At the intermediate frequencies of 0.3 – 1 GHz, emission by a warm and low-density ($n_e \sim 1 - 100$ cm^{-3}) ionized component dominates. Whereas at low frequencies ($\lesssim 250$ MHz), recombination lines arise in the cold-neutral medium stimulated by external radiation fields; carbon is typically the most prominent species, and spectral lines can be observed in either absorption or emission.

In this section we introduce and discuss the physics, primary use cases, and at times some historical information of the three different frequency regimes of RRL emission. Lastly, we give special emphasis to extragalactic RRLs with additional background. We refer the reader to "Radio Recombination Lines: Their Physics and Astronomical Applications" (Gordon & Sorochenko 2002) as well as Roelfsema & Goss (1992) for additional information.

1.3.1 High frequency recombination lines

The most common observations of RRLs are Galactic in origin and arise from spontaneous transitions found in optically-thin free-free emitting gas. When the electron temperature T_e of the thermal free-free emission from ionized gas accurately characterizes the relative populations of electrons in bound atomic levels, the system is well approximated by local thermodynamic equilibrium (LTE). Ionizing radiation is balanced by recombinations and line emission is described by the Saha-Boltzmann distribution. This description characterizes RRLs at frequencies of around 5 – 20 GHz arising from typical H II regions with densities of $10^3 - 10^4$ cm^{-3} and electron temperatures of ∼7000 K.

[2] $\hbar$ is the reduced Planck constant, Z is the atomic number, k_e is the Coulomb constant, e is the charge of an electron, and m_e is the mass of an electron.

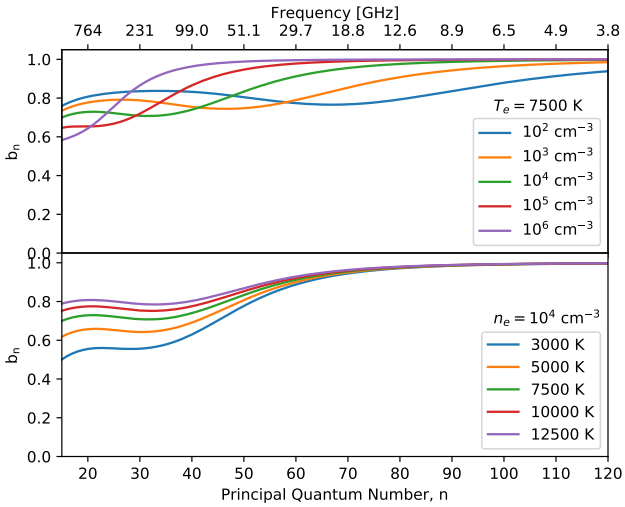


Figure 1.6: The departure (from LTE) coefficient of spontaneous emission, b_n, shown as a function of principal quantum number (frequency). Top shows the variation with density for a single temperature ($T_e = 7500$ K. Bottom shows the variation with temperature for a single density ($n_e = 10^4$ cm^{-3}).

As the recombination cross section of a Bohr atom rapidly decreases ($\propto \mathsf{n}^4$) towards smaller principal quantum numbers and thus higher frequencies, of order $\mathcal{O}(100$ GHz), collisions between protons and electrons become less important, and the observed line intensity is not exactly set by the kinetic motion of the electrons. Radiative processes, which dominate the smallest principal quantum numbers ($\mathsf{n} \lesssim 10$), influence the level populations; smaller n are underpopulated as compared with a Boltzmann distribution. The population rates into an energy level are not exactly balanced by rates out of the level, and a correction coefficient (b_n) for a departure from LTE is necessary. Multiplicative b_n factors to the line intensity, as shown in Figure 1.6, at a given n depend (mildly) on gas conditions found in typical H II regions; however for low-density gas, they become more important and variable across frequency.

The physics that is extracted through high-frequency recombination lines has been a major workhorse in studies of photo-ionized gas and massive star formation in the local universe. The intensity of emission from a RRL is dependent upon the electron temperature and the emission measure (EM)[3] of the ionized atoms. As the ionization balance is in steady state, recombinations are balanced by ionizations. (We refer the reader to Chapter 3 for more detailed equations.) The ionizing photon rate which maintains the ionized state of the gas is also directly proportional to the EM (and temperature through the recombination coefficient). Connecting these properties, the RRL intensity effectively measures the ionizing photon rate. In obscured regions, which may be in our Galactic plane or in extinguished regions in other (especially local, edge-on) galaxies, RRLs provide access to total ionizing luminosities (and thus star formation rates) free from dust extinction. High-frequency RRLs can also be used

[3] $EM = \int n_e n_i \, \mathrm{d}\ell$ where n_e and n_i are the volume number density of electrons and ions, respectively, and ℓ is the pathlength integral.

to extract the electron temperature when combined with continuum observations (at the same frequency); since the volume of gas emitting line and continuum emission is the same, dependencies on the EM, beam filling factor and distance to the source cancel out when taking the line to continuum ratio and only the temperature remains to be directly solved. With the kinematic information provided by the lines and if the region is spatially resolved, components can be distinguished and thus pathlengths determined in order to also infer electron densities. Deeper observations may also detect helium RRLs (offset by just 122 km s^{-1} from hydrogen lines) which follow these same physical principals for emission; by comparing the integrated line strengths, the helium abundance is determined.

1.3.2 Low frequency recombinations lines from cold partially ionized gas

Recombination lines at low frequencies ($\lesssim 250$ MHz) inherently arise from relatively cold gas, $T_e = 10 - 100$ K, with low electron densities, $n_e = 0.01 - 0.1$ cm^{-3} (Shaver 1975b) — characteristic conditions of the diffuse ISM. Since background radiation ($I_{\rm BC}$) typically dominates over the continuum emission of the cloud at low frequencies, stimulated effects dictate the RRL line intensity (Goldberg 1966), and the optical depth of the line is approximated by $-\tau_{\rm line} \approx I_{\rm line}/I_{\rm BC}$. Under these conditions, the observed emission depends on temperature as $T_e^{-2.5}$ (Shaver 1975b; Salgado et al. 2017b) and thus preferentially traces cold clouds. Low-frequency RRLs are quite faint — emitting/absorbing 0.1 – 0.01% of the continuum — which attributes to the difficulty in detecting them; however, as the lines appear more frequently at low frequencies (see Figure 1.9), line stacking (e.g., Emig et al. 2020a) can aid in their detection.

Carbon is typically the brightest RRL emitter at the lowest frequencies for a number of reasons. Carbon has a lower ionization potential ($E = 11.3$ eV) than hydrogen (13.6 eV) and remains ionized when the bulk of the gas (traced by hydrogen) is neutral. In the cold, diffuse ISM where ionizing radiation can not excite hydrogen, the [C II] $^2P_{1/2}$ – $^2P_{3/2}$ fine structure line at 158 μm serves as the dominant coolant (Tielens 2005) and temperatures stabilize near 70 K (Heiles & Troland 2003). Carbon is abundant (compared with other "metals") in the ISM (Cardelli et al. 1996). Ionized carbon at these temperatures also has the unique property of dielectronic capture (Watson et al. 1980), in which an electron recombines with a carbon ion by simultaneously exciting the [C II] $^2P_{1/2}$ – $^2P_{3/2}$ fine structure line, leaving additional (compared with LTE) carbon atoms with electrons in high Rydberg states. Dielectronic capture greatly enhances RRL emission at high quantum levels ($\mathsf{n} \gtrsim 250$; Salgado et al. 2017a).

Carbon RRLs are an unprecedented physical diagnostic of the diffuse ISM. Being maser-like, the distribution of the quantum level population, characterized by the correction factor for stimulated emission β_{n} and the departure coefficient b_{n}, is strongly dependent upon the temperature and density conditions of the gas (see Figure 1.7; Salgado et al. 2017a,b). This product, $b_{\mathsf{n}}\beta_{\mathsf{n}}$, which is directly proportional to the integrated line strength, is set by the relevant interactions which influence atomic physical processes; only recently have computing resources allowed for (i) the full

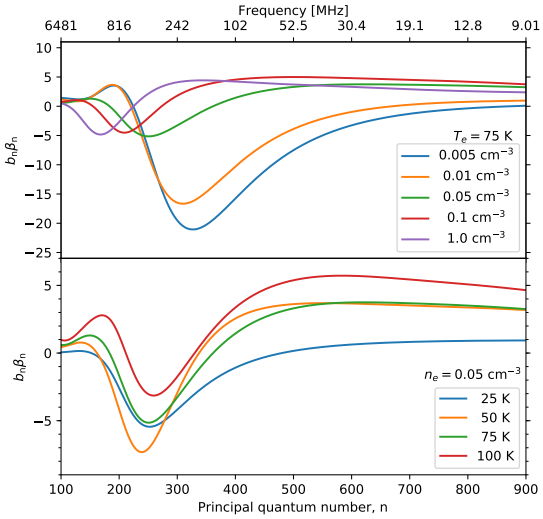


Figure 1.7: The strong dependence of carbon RRL emission (as a function of principal quantum number) on the physical conditions of cold diffuse clouds, for which the product of the departure and stimulation coefficients ($b_n\beta_n$) is directly proportional to the line intensity. Carbon RRLs show an enhancement in emission — where negative values of $b_n\beta_n$ indicate emission and positive values indicate absorption — at small principal quantum numbers (high frequencies), and they are often observed in absorption at the lowest frequencies.

angular momentum treatment for carbon and (ii) with transitions to large enough principal quantum number to be included (Salgado et al. 2017a). Because the density and temperature are constrained simultaneously by both the relation for the line intensity and the $b_n\beta_n$ coefficients, the pathlength can also be uniquely determined. (We refer the reader to Chapter 4 for the detailed equations.) The capability to derive the density, temperature and pathlength in the diffuse ISM is revolutionary because traditional probes of the ISM have not, apart from temperature, provided this physical information (e.g., Heiles & Troland 2003; Roshi & Kantharia 2011).

Detailed studies of gas along the line of sight to Cassiopeaia A, a supernova remnant in our Galaxy and one of the brightest radio sources in the sky at low frequencies, have demonstrated and shaped our understanding of carbon RRLs (Kantharia et al. 1998, and references therein) (Asgekar et al. 2013; Salas et al. 2017; Oonk et al. 2017; Salas et al. 2018; Chowdhury & Chengalur 2019). Large-area surveys in the Galaxy have also demonstrated the ubiquity of low-frequency carbon RRL emission, especially using the Ooty Radio Telescope (Anantharamaiah 1985a; Erickson et al. 1995; Kantharia & Anantharamaiah 2001; Roshi et al. 2002). However, because of the spatial resolutions (with the necessary sensitivity) and limited frequency coverage to constrain gas properties, a clear characterization of the origin of the emission in the ISM has not been made. Moreover, the important implication of stimulated emission possibly detectable in galaxies outside of the local universe (Shaver 1978) — since the line emission is proportionally to the continuum intensity of background radiation —

has not (yet) been fully utilized in extragalactic objects.

Ionized hydrogen may also be stimulated by an external radiation field at low-frequencies (e.g., Shaver 1975b,a; Anantharamaiah 1985a; Roshi & Anantharamaiah 2000, 2001). Indeed hydrogen RRLs have been observed from a cold medium (Gordon & Sorochenko 2002, and references therein) (Lockman 1989; Roshi & Anantharamaiah 1997, 2001; Oonk et al. 2017, 2019; Emig et al. 2020a). Cold ionized hydrogen layers are thought to originate in photoevaporating flows (typically when observed in conjunction with a broad hydrogen RRL of a typical H II region) or maintain ionization through cosmic rays thus probing cosmic ray ionization (for a review see Gordon & Sorochenko 2002).

1.3.3 Hydrogen RRLs from warm low-density ionized gas

At the intermediate frequencies of 0.3 – 1 GHz, a warm (T∼7000 K) and low-density ($n_e \lesssim 10^3$ cm^{-3}) ionized component dominates hydrogen RRL emission in the plane of the Galaxy (Shaver 1976; Lockman 1976; Anantharamaiah 1985b, 1986; Heiles et al. 1996b; Roshi & Anantharamaiah 2000; Alves et al. 2015). Firstly, it is possible this gas may be dominated by spontaneous emission from ionized gas that has a lower (electron) density than a typical H II region and would be characterized by emission measures of around $EM \lesssim 10^6$ pc cm^{-6}; in this case, non-LTE effects need to be considered. Secondly, the RRL emission may be dominated by stimulated effects — thus largely characterized by the properties of low-frequency RRLs as described in Section 1.3.2 — and would arise from gas of lower density than in the spontaneous case (i.e., $n_e \lesssim 10$ cm^{-3}). Thirdly, spontaneous and stimulated effects may both influence the line intensity. Without sampling the line properties across frequency and comparing it with continuum emission at a high enough resolution, it is difficult to distinguish between these cases (e.g. Shaver 1976). While RRLs have been detected with stimulated effects (Pedlar et al. 1978; Anantharamaiah 1985a; Heiles et al. 1996a; Roshi & Anantharamaiah 2000), in regions separated from ionizing sources (Lockman 1976; Heiles et al. 1996b), and at low frequencies (Shaver 1976; Anantharamaiah 1985a; Roshi & Anantharamaiah 2001), pinning down the origin and properties of gas probed in this frequency regime remains open.

1.3.4 Extragalactic observations

RRLs detected in external galaxies (beyond the Magellanic Clouds) have primarily been observed at frequencies above 1 GHz from classic H II-emitting regions in star-bursting regions, accounting for 21 galaxies detected to date[4]. We refer the reader to

[4]To the best of our knowledge, detections of RRLs above 1 GHz have been reported in the following extragalactic objects: M 82 (Shaver et al. 1977; Chaisson & Rodriguez 1977; Bell & Seaquist 1977, 1978; Shaver et al. 1978; Bell et al. 1984; Seaquist et al. 1985; Puxley et al. 1989; Roelfsema & Goss 1992; Seaquist et al. 1994, 1996; Rodriguez-Rico et al. 2004), NGC 253 (Seaquist & Bell 1977; Mebold et al. 1980; Anantharamaiah & Goss 1990, 1996; Puxley et al. 1997; Mohan et al. 2002, 2005; Rodriguez-Rico et al. 2006; Kepley et al. 2011; Bendo et al. 2015; Eisner et al. 2019), MRK 668 (OQ 208; Bell & Seaquist 1980; Bell et al. 1984), NGC 2146 (Puxley et al. 1991; Zhao et al. 1996), NGC 3628 (Anantharamaiah et al. 1993; Zhao et al. 1997), IC 694 (merging galaxy in ARP 299; Anantharamaiah et al. 1993; Zhao et al. 1997), NGC 1365 (Anantharamaiah et al. 1993), NGC 3690 (merging galaxy in ARP 299; Anantharamaiah et al. 1995; Zhao et al. 1997), ARP 220

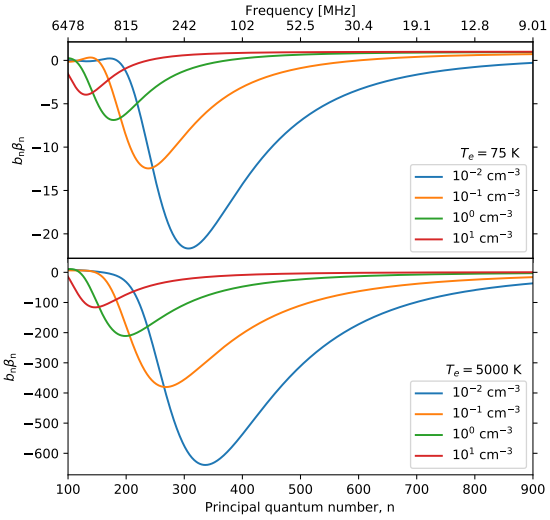


Figure 1.8: The strong dependence of the departure and stimulation coefficients ($b_n\beta_n$) from hydrogen RRL emission on the electron density of cold, partially ionized hydrogen (top, with $T_e = 75$ K) and warm, low-density ionized gas (bottom, with $T_e = 5000$ K).

Gordon & Sorochenko (2002) for a thorough discussion laying out the early advancements of probing and establishing the first extragalactic detections. Briefly, the first detections were established in M 82 by Shaver et al. (1977) and NGC 253 by Seaquist & Bell (1977), and only in the early 1990s were other galaxies detected, in large part due to renewed interest with the Very Large Array (e.g., Anantharamaiah et al. 1993). Through these and subsequent observations, star formation rates, ionized gas temperatures and densities, properties of super star clusters, and stellar kinematics have been established in the central starburst regions of nearby galaxies. Because the lines are relatively faint and trace starbursting regions (at galaxy centers), searches to detect high-frequency RRLs have not always been fruitful i.e., at low resolution, in the presence of a (Seyfert) AGN, and at frequencies below which (edge-on) starburst regions become optically thick to thermal continuum emission (e.g., Bell & Seaquist 1978; Bell et al. 1984; Anantharamaiah et al. 1995; Zhao et al. 1996; Phookun et al. 1998; Roy et al. 2008; Izumi et al. 2016; Luisi et al. 2018; Eisner et al. 2019).

Some of the extragalactic RRL observations at frequencies of about 1 – 10 GHz give support for line emission which is enhanced compared to LTE values as inferred from a comparison with thermal continuum components. In some sources (and at

(Anantharamaiah et al. 1995; Zhao et al. 1996; Anantharamaiah et al. 2000; Rodriguez-Rico et al. 2005), M 83 (Anantharamaiah et al. 1995; Zhao et al. 1996), NGC 4945 (Anantharamaiah et al. 1995; Roy et al. 2010; Bendo et al. 2016; Emig et al. 2020b), Circinus (ESO 97-G13; Anantharamaiah et al. 1995; Roy et al. 2008), NGC 660 (Phookun et al. 1998), NGC 5253 (Mohan et al. 2001; Rodriguez-Rico et al. 2007; Bendo et al. 2017), Henize 2-10 (Mohan et al. 2001), M 33 (toward NGC 604 Araya et al. 2004), NGC 3256 (Roy et al. 2005; Michiyama et al. 2020), NGC 1808 (Roy et al. 2008), IC 342 (Balser et al. 2017), M 51 and NGC 628 (Luisi et al. 2018).

particular frequencies) excess LTE values result once the free-free continuum starts to become optically thick and stimulates the line emission in a narrow frequency range (e.g., Anantharamaiah et al. 1993; Seaquist et al. 1996; Zhao et al. 1996, 1997; Phookun et al. 1998; Mohan et al. 2002, 2005; Rodriguez-Rico et al. 2006; Balser et al. 2017). Also, in a number of sources, a(n additional) low-density component which is stimulated by non-thermal background radiation is indicated by modeling the line and continuum emission. For example, the very first extragalactic detections in M 82 indicated stimulated emission was a major factor in the line intensities and originated from gas in front of a background non-thermal source (Shaver et al. 1977; Bell & Seaquist 1977; Shaver et al. 1978; Bell & Seaquist 1978; Roelfsema & Goss 1992; Rodriguez-Rico et al. 2004). Modeling the RRL emission from ∼100 GHz down to 1 GHz in NGC 253 and ARP 220, Mohan et al. (2005) and Anantharamaiah et al. (2000), respectively, showed that stimulation of a lower-density ionized component likely dominates emission below about 4 GHz (see also Mebold et al. 1980; Anantharamaiah & Goss 1990; Zhao et al. 1996; Mohan et al. 2002; Eisner et al. 2019). NGC 4945 is the brightest RRL emitter at and below 8 GHz, detected at 3″ with the Australia Telescope Compact Array (Roy et al. 2010); stimulated effects are possibly present. Detections of RRLs in the Circinus galaxy and NGC 1808 with the same observing set up as NGC 4945, find even stronger contributions of stimulated emission (≥90%; Roy et al. 2008). Early RRL observations of MRK 668 (OQ 208; Bell & Seaquist 1980; Bell et al. 1984) at 2.6′ (4.3′) resolution at 10.5 (6.1) GHz showed that stimulated RRL emission is detectable — with a peak line to continuum ratio of 0.3 (0.5) % — in a GHz-peaked radio source at a redshift of $z = 0.0763 \pm 0.0002$.

The RRL results coming from extragalactic observations in the first few years indicated that stimulated emission is important (Shaver et al. 1977; Bell & Seaquist 1977; Shaver et al. 1978; Bell & Seaquist 1978; Mebold et al. 1980; Bell & Seaquist 1980) and theory was formalized (Shaver 1976, 1978). The implication that RRLs could be observed in radio-bright high-z sources was realized. However, two additional (single dish) searches to deep sensitivities but with the narrow bandwidths at the time, totaling 33 sources in all, were not promising (Churchwell & Shaver 1979; Bell et al. 1984). Using the 305 m Arecibo Observatory, (Churchwell & Shaver 1979) targeted radio sources with redshifts in the range $z = (0.002 - 3.530)$ and which have at least one of the following qualities: (a) a low-frequency turnover in the SED, (b) low polarization, (c) low-frequency variability, (d) a fairly strong continuum, (e) a large optical emission-line region, (f) low electron density of ionized gas indicated by optical forbidden lines, (g) narrow optical lines, since the bandpass fidelity limits line detections to widths of ≤400 km s^{-1}, and (h) infrared emission. Observations of 21 sources at 1.4 GHz with a 10 MHz bandwidth receiver resulted in no detections[5] with a typical (median) 5σ limit to the peak line flux of 10 mJy and a peak line to continuum ratio of ∼ 0.1%. The observations had a 4.2 km s^{-1} channel width, but the upper limits consider a 400 km s^{-1} line width. In the same study, three sources were observed — 4C 28.25, 3C 270.1, and 3C 298 — at 430 MHz with a 2.5 MHz bandwidth receiver targeting carbon RRLs in intervening absorption systems. Upper limits for a 10 km s^{-1} line width are a peak line flux density of 7 – 18 mJy and a

[5]A tentative detection was found in the local source NGC 4631 but has not been followed up or confirmed since.

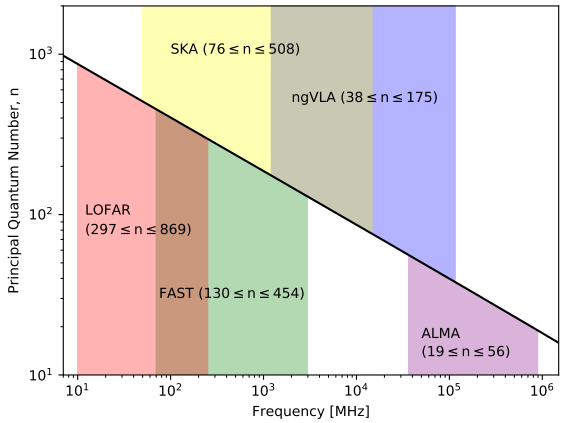


Figure 1.9: The principal quantum number, n, of the α-transitions ($\Delta\mathsf{n} = 1$) of radio recombination lines plotted as a function of frequency. We highlight the (planned) frequency ranges of telescopes that are currently operable or will be in the foreseeable future and that provide among the best sensitivity and resolution to characterize cold and warm diffuse gas and dense ionized gas in extragalactic sources through RRL observations.

peak line to continuum ratio of $(0.09 - 0.4)\%$. The other high-z search for RRLs was reported by Bell et al. (1984) using the Effelsberg 100 m dish and the 4.8 GHz receiver (thus $2.4'$ resolution) with two 40 MHz bandwidths. No ambiguous detections were reported in 9 sources which have redshifts ranging from $z = (0.0048 - 2.365)$. The 3σ upper limit to the peak line flux densities are $5 - 20$ mJy with a channel width of 20 km s^{-1}. Searches did not continue, likely because of the limited hardware at the time — inadequate resolution at low-frequencies (<1 GHz and especially below 500 MHz) and narrow bandwidth receivers.

1.4 Telescopes

The LOw Frequency ARray (LOFAR) and the Atacama Large Millimeter/submillimeter Array (ALMA) are two aperture synthesis telescopes which recently began operation, accessing portions of the electromagnetic spectrum with the sensitivity and resolution not previously capable. In this section, we briefly introduce these two telescopes. We discuss the new capabilities to observe radio recombination lines from these telescopes, thanks to more than 800 lines (per each element) populating frequencies below 950 GHz (see Figure 1.9). We also briefly introduce the advantages to studying thermal radio emission that these telescopes bring.

1.4.1 LOFAR

The LOw Frequency ARray (LOFAR; van Haarlem et al. 2013) is a new generation, digitally-intensive radio interferometer operating in the frequency range 10 – 250

Figure 1.10: *Top:* The yellow-filled circles mark the locations of LOFAR stations throughout Europe. LOFAR stations are centralized in the Netherlands and spread throughout nine European countries – the Netherlands, Germany, Poland, United Kingdom, France, Sweden, Ireland, Latvia, and Italy. *Bottom:* A zoom-in on LOFAR stations. The LOFAR "superterp" is the central circular patch of land containing multiple LOFAR stations that affords excellent short-spacing u,v coverage. Portions of three additional core stations are also seen in the peripheries of the image. High Band Antenna (HBA) stations are identified by consecutive segments of dark tiles. In the remote and international stations of LOFAR, a station consists of one consecutive grouping of HBA tiles; however, for the core stations (and as shown in this image), an HBA station is made up of two "ears" or two groupings of HBA tiles. Low Band Antenna (LBA) stations are shown in the image as dense groups of individual antenna elements interspersed within the HBA stations.

MHz. The array consists of a large number of simple, inexpensive dipole antennas that are not steerable, but fixed in place. Two types of antennas provide coverage in two bands: the low band antennas (LBA) at 10 – 90 MHz and high band antennas (HBA) at 110 – 250 MHz. LOFAR is the first telescope of its kind in the Northern Hemisphere and will uniquely remain so for the foreseeable future.

The heart of LOFAR is located in the Netherlands — as the telescope was designed and built by the Netherlands Institute for Radio Astronomy (ASTRON) — with stations currently spread across nine European countries (see Figure 1.10). The array is configured with 24 core stations (CS) within a radius of 2 km, and 6 of those stations reside on raised land 320 m diameter referred to as the "superterp" (see Figure 1.10). The 14 remaining stations of the full Dutch array, referred to as the remote stations (RS), extend out to a radius of 90 km. The international stations of LOFAR have baselines of more than 1800 km.

LOFAR is an extremely flexible telescope, offering multiple observing modes (beam-formed, interferometric) and vast ranges of spectral, timing and spatial resolutions. Through observing the low-frequency sky with unprecedented resolution, depth, and fidelity, the key science areas where LOFAR is making an impact include: H I 21 cm observations of the epoch of reionization; northern sky continuum surveys of AGN, star-forming galaxies, and galaxy clusters; pulsars; ultra high energy cosmic rays; (Galactic) ISM studies of supernovae, low-density ionized gas, and RRLs; transient (stellar) sources; magnetism; and, solar physics and space weather.

In particular, LOFAR is allowing for the exploration of RRLs in the Galaxy and in extragalactic sources by meeting the following technical requirements: 1) sensitivity to frequencies (principal quantum numbers) where stimulated (enhanced) emission is prevalent, 2) large fractional bandwidths that observe 50 to 100s of RRLs simultaneously, enabling changes in line properties to be identified and/or more sensitive searches through line stacking, 3) adequate and flexible spectral resolutions for Galactic ($\sim$1 km s^{-1}) and extragalactic ($\sim$10 km s^{-1}) observations, and 4) spatial coverage and resolution to large clouds while also resolving the $\lesssim$1–100 pc emitting regions. These requirements have inhibited wide-spread, in-depth studies of low-frequency RRLs in the past, largely due to the low spatial resolutions and the narrow bandwidths of traditional low-frequency instruments — owing to the difficulty of calibrating low frequency observations affected by the ionosphere. Using LOFAR, the unprecedented science of RRLs is being demonstrated (Asgekar et al. 2013; Oonk et al. 2014; Morabito et al. 2014; Oonk et al. 2017; Salas et al. 2017, 2018, 2019; Emig et al. 2019, 2020a).

Continuum studies of thermal radio emission have uncovered and characterized phases of warm, low-density ionized gas in the past (e.g., Westerhout 1958; Hoyle & Ellis 1963). Low-frequency observations provide a crucial probe of the turnover frequency of low-(electron) density ionized gas where the free-free emission becomes optically thick. LOFAR provides the unique resolution to resolve differences in cloud properties in order to investigate influences of energy and momentum injected by stellar feedback in warm ionized gas, the transport and leakage of ionization radiation, and turbulence of ionized gas.

Figure 1.11: Antennas of the Atacama Large Millimeter/submillimeter Array (ALMA). Fifty of these antennas have 12 m diameters and make up the "12 m array", used for sensitive, high-resolution imaging. Green circles in this image mark four additional 12 m antennas that are used for total power observation (the "TP array"), and interior to those are twelve closely spaced 7 m antennas (the "7 m array"). The 7 m and TP arrays enhance wide-field imaging of extended structures and are collectively referred to as the Atacama Compact Array.

1.4.2 ALMA

The Atacama Large Millimeter/submillimeter Array (ALMA)[6,7] currently operates in the frequency range 84 – 950 GHz (3.6 – 0.32 mm). The array is located on the Chajnantor plain of the Chilean Andes (lat.= $-23.02917°$, long.= $-67.754649°$), a region with exceptionally dry and clear sky conditions. The array consists of 66 antennas that can be positioned in different configurations over 192 antenna foundations with baselines as large as 16 km. A portion of the antennas are shown in Figure 1.11. Since first light in 2010 and becoming fully operational in 2013, ALMA has had a revolutionary impact on many fields of astrophysical research, notably star formation – from resolving individual stellar systems to galaxy scale properties and at the highest redshifts. The high-sensitivity and high-resolution capabilities have enabled the plethora of atomic and molecular species in this wavelength range to be explored together with dust and free-free continua.

In particular, ALMA affords a breakthrough in studying the most massive (and densest) forming stellar clusters. Direct optical and even near-infrared observations of forming clusters are complicated by large amounts of extinction present in the recently formed, largely embedded regions. Analyses of long wavelength free-free emission and hydrogen recombination lines of star clusters offer an alternative, extinction-

[6] `https://www.almaobservatory.org`

[7] A comprehensive description of the ALMA observatory and its components can be found in the ALMA Technical Handbook: `https://almascience.org/documents-and-tools/cycle8/alma-technical-handbook`

free probe of the ionizing gas (and therefore ionizing radiation and stellar mass) surrounding young star clusters (Condon 1992; Roelfsema & Goss 1992; Murphy et al. 2018). However, achieving a spatial resolution matched to the size of young clusters $\mathcal{O}(1$ pc) (Ryon et al. 2017) in galaxies at the necessary frequencies and sensitivities has only recently become possible thanks to the ALMA. Together with molecular species and dust continuum observations, a comprehensive picture can be assessed towards understanding their formation, identifying the dominant feedback processes at each stage of cluster evolution, determining which clusters survive as gravitationally bound objects, and linking all of these processes to the galactic environment (Turner et al. 2017; Oey et al. 2017; Leroy et al. 2018; Rico-Villas et al. 2020; Krieger et al. 2020; Emig et al. 2020b; Levy et al. 2021; Villas et al. 2020).

On larger (unresolved) scales, observations of RRLs with ALMA are providing tracers of total star formation rates and characterizing physical properties of ionized gas. At low ALMA frequencies affects due to stimulated emission which may be present at traditionally employed radio frequencies, 1–10 GHz, do not complicate line observations and the continuum is less contaminated by synchrotron emission. Sources including NGC 253 (Bendo et al. 2015, Mills et al., in preparation), NGC 3256 (Michiyama et al. 2020), NGC 4945 (Bendo et al. 2016; Emig et al. 2020b), and NGC 5253 (Bendo et al. 2017) have been investigated through RRLs.

1.5 In this thesis

This thesis is focused on investigating the interstellar medium and the influence of massive star formation on global (galactic) scales to sub-cloud pc scales. We focus on the influence of massive stars through the warm gas which they ionize in two different environments: a massive excavated region of relatively low surface mass density star formation (Chapter 2), and a massive clustered mode of efficient star formation (Chapter 3). We explore new methods of investigating the interstellar medium through RRL observations, resulting in the first detection at cosmological distances (Chapter 4 and 5).

In **Chapter 2**, we investigate the influence of massive stars on their environment by tracing low-density photoionized in the Cygnus X region. The Cygnus X region ($d \sim 1.5$ kpc) is one of the most massive star forming complexes in our Galaxy, in which the Cyg OB2 association (age of 3-5 Myr and stellar mass 2×10^4 $M_\odot$) has a dominant influence. We observe the Cygnus X region at 142 MHz using the LOFAR and correct for missing short-spacing information during image deconvolution. Together with data from the Canadian Galactic Plane Survey, we investigate the morphology, distribution, and physical conditions of low-density ionized gas in a $4^\circ \times 4^\circ$ ($\sim$100 pc $\times$ 100 pc) region at a resolution of $2'$ (0.9 pc). The radio emission in the region analyzed is almost entirely thermal (free-free) at 142 MHz, with emission measures (EM) of $10^3 < EM$ [pc cm^{-6}] $< 10^6$. As filamentary structure is a prominent feature of the emission, we use `DisPerSE` and `FilChaP` to characterize their radial profile distributions. The distribution of profiles has a characteristic filament width of 3.6 pc and a power-law distribution ($\beta = -1.8 \pm 0.1$) in peak EM down to our completeness limit of 4200 pc cm^{-6}. The electron densities in the filaments range

between $10 \lesssim n_e$ [cm^{-3}] $\lesssim 400$ with a median value of 38 cm^{-3}, remarkably similar to [N II] surveys of ionized gas. Cyg OB2 may ionize up to two-thirds of the total ionized gas and the ionized gas in filaments. Indeed we find the majority of filaments are likely photoevaporating surfaces flowing into a surrounding diffuse ($\sim$5 cm^{-3}) medium. However, stellar winds of Cyg OB2 may create a minority of the ionized filaments as dissipated turbulence or shock heated ionized gas. We discuss Cygnus X as a source of extended low-density (ELD) ionized gas, replenished over the lifetime of Cyg OB2 by ten of the typical filaments we characterize.

In **Chapter 3**, we turn to NGC 4945, a nearby (3.8 Mpc) galaxy hosting a nuclear starburst and Seyfert Type 2 AGN. We use the Atacama Large Millimeter/submillimeter Array (ALMA) to image the 93 GHz (3.2 mm) free-free continuum and hydrogen recombination line emission (H40α and H42α) at 2.2 pc (0.12″) resolution. Our observations reveal 27 bright, compact sources with FWHM sizes of 1.4–4.0 pc, which we identify as candidate super star clusters. Recombination line emission, tracing the ionizing photon rate of the candidate clusters, is detected in 15 sources, 6 of which have a significant synchrotron component to the 93 GHz continuum. Adopting an age of $\sim$5 Myr, the stellar masses implied by the ionizing photon luminosities are $\log_{10}(M_\star/\mathrm{M}_\odot) \approx 4.7$–6.1. We fit a slope to the cluster mass distribution and find $\beta = -1.8 \pm 0.4$. The gas masses associated with these clusters, derived from the dust continuum at 350 GHz, are typically an order of magnitude lower than the stellar mass. These candidate clusters appear to have already converted a large fraction of their dense natal material into stars and, given their small free-fall times of $\sim$0.05 Myr, are surviving an early volatile phase. We identify a point-like source in 93 GHz continuum emission which is presumed to be the AGN. We do not detect recombination line emission from the AGN and place an upper limit on the ionizing photons which leak into the starburst region of $Q_0 < 10^{52}$ s^{-1}.

Chapter 4 reports the first detection of RRLs at cosmological distances with observations of the radio quasar 3C 190 ($z = 1.1946$. 3C 190 was observed with the HBA of the LOFAR and processed using newly developed techniques for spectral analysis. We report the detection of RRLs in the frequency range 112 MHz – 163 MHz in the spectrum of 3C 190. Stacking 13 α-transitions with principal quantum numbers $\mathsf{n} = 266 - 301$, a peak 6σ feature of optical depth $\tau_{\mathrm{peak}} = (1.0 \pm 0.2) \times 10^{-3}$ and FWHM $= 31.2 \pm 8.3$ km s^{-1} was found at $z = 1.124$. This corresponds to a velocity offset of -9965 km s^{-1} with respect to the systemic redshift of 3C 190. We consider three interpretations of the origin of the RRL emission: an intervening dwarf-like galaxy, an active galactic nucleus-driven (AGN) outflow, and the inter-galactic medium. We argue that the recombination lines most likely originate in a dwarf-like galaxy ($M \sim 10^9$ M$_\odot$) along the line of sight, although we cannot rule out an AGN-driven outflow. We do find the RRLs to be inconsistent with an inter-galactic medium origin. With this detection, we have opened up a new way to study the physical properties of cool, diffuse gas out to cosmological distances.

In **Chapter 5**, we lay out the technical methods of searching for faint RRLs, both in processing and preparing new spectral line observations with LOFAR and how a spectrum can be searched in redshift space for RRLs. RRLs have been largely unexplored outside of our Galaxy. Next-generation low-frequency interferometers such as LOFAR, MWA, and the future SKA will, with unprecedented sensitivity,

resolution, and large fractional bandwidths, enable the exploration of the extragalactic RRL universe. We describe methods used to (1) process LOFAR high band antenna (HBA) observations for RRL analysis, and (2) search spectra for RRLs blindly in redshift space. We observed the radio quasar 3C 190 ($z \approx 1.2$) with the LOFAR HBA. In reducing these data for spectroscopic analysis, we placed special emphasis on bandpass calibration. We devised cross-correlation techniques that utilize the unique frequency spacing between RRLs to significantly identify RRLs in a low-frequency spectrum. We demonstrate the utility of this method by applying it to existing low-frequency spectra of Cassiopeia A and M 82, and to the new observations of 3C 190. Radio recombination lines have been detected in the foreground of 3C 190 at $z = 1.12355$ (assuming a carbon origin) owing to the first detection of RRLs outside of the local universe (first reported in Emig et al. 2019). Toward the Galactic supernova remnant Cassiopeia A, we uncover three new detections: (1) stimulated Cϵ transitions ($\Delta \mathsf{n} = 5$) for the first time at low radio frequencies, (2) Hα transitions at 64 MHz with a full width at half-maximum of 3.1 km s^{-1}, the most narrow and one of the lowest frequency detections of hydrogen to date, and (3) Cα at $v_{\rm LSR} \approx 0$ km s^{-1} in the frequency range 55–78 MHz for the first time. Additionally, we recover Cα, Cβ, Cγ, and Cδ from the -47 km s^{-1} and -38 km s^{-1} components. In the nearby starburst galaxy M 82, we do not find a significant feature. With previously used techniques, we reproduce the previously reported line properties. RRLs have been blindly searched and successfully identified in Galactic (to high-order transitions) and extragalactic (to high redshift) observations with our spectral searching method. Our current searches for RRLs in LOFAR observations are limited to narrow (< 100 km s^{-1}) features, owing to the relatively small number of channels available for continuum estimation. Future strategies making use of a wider band (covering multiple LOFAR subbands) or designs with larger contiguous frequency chunks would aid calibration to deeper sensitivities and broader features.

1.6 Future Outlook

A major take away from this thesis is that the physical conditions in the diffuse ISM can be explored through observations of RRLs in and/or against radio bright galaxies out to high redshift. This opens the door to investigations in large populations of external galaxies of cold, diffuse gas and/or of diffuse ionized gas using existing facilities like the LOFAR, VLA, GMRT, WSRT, MeerKAT, and ASKAP. In Figure 1.12, we plot the (brightest) galaxies at 178 MHz observed with the 3C(R) catalog that will enable the deepest searches at low frequencies. In the immediate future, it is possible to follow up the 3C 190 analysis with RRL observations at lower frequencies (i.e., with LOFAR's LBA) and higher frequencies (P-band, 300–500 MHz) to confirm the detection and constrain physical properties of the gas. It is also clear that M 82 needs to be re-observed to significantly deeper sensitivities (50 hours) and at higher spectral resolution using LOFAR.

In the future, the Square Kilometer Array (SKA)[8] can revolutionize (extragalactic) radio recombination lines (Peters et al. 2011; Oonk et al. 2015; Manti et al. 2016).

[8] `https://www.skatelescope.org`

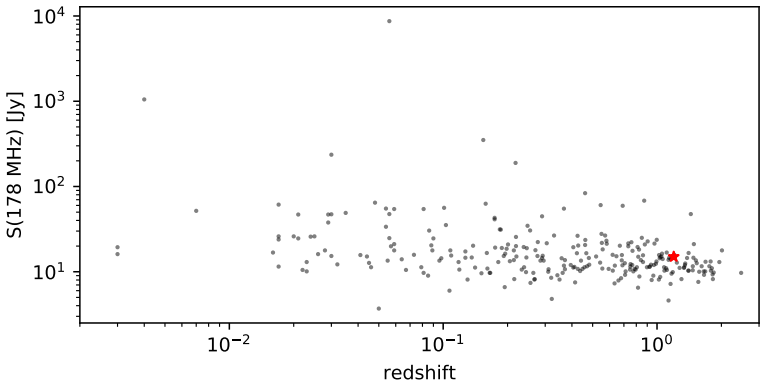


Figure 1.12: 278 (of the 298) sources of the revised third Cambridge catalog (3CR) — which includes radio sources in the Northern Hemisphere with flux densities of $S(178\ \mathrm{MHz}) \geq 9$ Jy — with known redshifts as reported in Spinrad et al. (1985). The 3CR catalog consists of the brightest sources in the low-frequency sky that will enable the deepest searches of RRLs from the diffuse ISM with current telescopes. In the future, SKA will enable deep searches in more than 10 000 galaxies. 3C 190, the source from which an RRL detection was reported by Emig et al. (2019), is shown with a red star at $z = 1.2$.

It will be a radio interferometer with the world's largest collecting area and fastest survey speed (expected operational for the community by 2028). With two different arrays, the SKA1-low ($50 - 350$ MHz) located in Australia will have a maximum baseline of 65 km ($6''$ resolution at 200 MHz) and deliver a sensitivity approximately eight times deeper than that of LOFAR, and the SKA1-mid ($0.35 - 15$ GHz) will be operated from South Africa (current MeerKAT site) with a maximum baseline of 150 km ($0.1''$ resolution at 5 GHz) with a sensitivity about six times deeper than the VLA. Additional updates (planned as SKA2) would add another order of magnitude in sensitivity. With LOFAR, clouds with column densities of 3×10^{20} cm^{-2} on scales of $>10'$ can be probed. With SKA, we can detect clouds with $N(\mathrm{H}) > 5 \times 10^{19}$ cm^{-2} on scales $>3'$. Or to put this in a different perspective, with LOFAR we can study absorption towards the sources in the 3C catalog while with SKA, all sources in the 7C catalog come into play (Oonk et al. 2015).

NGC 4945 and NGC 253 are the first galaxies in which super star clusters are being uncovered and resolved with ALMA (Leroy et al. 2018; Emig et al. 2020b) and being directly probed with RRLs using ALMA (Emig et al. 2020b, Mills et al., in preparation). ALMA has great potential to uncover many more SSCs and constrain (i) the fraction of star formation that occurs in this clustered mode, (ii) the timescales for the onset and completion of cluster formation, (iii) the net efficiency (fraction of gas converted to stars) and efficiency per free-fall time, (iv) the dominant feedback and support mechanisms for protoclusters as a function of time (Levy et al. 2021), and (v) the mechanism that ends cluster formation. A significant fraction of galaxies with existing RRL detections at lower resolution or lower frequency seem to harbor SSCs; populations of young SSCs likely remain to be uncover from the currently remaining extragalactic sources with RRL detections.

2 | Low-Frequency Observations of Diffuse Ionized Gas in Cygnus X

Abstract

Photoionized gas probes the influence of massive stars on their environment. The Cygnus X region ($d \sim 1.5$ kpc) is one of the most massive star forming complexes in our Galaxy, in which the Cyg OB2 association (age of 3-5 Myr and stellar mass 2×10^4 $M_\odot$) has a dominant influence. We observe the Cygnus X region at 142 MHz using the Low Frequency Array (LOFAR) and take into account short-spacing information during image deconvolution. Together with data from the Canadian Galactic Plane Survey, we investigate the morphology, distribution, and physical conditions of low-density ionized gas in a $4^\circ \times 4^\circ$ ($\sim$100 pc $\times$ 100 pc) region at a resolution of $2'$ (0.9 pc). The radio emission in the region analyzed is almost entirely thermal (free-free) at 142 MHz, with emission measures (EM) of $10^3 < EM$ [pc cm^{-6}] $< 10^6$. As filamentary structure is a prominent feature of the emission, we use `DisPerSE` and `FilChaP` to identify filaments and characterize their radial (EM) profiles. The distribution of profiles has a characteristic filament width of 3.6 pc and a power-law distribution ($\beta = -1.8 \pm 0.1$) in peak EM down to our completeness limit of 4200 pc cm^{-6}. The electron densities in the filaments range between $10 \lesssim n_e$ [cm^{-3}] $\lesssim 400$ with a median value of 38 cm^{-3}, remarkably similar to [N II] surveys of ionized gas. Cyg OB2 may ionize up to two-thirds of the total ionized gas and the ionized gas in filaments. Indeed we find the majority of filaments are likely photoevaporating surfaces flowing into a surrounding diffuse ($\sim$5 cm^{-3}) medium. However, stellar winds of Cyg OB2 may create a minority of the ionized filaments through dissipated turbulence or shock heated ionized gas. We discuss Cygnus X as a source of "extended low-density" (ELD) ionized gas; this gas is not well confined to the region, but can be replenished over the lifetime of Cyg OB2 by ten of the typical photoevaporating filaments we characterize.

K. L. Emig, G. J. White, P. Salas, R. L. Karim, R. van Weeren, P. Teuben, H. J. A. Röttgering, A. G. G. M. Tielens, *et. al*
to be submitted to Astronomy & Astrophysics

2.1 Introduction

The interaction of massive stars with their environment has a profound impact on the evolution of galaxies (Hopkins et al. 2014, 2018) through the collective effects of protostellar outflows (Bally 2016), ionizing radiation (Matzner 2002), stellar winds and supernovae (SN; Yorke et al. 1989). One way to investigate their impact is through photoionized gas, in which stars of mass $M_\star \gtrsim 7$ M$_\odot$, all O types and earlier than B3, produce extreme ultraviolet (EUV) radiation of $E \geq 13.6$ eV capable of photoionizing hydrogen in the surrounding medium.

Early in their lifetimes, massive stars dissociate and ionize their immediate environment. Dense H II regions form as pockets of ionized gas, increasing the thermal gas pressure within the molecular cloud by three orders of magnitude. The subsequent expansion of an H II region can mechanically unbind the parent cloud and induce turbulent motions (e.g., Walch et al. 2012) in the interstellar medium (ISM). As an H II region grows, its volume density diminishes. Stellar winds and radiation pressure are also important contributors to the expansion of the ionized gas volume (e.g., Pabst et al. 2019, 2020; Olivier et al. 2020). As a result of peculiar motion and/or inhomogeneities in the medium, within a few Myr (e.g., Mezger 1978), the star, its photons and the gas it ionizes enter a surrounding low-density ($n_e \sim 0.1 - 100$ cm^{-3}) medium.

Within the plane of the Galaxy, photoionized gas is found in a variety of environments (and referred to with a variety of different names). Dense ($n_e > 10^3$ cm^{-3}) ionized gas pervades compact, ionization-bounded H II regions. From leaky H II regions, ionizing photons escaping through porous material create (partially) ionized gas ($1 - 10$ cm^{-3}) in the envelopes of H II regions; alternatively these envelopes may be ionized from the outside. Transitioning to density-bounded H II regions, ionized gas ($1 - 100$ cm^{-3}) permeates to larger volumes in blister H II regions aided by champagne flows (Tenorio-Tagle 1979) of ionized gas into lower-density regions. When stellar winds are influential in an H II region, the medium stratifies with ionized gas at larger radii and hot gas filling the inner regions (Weaver et al. 1977; Churchwell et al. 2006). Assisted by supernova explosions, massive stars create large excavated regions or plasma tunnels containing fully ionized gas ($1 - 10$ cm^{-3}). Over-pressured — thus denser, brighter, and more readily detected — photoevaporating ionized gas (of smaller path length) is frequently observed in a number of these scenarios as ionization fronts propagate into local neutral material. Ionizing photons which escape in these scenarios provide a surplus of the ionizing photon budget required to maintain (e.g., Reynolds 1984) the pervasive (volume filling factor $\phi \sim 0.25$, Kulkarni & Heiles 1988) warm ionized component of the ISM (WIM), that resides both in the Galactic plane ($n_e \sim 0.1$ cm^{-3}) and which is characterized by large ($z \sim 1$ kpc) scale heights (for a review see Haffner et al. 2009).

Blind surveys and large targeted samples with thermal radio continua (Mezger 1978; Murray & Rahman 2010), radio recombination line emission (Shaver 1976; Lockman 1976; Anantharamaiah 1985b, 1986; Roshi & Anantharamaiah 2000; Heiles et al. 1996b; Alves et al. 2015), FIR fine structure line emission from [N II] (Bennett et al. 1994; Goldsmith et al. 2015) and pulsar dispersion measures (Berkhuijsen et al. 2006, and references therein) have brought to light properties of low-density ionized gas

within the Galactic plane. Mezger (1978) estimated that 84% of ionizing photons are emitted by O stars outside of compact H II regions, in gas characterized by densities of $n_e \approx 5-10$ cm^{-3} dubbed extended low-density (ELD) H II gas. However, a number of different conclusions have been reached regarding the dominant origin of this gas: (i) envelopes of H II regions (Shaver 1976; Anantharamaiah 1986; McKee & Williams 1997), (ii) a pervasive component (Heiles et al. 1996b) which may be a continuation of the WIM (Bennett et al. 1994; Berkhuijsen et al. 2006), and (iii) from just a handful of the most luminous sources (Murray & Rahman 2010). High-resolution pinhole surveys of FIR fine structure lines (Goldsmith et al. 2015; Pineda et al. 2019) trace somewhat denser gas (with mean values of $n_e \approx 30 - 40$ cm^{-3}) and could plausibly be tied with these three possibilities.

The detailed nature of low-density gas has been revealed in individual nebulae, some of which show ridge-like and filament morphologies. For example, IC 434, the ionized gas associated with the Horsehead area in Orion, is a low-density (~ 100 cm^{-3}) ionization front created by an O star (σ Ori) drifting away (Ochsendorf et al. 2014). The California Nebula (NGC 1499) is low-density ionized gas created as a chance encounter with a run away star (Bohnenstengel & Wendker 1976). Barnard's Loop is low-density ($\sim$3 cm^{-3}) ionized gas swept up and transported away by a SN explosion within the Orion-Eridanus bubble and ionized by the Orion Belt stars whose ionizing photons can travel some 10's of pc inside the Orion-Eridanus bubble (Ochsendorf et al. 2015; Abdullah & Tielens 2020). The massive Arches star cluster, in a region dominated by stellar winds, maintains the prominent photoionized Arched Filaments (Lang et al. 2001).

The Cygnus X region is a massive complex which also displays filamentary-like structure in low-density ionized gas (Wendker et al. 1991). "Cygnus X" refers to a $\sim 10^\circ$ wide region in the galactic plane with enhanced radio emission (Piddington & Minnett 1952) – for an overview of the region see Reipurth & Schneider (2008). This coherent region of massive star formation appears to lie at $\sim$1.5 kpc (Schneider et al. 2006; Rygl et al. 2012) and dominates the observed emission, despite the view looking down a spiral arm in this direction. Open and massive OB associations are seen in this direction at a similar distance – Cyg OB1, OB2, OB6, OB9 (Uyaniker et al. 2001). Yet (a subset of) the region still retains a large reservoir ($M > 10^6$ M$_\odot$) of molecular gas (Schneider et al. 2006, 2011). Massive post main-sequence stars indicate that star formation began in the region $\sim$15 Myr ago (Comerón & Pasquali 2012; Comerón et al. 2016, 2020). Though the nature of the rarefied region is debated, it is consistent with the superbubble formalism (Mckee & Ostriker 1977) indicating that stellar winds and/or supernovae have contributed to excavating the parent cloud(s) (Bochkarev & Sitnik 1985; Ackermann et al. 2011).

Part of the region that is referred to as Cygnus X North (Schneider et al. 2006) harbors Cyg OB2 and other smaller clusters and associations, some of which are actively forming stars (Cong 1977; Odenwald et al. 1990; Comeron & Torra 1999; Comerón & Torra 2001; Le Duigou & Knodlseder 2002; Marston et al. 2004; Motte et al. 2007; Beerer et al. 2010; Panwar et al. 2020). The Cyg OB2 association strongly influences the medium, largely characterized through pillars and globules (Wright et al. 2012; Schneider et al. 2016; Deb et al. 2018). Cyg OB2 has a total stellar mass of $M_\star = 1.7^{+0.4}_{-0.3} \times 10^4$ M$_\odot$ (Wright et al. 2015). It is not bound gravitationally and

likely formed that way, in a relatively low density environment given its mass (Wright et al. 2014), during bursts of star formation 3 and 5 Myr ago (Wright et al. 2010; Berlanas et al. 2020).

In this article, we investigate low-density ionized gas in Cygnus X (North) with 142 MHz continuum observations using the Low Frequency Array (LOFAR; van Haarlem et al. 2013). We observe this region, encompassing 16 square degrees ($\sim$10 kpc^2) within a single pointing, around the Cyg OB2 association as it contains a wide variety of evolutionary stages of star-formation with low-density ionized gas. We aim to characterize the influence of massive stars on their environment and explore the connection between low-density ionized gas and ELD H II gas. This is the first analysis with LOFAR which focuses on (extended) thermal emission in the Galactic plane.

We adopt a distance of $d = 1.5 \pm 0.1$ kpc to the Cygnus X region and Cyg OB2 following (Comerón et al. 2020, see their Sec. 2.2 for a detailed discussion). After an analysis of molecular line emission linked the region as a coherent structure influenced by Cyg OB2 (Schneider et al. 2006, 2016), the distance to the Cygnus X complex was determined as 1.40 ± 0.08 kpc through maser observations of massive stars (Rygl et al. 2012). A recent *Gaia* analysis (Berlanas et al. 2019) places a main subgroup (80% of the OB population) of Cyg OB2 at $1.76^{+0.37}_{-0.26}$ kpc with a smaller subgroup at 1.35 ± 0.2 kpc. While a clear separation in distance distinguishes these two subgroups, Comerón et al. (2020) points out that a systematic offset in parallax may affect (both of) the distance determinations. At $d = 1.5$ kpc, the physical scale is $1' \approx 0.44$ pc.

2.2 Data

2.2.1 LOFAR observations & data processing

In this section we describe the data processing of LOFAR interferometric observations of one pointing centered approximately on the massive star forming region DR 21 at $(\alpha_{\rm RA}, \delta_{\rm dec}) = (309.5500°, +42.0708°)$ in J2000 coordinates. These data were obtained with the high band antennas (HBA) covering 110–190 MHz on July 19, 2013 under project LC0_032 (PI: G. White). Continuous frequency coverage was obtained between 126 – 165 MHz. 23 stations of the full Dutch array recorded data during these observations, which have a maximum baseline of 120 km ($4''$) and minimum baseline lengths of 70 m corresponding to largest angular scales of $96'$ (1.6°). The on-source integration time is 6 hours and 50 minutes.[1] In this analysis, we make use of 2 MHz bandwidth that is split across 10 sub-bands and centered on 142 MHz. We use a small fraction (5%) of the available bandwidth and with our results, inform on the prospects that deeper LOFAR observations hold.

We recorded data throughout the observation also in the direction of Cygnus A (299.8682°, 40.7339°), a bright ($\sim 10^4$ Jy), well-modeled source 7.4° away from the target center. The flexibility of digitally pointing the LOFAR beam allows HBA data to be recorded towards multiple phase centers within the station beam of $\sim$20°. Because of the close proximity and simultaneous observation, we use Cygnus A as the

[1] A plot of the *uv* coverage of a typical LOFAR observation can be found in e.g., Figure 17 of van Haarlem et al. (2013).

primary calibrator source to derive flux, bandpass and phase calibrations to transfer to the target pointing.

Calibrating the visibilities

We describe the calibration of the visibility data in this section. We derive flux, bandpass and phase calibrations of the primary calibrator pointing. We then apply these to the target pointing. Lastly, we mitigate side lobe contribution from Cygnus A to the target pointing by subtracting Cygnus A visibilities from the target visibilities; the trade-off to LOFAR's large field of view is its susceptibility to strong side lobe contamination.

To begin processing of both the primary calibrator and target Measurement Sets (MS), we flag the correlations between the two "ears" of each HBA core station (van Haarlem et al. 2013) and flag for radio frequency interference (RFI) with `AOFlagger` (Offringa et al. 2012) at full time and frequency resolution (0.763 kHz channel) using the default HBA strategy (Offringa et al. 2012). We also flag station CS013 as its antennas were not properly phased up at the time of the observations. Then we average (by a factor of 64) the data to a channel resolution of 48.8 kHz. We run these steps using version 2.20_2 of the Default Pipeline Processing Platform (`DPPP`; van Diepen & Dijkema 2018). This version of `DPPP` outputs the data with minor updates to the metadata of the Measurement Set – as these data were recorded in Cycle 0 of LOFAR's operation – that are necessary in order to use later versions of `DPPP`. In subsequent processing, we make use of the most recent `DPPP` versions.

For the primary calibrator, we solve for the diagonal (XX and YY) gain, phase and amplitude, at full frequency resolution with `DPPP`. We use a LOFAR model of Cygnus A at 150 MHz consisting of 33 000 components (courtesy J. McKean; McKean et al. 2016). The solutions have cleanly converged, and they also indicate the ionosphere was relatively mild. We apply only the phase solutions. Then, we solve again for the diagonal gain, phase and amplitude. Using the LOFAR Solutions Tool (LoSoTo; de Gasperin et al. 2019), we flag spurious jumps in the amplitude solutions of each channel – likely due to instrumental instabilities – as well as any residual RFI, which are identified as 7σ outliers averaged over a sliding window of 40 solution intervals in time.

For the target pointing, we use `DPPP` to apply the phase solutions of the first iteration of gain calibration of the primary calibrator and both the amplitude and phase solutions of the second iteration. The time steps of the visibility data which have flagged amplitude solutions will be flagged as the solutions are applied. Phase offsets due to asynchronized — on the order of $\mathcal{O}$(10 nanoseconds) — station clocks of the remote stations (van Weeren et al. 2016) have been absorbed into the primary calibrator solutions. Furthermore, a calm ionosphere may be characterized by large patches of coherence in the total electron content (Intema et al. 2009), thus inducing similar phase errors (and Faraday rotation) towards the target and nearby primary calibrator. Phase (and amplitude) errors due to the ionosphere are absorbed into these solutions as well.

In the next step, we use `DPPP` to subtract Cygnus A from the target visibilities because of the close proximity of Cygnus A to our pointing center. We phase shift

the target visibilities to the location of Cygnus A and then predict the 33 000 source model of Cygnus A into the MODEL column. Next we smooth our visibilities with a baseline dependent averager (de Gasperin et al. 2019) and solve for the diagonal (XX and YY) gain at a 195.3 kHz frequency resolution and 1 minute time resolution, attempting to smear the visibility structure of the target. We corrupt the MODEL column with the solutions. Then we subtract the corrupted MODEL from the DATA. We phase-shift these visibilities back to the target phase center of the observation. In this manner, we have subtracted the off-axis signal of Cygnus A from the data. We find this approach produces more coherent solutions with a larger number of stations converging than with the standard "demixing" (van der Tol et al. 2007) capabilities of `DPPP`.

At this point, the target data are well calibrated against a source which has a peak flux 1000 brighter than the apparent flux of any source in the target field. We tested self-calibrating the data, in which we image (with the parameters specified in Section 2.2.1), create a sky model, and solve for an amplitude and phase calibration against the model. We see minor improvements for long baselines. However, our science aim is to investigate extended diffuse emission that is well characterized with low-resolution ($2'$) imaging. Self-calibration does not significantly improve our short baseline calibration. We thus opted not to perform the computationally intensive self-calibration.

Constructing a short-spacing map

The LOFAR data have excellent *uv* coverage to large scale emission with baselines as short as 70 m providing sensitivity to a largest angular size of $96'$. However, for observations near the Galactic plane, where large scale emission surpasses tens of degrees, information on the smallest *uv* scales is necessary to properly deconvolve and obtain the flux density of the extended $\mathcal{O}(1^\circ)$ structures in the Cygnus X region.

We construct a short-spacing map at 142 MHz and with $72'$ resolution that is interpolated from multi-frequency fits across 52 MHz, 408 MHz, 820 MHz, and 1420 MHz. To ensure that *uv* coverage of the short-spacing map overlaps with the *uv* coverage of the LOFAR observations, we targeted survey data with resolutions of approximately $\sim 1^\circ$ or less and which covered galactic latitudes of $|b| \lesssim 7^\circ$.

We compiled the following data. We make use of 52.224 MHz survey data obtained with the Owens Valley Radio Observatory Long Wavelength Array (OVRA-LWA; Eastwood et al. 2018). It has a native resolution of $16.2' \times 15.0'$ and an rms noise of 418 K. Inspection of the data in this region shows that it is calibrated to within $\sim$20% as compared to lower resolution surveys. We use the 408 MHz survey of Haslam et al. (1982) that has been de-stripped (to <1 K) of large-scale striations (Remazeilles et al. 2015). At 408 MHz, the data have a native resolution of $51'$, rms noise of 3.2 K, and zero level calibration of 3 K (Haslam et al. 1974). The 820 MHz survey with the Dwingeloo Telescope (Berkhuijsen 1972) has a native resolution of $72'$, rms noise of 0.2 K and zero level calibration of 0.6 K. The 820 MHz data limits the resolution of our frequency-interpolated map as it has the coarsest resolution of the data used. Lastly, the data used from the Stockert 25 m survey at 1420 MHz (Reich 1982) has a native resolution of $35'$, rms noise of 0.01 K and zero level calibration of 0.05 K.

The procedure we employ first involves creating an image cutout of $12^\circ \times 12^\circ$ centered on DR 21 from each survey. Then we smooth the cutouts to the common resolution of $72'$ and re-grid the images using CASA (McMullin et al. 2007; Emonts et al. 2019) to a common pixel grid and pixel size of $4.7'$. Next we convert the intensity scales from temperature brightness units to Jy beam^{-1}, as the imager we use, `wsclean` (Offringa et al. 2014; Offringa & Smirnov 2017), currently only accepts Jy units. Pixel by pixel, we fit a power-law[2] to the "high-frequency" data points at 408, 820 and 1420 MHz, as previous studies have found this region largely consists of thermal, free-free emission down to 408 MHz (Landecker 1984; Xu et al. 2013). Then, to the 52 MHz data point, we fit for either a free-free turnover or a synchrotron component. We do this by extrapolating the high-frequency fit to 52 MHz. If the flux density at 52 MHz is *less* than the extrapolated value of the fit at that frequency, we fit for a free-free turnover,

$$S(\nu) = \frac{S_{0,\mathrm{ff}}}{\tau_0} \left(\frac{\nu}{\nu_0}\right)^2 \left(1 - \exp\left[-\tau_0 \left(\frac{\nu}{\nu_0}\right)^{\alpha-2}\right]\right) \tag{2.1}$$

where α is the spectral index fit to high-frequency data, and $\nu_0 = 142$ MHz such that $S_{0,\mathrm{ff}}$ is the free-free component of the flux density at ν_0 in the optically thin (unabsorbed) regime and τ_0, the only free parameter in this fit, is the optical depth at ν_0. Otherwise, if the flux density at 52 MHz is *greater* than the extrapolated value of the power-law fit, we add-in a synchrotron component with spectral index -0.7 and fit,

$$S(\nu) = S_{0,\mathrm{ff}} \left(\frac{\nu}{\nu_0}\right)^{\alpha} + S_{0,\mathrm{s}} \left(\frac{\nu}{\nu_0}\right)^{-0.7} \tag{2.2}$$

where $S_{0,s}$ is the synchrotron component of the flux density at ν_0 and the only free parameter in this fit. Examples of pixel locations fit with a free-free turnover and with a synchrotron component are shown in Figure 2.1. Although frequency-interpolation is performed at each pixel independently, point-to-point scatter is not introduced during this process. This is due to the coherence in intensity of neighboring pixels which have been smoothed over by a Gaussian beam that is 235 square pixels in area.

The short-spacing map frequency-interpolated to 142 MHz is shown in Figure 2.2,*center* for the region of sky falling within the 20% power of the LOFAR primary beam of the observation. We compare the frequency-interpolated short-spacing map at 142 MHz to an all-sky map at 150 MHz with 5° resolution (Landecker & Wielebinski 1970). The survey data in this region of the sky were originally observed at 178 MHz with $5^\circ \times 1.25^\circ$ resolution (Turtle & Baldwin 1962) and then scaled to 150 MHz with a spectral index[3] of $\beta = -2.6$. Since, the survey data have been re-calibrated to attain an 8 K zero level calibration and a scaling error of 1% (Patra et al. 2015). We take a cutout from the survey data which covers our region of interest and convert the intensity units from temperature brightness to Jy beam^{-1}. We convolve our short-spacing map to 5° resolution and re-grid it to match the 0.23° pixel scale. Taking the intensity ratio of the 150 MHz survey image over the convolved short-spacing map at

[2] $S(\nu) = S_0(\nu/\nu_0)^\alpha$
[3] $T_b \propto \nu^\beta$

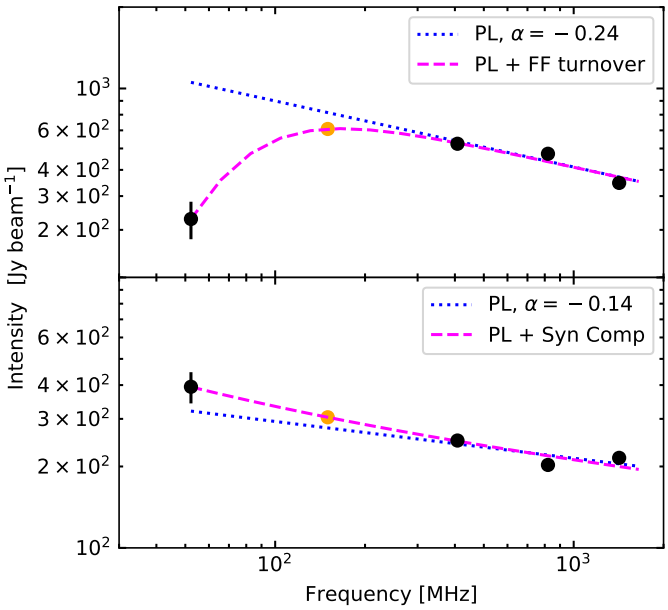


Figure 2.1: Examples of the multi-frequency fits (done on a pixel by pixel basis) to frequency-interpolate the "short-spacing" map – a low resolution image with sensitivity to large angular scales – at 142 MHz. The black data points represent the survey data used in the fit. The orange data point represents the interpolated value at 142 MHz. The dotted blue line represents the power-law fit to the high-frequency data points at 408, 820, and 1420 MHz. The dashed magenta line represents the final SED fit – where either a free-free turnover (*top*) is fit (Equation 2.1) or a synchrotron component (*bottom*) is added (Equation 2.2) to match the low-frequency point at 52 MHz.

142 MHz results in intensity ratios that have a median value of 1.25. This is within sensible agreement as a 25% error falls within the uncertainty of our survey data.

Imaging

The final step in processing the LOFAR observations is imaging the data. We use the frequency-interpolated short-spacing map (Section 2.2.1) as a starting model during imaging in order to deconvolve the interferometric data with zero-level and total flux density information.

To prepare a template for the short-spacing map for `wsclean` (Offringa et al. 2014), we first run `wsclean` with the image weight settings (see below) and 1 iteration so that the pixel grid and synthesized beam size are set. Next we convert the short-spacing map into intensity units of Jy pixel^{-1}according to the synthesized beam size of the `wsclean` image. This is done with the following routine. We re-grid the short-spacing map to match the `wsclean` output pixel grid and size of $18''$. Then we scale its intensity by the ratio of the beam areas, ($A_{\texttt{wsclean}}/A_{ss}$), to convert units to Jy beam^{-1}with respect to the `wsclean` output synthesized beam. Next we convert the intensity to units of Jy pixel^{-1} by dividing by the beam area in units of pixels. At this point the short-spacing model is prepared for `wsclean` in true-sky flux density units. `wsclean` requires that we input two model images: one which is attenuated by the primary beam and one which is not. Therefore, we create a copy of our short-spacing model image, but we also apply the primary beam response. We set the primary beam model to zero outside of the first null as we do not image beyond that. With these steps[4], the short-spacing map is prepared to be input to `wsclean`.

We run `wsclean` using the `-predict` option to predict the short-spacing model image into the MODEL column of the measurement set. We then run `wsclean` a second time, now with the `-continue` option to deconvolve the interferometric data. We image out to the first null of the primary beam over an area of $10^\circ \times 10^\circ$. We use Briggs image weighting with a robust parameter of -0.5 and a circular Gaussian taper of $1.5'$, effectively smoothing out the long baselines. We CLEAN with multi-scale cleaning (Offringa & Smirnov 2017) on pixel scales of [0,18,36,72,144] and to a threshold of 4σ where σ is scale dependent and internally calculated.

In Figure 2.2*right*, we show the output short-spacing corrected image. In this paper, we analyze a $4^\circ \times 4^\circ$ region covered by the LOFAR observations, which approximately encompasses the full-width half power of the primary beam. We smoothed the image to a circular beam of FWHM of $2'$. The final image has a noise of 25 mJy beam^{-1} and effective frequency of 142.19 MHz.

For comparison, we also image the LOFAR data without the short-spacing map but using the same imaging parameters and show the results in Figure 2.2*left*. Additionally, we checked our imaging results by convolving the final image to the $72'$ resolution of the initial short-spacing map and verified that the total integrated fluxes are consistent. Lastly, we compare our imaging results by "feathering" (Stanimirovic 2002; Cotton 2017) the short-spacing map with the LOFAR data imaged without

[4]We also verify our template preparation procedure with the CASA preparation procedure outlined by J. Kauffmann at `https://sites.google.com/site/jenskauffmann/research-notes/adding-zero-spa` .

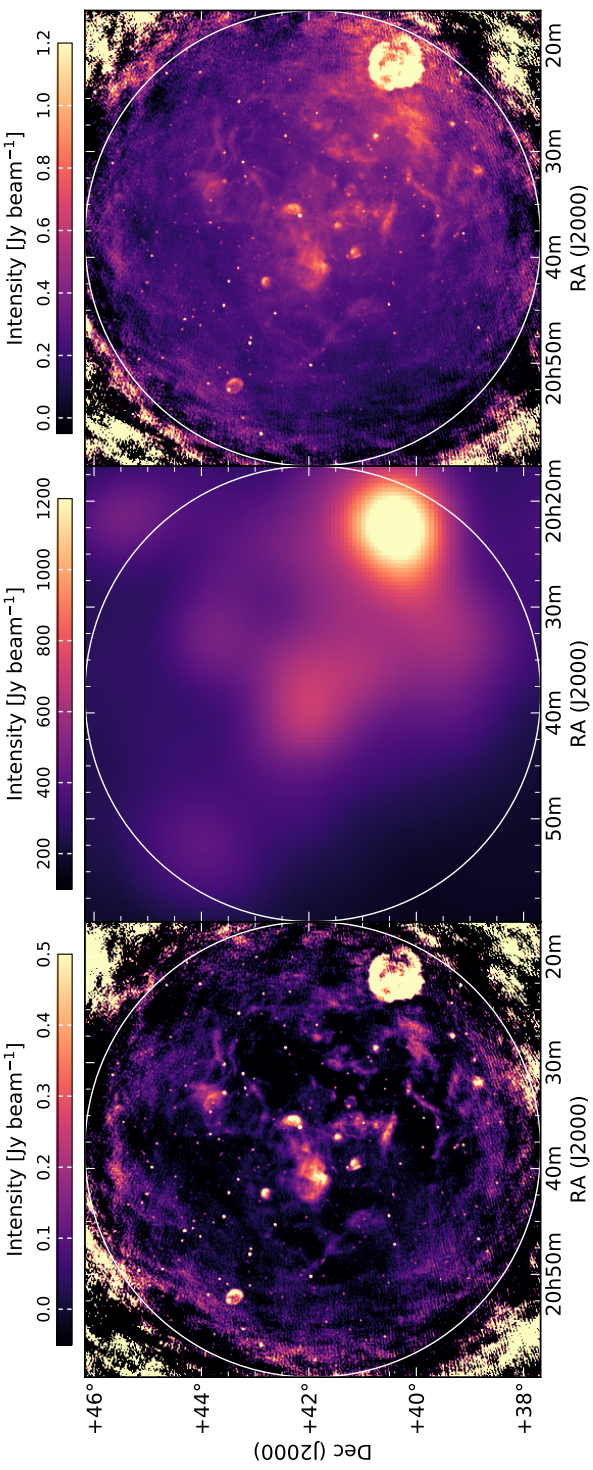


Figure 2.2: Imaging LOFAR 142 MHz observations of the Cygnus X region. The white circle shows the 20% power of the primary beam. We draw attention to the different intensity scales among the images. For reference, Cygnus A is located about 7.4°due west from the center of the pointing. *Left:* Deconvolved LOFAR image without short-spacing information. While these data are sensitive to emission on angular scales as large as 96′, negative (un-physical) emission is present induced by large scale emission in the Galactic plane. *Center:* The frequency-interpolated short-spacing map at 142 MHz (see Section 2.2.1) and 72′ resolution. *Right:* LOFAR data imaged with the short-spacing map as an initial model (see Section 2.2.1).

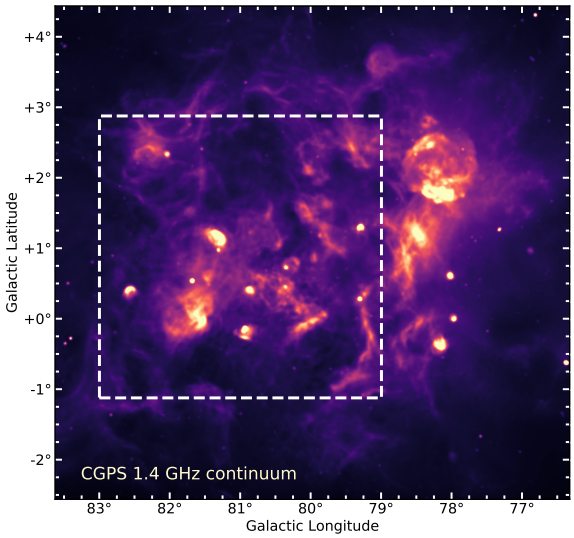


Figure 2.3: The full Cygnus X star-forming region shown in CGPS 1.4 GHz continuum intensity (Taylor et al. 2003). The white dashed lines encompass the region of interest (4° × 4°, or ∼100 pc × 100 pc) that we analyze with LOFAR 142 MHz observations.

short-spacings. Through feathering, images are combined in the domain of their Fourier transforms by a weighted average in order to extract the most appropriate spatial frequencies from each image. Taking the intensity ratio of the `CASAfeather` and the `wsclean` created images, we find a mean and median value of 1.06 computed over the 4° × 4° region.

2.2.2 Ancillary Data

We compare the LOFAR data with observations compiled through the Canadian Galactic Plane Survey (CGPS; Taylor et al. 2003) at 1420 MHz. These data were observed with the Synthesis Telescope at the Dominion Radio Astrophysical Observatory with short-spacing corrections, in this region, using the Stockert 25 m survey (Reich 1982). Since the effective resolution varies across the mosaic data products – the synthesized beam is declination dependent – we processed the mosaics to attain a common resolution of 2′. Then we stitch the smoothed mosaics together. The standard deviation in a relatively low emission region of the image is $\sigma = 0.03$ K (0.7 mJy beam^{-1}).

2.3 Continuum Emission

The Cygnus X star-forming complex encompasses an area of more than 50 square degrees as shown in Figure 2.3. The region of interest analyzed with our LOFAR pointing, 4° × 4° (100 pc × 100 pc) is depicted with white dashed lines.

Figure 2.4 shows the continuum emission observed at 142 MHz with LOFAR and which is corrected for missing short-baseline information (Section 2.2.1). The 142 MHz emission in this region primarily traces thermal free-free emission from ionized gas with low density and small emission measure (see Section 2.4 and e.g., Wendker et al. 1991). Thermal radiation referred to as "free-free" emission is bremsstrahlung radiation emitted as the paths of free electrons are deflected in the presence of free ions.

The morphology of the 142 MHz emission includes extended (on degrees scales), resolved regions of photoionized gas in the vicinity of massive stars and star clusters. Filaments are a prominent feature in bright and also faint emission regions. Shell-like regions also appear. An additional diffuse component surrounds much of the extended structures. Extra-galactic radio galaxies with bright synchrotron emission at these frequencies appear as point-like sources. Analyzing the spectral energy distribution (SED; see Section 2.4) of unresolved objects is necessary to pull-out synchrotron dominated sources from the regions of active star-formation embedded within a dense medium.

The locations of massive stars and star clusters are shown in Figure 2.5. We label the radio continuum features which were identified in Downes & Rinehart (1966) with 4.8 GHz observations – DR 7, 10, 11, 16, 18, 19, 20, 21, 22, 23. Only DR 7 is not associated with the Cygnus X region; instead it is located more than 3.3 kpc distant in the Perseus Arm. The Cygnus X region has been extensively surveyed for massive stars and clusters (Comerón & Pasquali 2012; Wright et al. 2015; Berlanas et al. 2018; Comerón et al. 2020). Confirmed OB stars within the Cyg OB2 association and in the field are marked by small green stars (Berlanas et al. 2018) and blue stars mark massive supergiants (Comerón et al. 2020). Large green stars mark candidate star clusters (Le Duigou & Knodlseder 2002).

The 1.4 GHz continuum emission from the CGPS, also shown in Figure 2.5, is strikingly similar to the LOFAR 142 MHz continuum emission. The morphologies and relative intensities of the extended, resolved emission (including filaments) are comparable. Even regions which are devoid of emission seem to agree. The low-level, diffuse emission surrounding extended structures is relatively brighter at lower frequencies. Similarities attest to the primarily thermal nature of the emission (Xu et al. 2013; Wendker et al. 1991). Differences in relative intensities arise in compact sources and are attributed to two effects. One, compact regions which appear relatively fainter at 142 MHz have high emission measures ($EM > 10^6$ pc cm^{-6}) and large optical depths ($\tau_{\rm ff} > 10$). Two, at 1.4 GHz the synchrotron emission from extra-galactic sources is faint compared with the thermal component from galactic regions.

Mid-infrared (MIR) emission at 3.6, 8.0, and 24 μm is compiled as an RGB image in Figure 2.5. Emission at 3.6 μm in blue and observed with the *Spitzer* Space Telescope[5] shows the stellar population of young massive stars as point-like sources of emission (Beerer et al. 2010). 8 μm emission observed with the Midcourse Space Experiment (MSX; Price et al. 2001) and described by Schneider et al. (2006) especially with respect to CO, mainly traces UV heated polycyclic aromatic hydrocarbons

[5] *Spitzer* data was acquired from the archive at `https://irsa.ipac.caltech.edu/data/SPITZER/Cygnus-X/` .

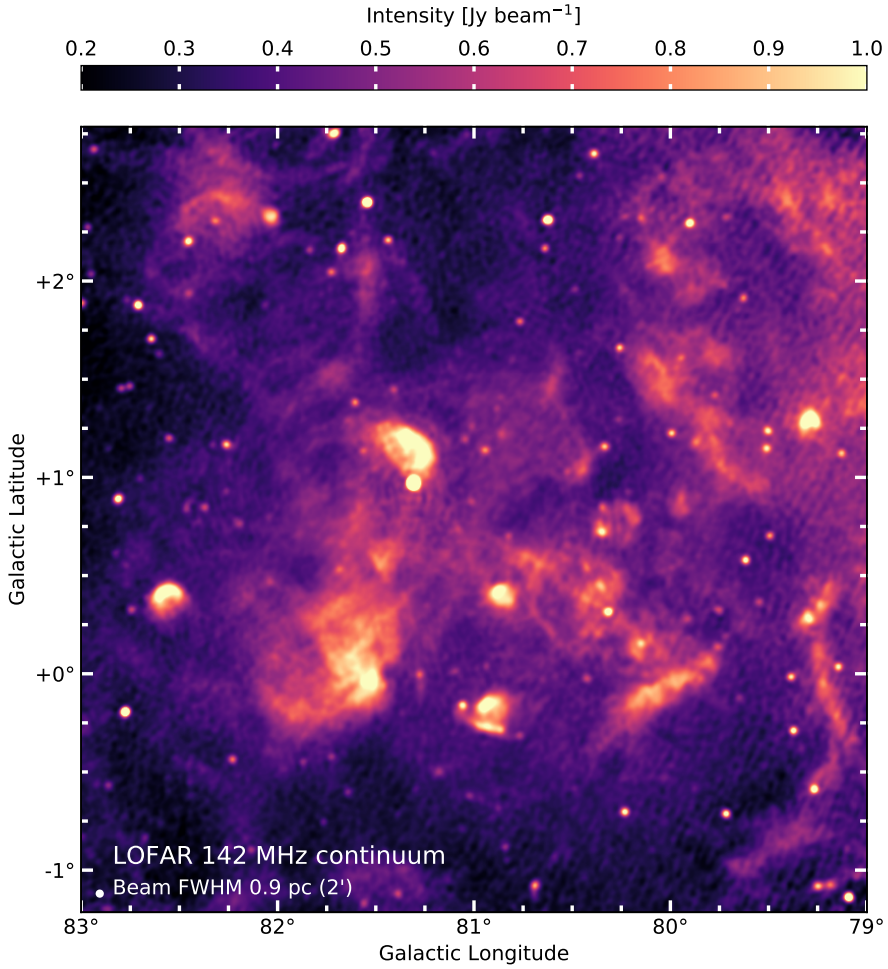


Figure 2.4: LOFAR 142 MHz continuum emission of the 4° × 4° (∼100 pc × 100 pc) area in the Cygnus X region that we analyze at 2′ (0.9 pc) resolution. The emission in this region is primarily thermal, free-free emission from low-density photoionized gas. A number of extra-galactic radio sources are present as bright, point sources. The noise away from bright emission is σ = 25 mJy beam^{-1}.

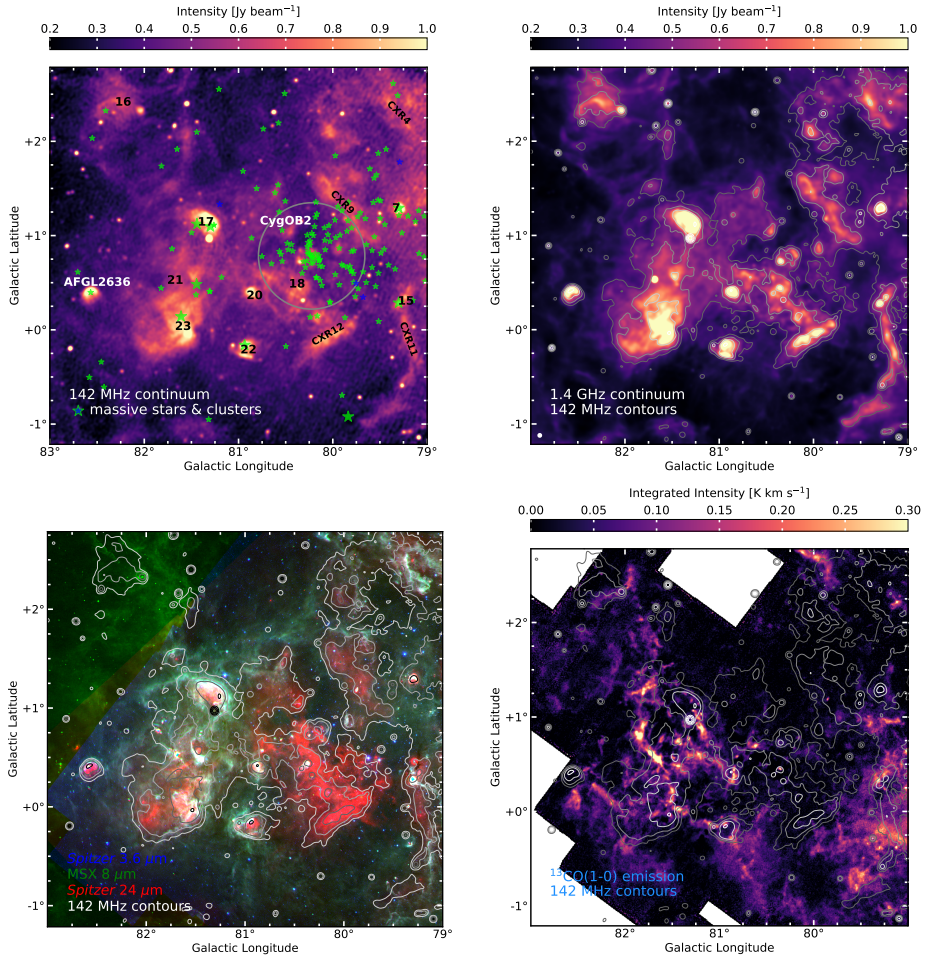


Figure 2.5: *Top Left:* The location of massive stars and star clusters are shown overlaid on LOFAR 142 MHz continuum emission. Small green stars represent massive OB stars of the Cyg OB2 association and within the field (Berlanas et al. 2018), and small blue stars mark the locations of supergiants (Comerón et al. 2020). The gray circle represents the core of Cyg OB2 as identified by Wright et al. (2015). Large green stars mark the location of (open) clusters in the region (Le Duigou & Knodlseder 2002). Additionally, numbers in black denote the position of radio continuum sources as identified by Downes & Rinehart (1966). Filaments identified by Wendker et al. (1991) are designated with "CXR". *Top Right:* 1.4 GHz continuum emission observed through the CGPS (Taylor et al. 2003) and smoothed to a common resolution of 2′. Contours from the LOFAR 142 MHz continuum are overlaid at $0.3 + [3, 3\sqrt{3}, 9, ...]\sigma$ where $\sigma = 0.05$ Jy beam^{-1}. The morphology in the majority of extended features in strikingly similar between 142 GHz and 1.4 GHz, consistent with thermal continuum from low-density, ionized gas; however, optical depth effects present in bright extended regions become apparent as the relative intensity of the emission is brighter at 1.4 GHz (deviating from the $S \propto \nu^{-0.12}$ dependence of optically thin emission). High-density regions such as DR 21 are brighter in 1.4 GHz emission, whereas the point-like sources which are dominated by synchrotron emission, primarily of extra-galactic background galaxies, are fainter in 1.4 GHz emission. *Bottom Left:* MIR emission. In blue, *Spitzer* 3.6 μm tracing the stellar population. In green, MSX 8 μm emission primarily from PAHs heated by UV radiation from massive stars tracing the PDR surfaces. In red, *Spitzer* 24 μm emission of warm dust, corresponding well with diffuse ionized gas observed at 142 MHz. *Bottom Right:*. ^{13}CO (1-0) integrated intensity observed with FCRAO (Schneider et al. 2011) tracing the bulk of the molecular gas.

(PAHs) in photodissociation regions (PDRs), thereby emphasizing interfaces between molecular clouds and intense UV fields from young massive stars. *Spitzer* 24 μm emission is predominantly produced by thermal emission from warm dust co-spatial with photoioinized gas (Churchwell et al. 2006; Calzetti et al. 2007; Salgado et al. 2012). Bright 24 μm emission in red corresponds well with the ionized gas traced by low-frequency radio emission. A correspondence is present for H II regions (partially) surrounded by PDR envelopes as well as for ionized regions without higher density PDRs at their edges.

Molecular cloud emission as traced by ^{13}CO(1-0) (Schneider et al. 2011) shows the cold gaseous reservoir with respect to massive stars in Figure 2.5. The full Cygnus X complex (see Figure 2.3) contains 4.7×10^6 M$_\odot$ of molecular gas with an average density of $\sim$60 cm^{-3} (Schneider et al. 2006). High-contrast elongated filaments are observed as well as diffuse emission from higher density clouds. Embedded regions like DR 21, DR 15 and DR 20 are sites of active star-formation coincident with radio emission, whereas DR 17, DR 18, DR 22, and DR 23 are located at the edges of or well-separated from molecular clouds. Several arc or shell-like structures are also visible in CO.

2.4 Mapping the physical properties of ionized gas

The observed intensity of optically thin, free-free emission is a function of frequency (e.g., Condon 1992; Emig et al. 2020b),

$$I_{\rm ff}(\nu) = \left(1.034 \times 10^5 \text{ Jy beam}^{-1}\right) \left(\frac{A}{\rm sr}\right)^{-1} \left(\frac{EM_+}{10^3 \text{ cm}^{-6} \text{ pc}}\right) \left(\frac{T_e}{10^4 \text{ K}}\right)^{-0.323} \times \left(\frac{\nu}{\rm GHz}\right)^{-0.118} \tag{2.3}$$

and which we express as being dependent on the area of the beam in steradians, A, the electron temperature, T_e, and the emission measure, EM_+, of the ionized medium. The emission measure,

$$EM_+ = \int n_e n_+ {\rm d}l \tag{2.4}$$

is defined by the electron density, n_e, by the ion density, n_+, and by the pathlength integral of the emitting region. Fitting the free-free SED at frequencies where the radiation becomes optically thick and will thus be self-absorbed — such that the optical depth, $\tau_{\rm ff}$, is given by (Condon 1992; Emig et al. 2020b),

$$\tau_{\rm ff}(\nu) = 3.37 \times 10^{-4} \left(\frac{EM_+}{10^3 \text{ cm}^{-6} \text{ pc}}\right) \left(\frac{T_e}{10^4 \text{ K}}\right)^{-1.323} \left(\frac{\nu}{1 \text{ GHz}}\right)^{-2.118} \tag{2.5}$$

— allows for the electron temperature and emission measure to be determined.

In this section we use the LOFAR image to map out the physical properties of the free-free emitting gas in order to better understand the influence of ionized gas (e.g., through gas pressure) and how its ionization is maintained. We fit for the optical depth at 142 MHz (Figure 2.6). From the free-free turnover and SED, we derive the emission measure (Figure 2.7).

2.4.1 Fitting the free-free optical depth at 142 MHz

The radio emission arising from the extended, resolved features in the Cygnus X region is consistent with optically thin thermal, free-free emission at frequencies as low as 408 MHz (Wendker et al. 1991; Xu et al. 2013). A dominating component of thermal emission is further supported by the relative intensities between the 142 MHz and 1.4 GHz image, as discussed in Section 2.3. Therefore, fitting the regions of emission as single components between 1.4 GHz and 142 MHz is a reasonable assumption.

We fit for the optical depth at 142 MHz using the LOFAR data, the 1.4 GHz CGPS data, and the expected frequency dependence of free-free emission. We first prepare the CGPS data for processing by regridding the image to the same pixel grid as the LOFAR image. We use a similar procedure to fit the SED as described in Section 2.2.1. For each pixel, we use the 1.4 GHz data point as the anchoring intensity measurement. Assuming the intensity is thermal and optically thin, we extrapolate it to 142 MHz. If the intensity at 142 MHz is less than the extrapolated value, we fit for a free-free turnover as in Equation 2.1. Otherwise, if the intensity at 142 MHz is greater than the extrapolated value, we assign $\tau = 0$ for that pixel. An example of the free-free fitting process is shown in Figure 2.1. The error of the optical depth is determined from the formal fit errors.

Maps of the free-free optical depth at 142 MHz and its error are shown in Figure 2.6. We find measurable optical depths ($\tau > 0.01$) in a large portion (77% of pixels) of the image. The extended resolved structures have optical depths up to $\tau_{\rm ff} = 1.5$. The largest optical depths are found in the compact sources exceeding $\tau_{\rm ff} > 5$. In the densest regions of AFGL 2637, DR 15, DR 21, and DR 22, optical depth effects are already important (e.g., $\tau > 0.1$) at 1.4 GHz (Wendker et al. 1991); since their intensity at 1.4 GHz is not consistent with the optically thin assumption, the values we determine here underestimate their true optical depths at 142 MHz.

Shallower sensitives and imperfections in the beam model towards the image outskirts result in less reliable intensities and thus optical depths determined there. However, the comparison between our final image and that obtained by feathering (Section 2.2.1) indicate that our results are reliable towards all resolved (DR) sources and in the diffuse emission within $\sim$1.5° from the center of the pointing.

2.4.2 Mapping the free-free emission measure

From the estimated optical depth, we use Equation 2.5 to derive maps of the free-free emission measure. At each pixel, we take the optical depth value, set $T_e = 7400$ K and $\nu = 142$ MHz, and solve for the EM. The results are shown in Figure 2.7. A temperature of $T_e = 7400$ K is the median temperature derived from radio recombination line observations at 4.8 GHz and 2.6′ resolution (Piepenbrink & Wendker 1988) and is consistent with the value of $T_e \approx 7500$ K derived in the DR 21 region with 2.5″ resolution observations (Harris 1973). At 2′ resolution, we find emission measures ranging from $EM_{\rm ff} = (2 - 40) \times 10^3$ pc cm^{-6} in the extended features, and in compact unresolved sources, $EM_{\rm ff} = 4 \times 10^4$ pc cm^{-6} to $EM_{\rm ff} \leq 3 \times 10^5$ pc cm^{-6}.

From Equation 2.5, we have that at 1.4 GHz a free-free optical depth of $\tau_{\rm ff} = 0.1$ (1) corresponds to $EM = 4 \times 10^5$ (10^6) pc cm^{-6}. Thus we expect to measure an upper limit of $EM \lesssim 4 \times 10^5$ pc cm^{-6}. Regions with $EM \gtrsim 4 \times 10^5$ tend to be compact,

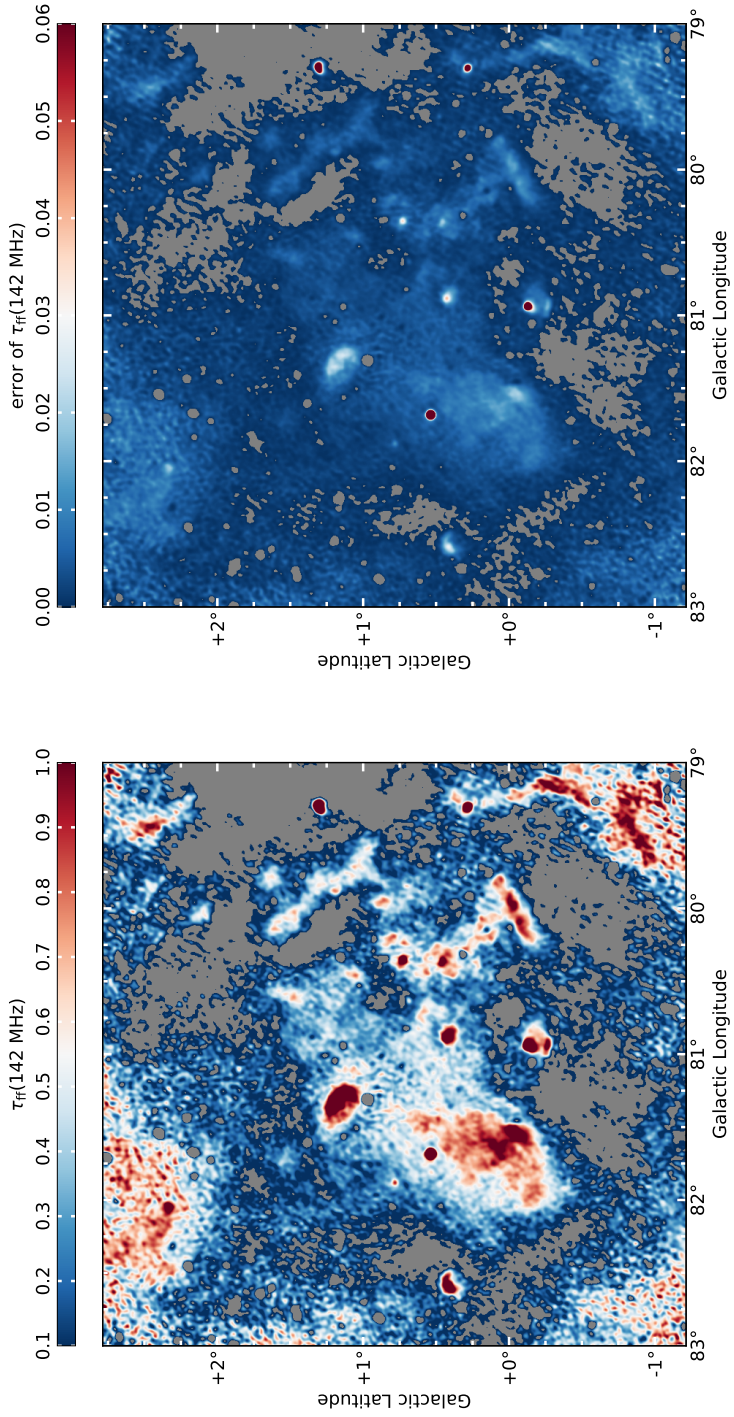


Figure 2.6: The free-free optical depth, $\tau_{\rm ff}$, at 142 MHz. Regions which are optically thick ($\tau_{\rm ff} > 1$) are dark red. Pixels for which a free-free turnover was not fit are shown in gray. Extra-galactic radio galaxies, dominated by synchrotron emission, are present in the region as unresolved point-like sources (in gray) with negligible free-free optical depths. *Right:* The error map of the fitted optical depth at 142 MHz. The error on the optical depth is typically 10%.

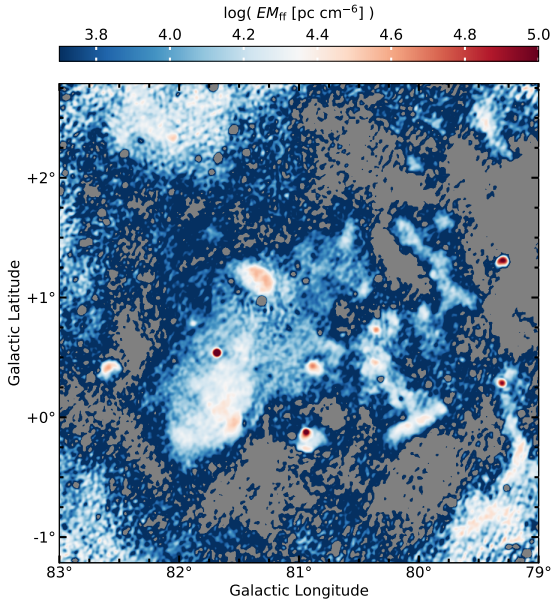


Figure 2.7: Free-free emission measure, $EM_{\rm ff}$, shown in log stretch, of photoionized gas. Tests of the imaging process (Section 2.2.1) find that our results are reliable towards all DR sources and in the diffuse emission within $\sim$1.5° from the center of the pointing. Shallower sensitives and an imperfect beam model towards the image outskirts result in less robust estimates.

dense regions where massive stars have yet to shed their natal material. For example, regions of $\sim$1 pc would have electron densities of $n_e \gtrsim 600$ cm^{-3} at this emission measure limit. Following the same procedure for the lower limit, we have that at 142 MHz a free-free optical depth of $\tau_{\rm ff} = 0.1$ corresponds to $EM = 3 \times 10^3$ pc cm^{-6}. These limits are consistent with what we derive.

The emission measures we map out are consistent with previous lower resolution findings (Downes & Rinehart 1966) and previous estimates of individual regions (Wendker et al. 1991). However, with the higher resolution of our analysis, we do see spatial variations of the emission measure which span a larger range of EM values. These emission measures also agree well with the emission measures derived from optically thin 1.4 GHz continuum. As the sensitivity of the CGPS 1.4 GHz data is significantly deeper for the current bandwidth of LOFAR data employed, we use the CGPS data in the remainder of this analysis.

2.5 Analyzing filaments

Filamentary structure is a prominent feature of the low-density ($n_e \lesssim 100$ cm^{-3}) ionized gas in the Cygnus X region, as we point out in Section 2.3 and which is also seen over a larger area of the Cygnus X region in Figure 2.3. We are motivated to derive properties of these filaments to investigate how ionizing radiation interacts

with and shapes its environment. To briefly summarize the contents of this section, Figure 2.8 shows the 906 filaments identified with the `DisPerSE` algorithm (Sousbie 2011). Figure 2.9 demonstrates the processing and fitting of the radial distribution of a filament profile using `FilChaP` (Suri et al. 2019). We plot the distributions of the peak EM fit to the radial profiles, the best-fit widths, and the inferred electron densities in Figure 2.10. The peak EMs of the filament profiles follow a power law in their number distribution down to the estimated completeness limit. The widths of the filament profiles show a characteristic peak at a median value of 3.6 pc (noting that our beam resolution is 0.9 pc). The median electron density within filaments is $n_e = 38$ cm^{-3}, with densities spanning $n_e = 10 - 500$ cm^{-3}. Figure 2.11 and 2.12 compile and compare the peak EM, width and density.

2.5.1 Identifying filaments

Using the Discrete Persistent Structures Extractor (`DisPerSE`; Sousbie 2011), we identify 906 filaments in the Cygnus X Region as shown in Figure 2.8. `DisPerSE` uses discrete Morse theory to derive information on the topology of a given data set. Filaments are identified as the set of arcs joining maxima and saddle points. Persistence theory is used to filter out and identify filaments with significance. We refer the reader to Sousbie (2011) for a detailed description of the procedures. `DisPerSE` has been widely used to extract filamentary structures both in observational and simulated datasets (e.g., André et al. 2014, and references therein).

We apply the `DisPerSE` algorithm to a map of the emission measure derived from the CGPS 1.4 GHz intensity using Equation 2.3. We filter the filament identification output of `DisPerSE` using the `-breakdown` option to merge overlapping filament segments and with `-trimBelow` to remove arcs below a robust persistence[6] of 5300 pc cm^{-6}. The rms in low intensity regions of the EM map, interpreted as the offset value or background contribution, is 5300 pc cm^{-6} and the standard deviation, interpreted as the noise, is $\sigma \approx 53$ pc cm^{-6}. Our input parameters to `DisPerSE` result in local maximum and saddle point peak EM intensities of $\gtrsim 9500$ pc cm^{-6} being identified. Filaments are output as lists of image coordinates sampled with points at each half pixel shift in direction.

2.5.2 Fitting filament profiles

We characterize properties of filaments using the python-based Filament Characterization Package (`FilChaP`; Suri et al. 2019) which was designed to work together with `DisPerSE`. Here we give a brief explanation of the procedures implemented in `FilChaP` and refer the reader to Suri et al. (2019) for more detailed descriptions. `FilChaP` takes as input a list of filament coordinates. At each sample point along the filament, a radial profile perpendicular to the filament spine is extracted. The radial profile extends to ± 130 pixels, which we choose so that a baseline estimate includes true background emission at any location in the image.[7] The radial profiles extracted

[6] for the specific implementation see `http://www2.iap.fr/users/sousbie/web/html/index959e.html?post/definitions`

[7] We padded the image with true EM intensity so that filament profiles at edges of the region would be fully sampled.

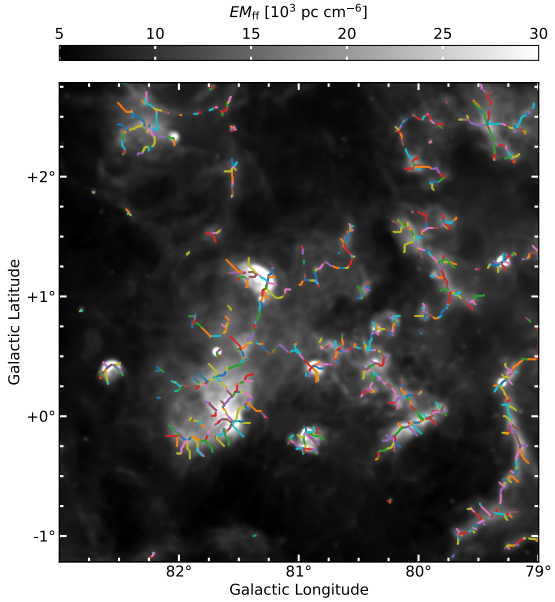


Figure 2.8: The 906 filaments identified with `DisPerSE`. Each filament is displayed in a different (arbitrary) color. The grayscale image is the emission measure, computed from the 1.4 GHz continuum intensity assuming optically thin thermal emission, and is the image on which the filaments are identified.

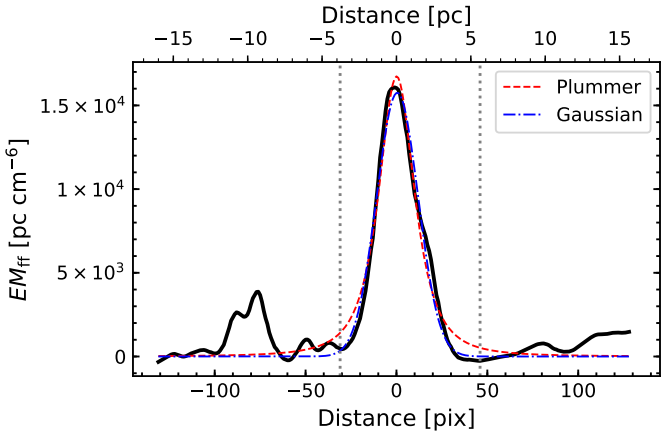


Figure 2.9: An *EM* radial profile of a filament segment, shown with a black solid line. The radial profile has been baseline (and background) subtracted. Negative distances from the spine point to the northeast direction, and positive distances to southwest of the filament. The red, dashed line is the best fit Plummer, $p = 4$ profile used to compute the filament profile properties. The blue dot-dashed line shows the best fit Gaussian for comparison.

at four consecutive sample points of the filament are averaged together. This mean radial profile is baseline subtracted, effectively removing any background emission and baseline gradient. To the mean radial profile, `FilChaP` fits three line profiles — a Gaussian and two Plummer profiles — and computes the second moment[8] width. We demonstrate a radial profile processed and fit with `FilChaP` in Figure 2.9.

Plummer-like functions have been used to describe the column density of a filament with a dense and flat inner portion and a power-law decline at larger radii (e.g., Arzoumanian et al. 2011):

$$\Sigma_p(r) = A_p \frac{\rho_c R_{\rm flat}}{\left[1 + \left(\frac{r}{R_{\rm flat}}\right)\right]^{\left(\frac{p-1}{2}\right)}} \tag{2.6}$$

where ρ_c is the critical density, $R_{\rm flat}$ is the radius within which the density is uniform, p is the power law index, and A_p is a finite constant that is related to the filament's inclination compared to the plane of the sky (here we assume that this angle is zero). Two literature values of the index are fit in `FilChaP`, $p = 4$ for the density of a homogeneous isothermal cylinder (Ostriker 1964) and $p = 2$ as the value derived from embedded star-forming filaments identified with *Herschel* observations (Arzoumanian et al. 2011).

The filament properties that we report are consistent with the four different types of fits and determinations – as we show in Appendix 2.A. Since a Plummer profile with $p = 4$ results in a slightly lower reduced chi-squared, we use the properties derived from this fit to represent our results. Through the line fitting process, 2073 mean radial profiles are constructed. We remove poor fits, with widths less than 3 pixels, resulting in 1934 filament profiles that we analyze.

2.5.3 Filament properties

Peak EM of filaments

We determine the peak of the EM radial profile (which has been background and baseline subtracted) from the best fit of a Plummer $p = 4$ line profile to 1934 mean radial profiles extracted within 906 filaments. In Figure 2.10, we plot a histogram of the peak EMs. We find peak EMs ranging from 500 pc cm^{-6} to 10^5 pc cm^{-6}, with a median value of 5.5×10^3 pc cm^{-6}.

We discuss the completeness of the distribution. We previously mention, in Section 2.4.2, the upper limit to our EM sensitivity of 4×10^5 pc cm^{-6}. Since the rms in low intensity regions is equal to 5.3×10^3 pc cm^{-6}, we expect an upper bound to fully sampled emission measures of $< 3.5 \times 10^5$ pc cm^{-6}. Furthermore, we applied a cut to select filaments with a peak $EM > 9.5 \times 10^3$ pc cm^{-6}. Subtracting the "background" EM, which will be greater than or equal to 5.3×10^3 pc cm^{-6}, results in an estimated completeness of 4.2×10^3 pc cm^{-6}. Therefore, filaments with a fitted peak EM below 4.2×10^3 pc cm^{-6} are not fully sampled.

[8]The nth moment of a distribution I is given by, $m_n = (N\sigma^n)^{-1}\Sigma_i^N I_i(x_i - \bar{x})^n$, where $\bar{x}$ is the intensity weighted mean position of the profile, I_i is the intensity at position x_i, and σ is the intensity weighted standard deviation of the profile.

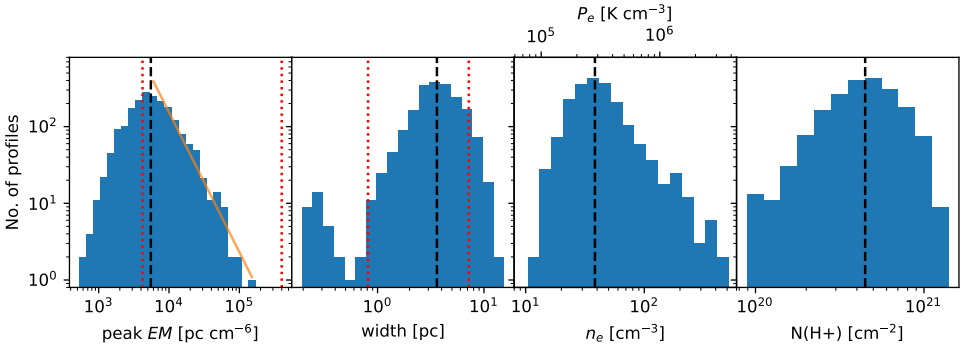


Figure 2.10: Histograms of filament properties inferred from fits to their EM radial profiles. We fit a Plummer, $p = 4$ line profile to the radial distributions. Black dashed lines mark the median values. Red dotted lines mark the estimated completeness limits in the EM and width plots. *Left:* Peak emission measure (EM) of the best fit in units of pc cm^{-6}. The median EM of the filaments identified in this analysis is 5530 pc cm^{-6}. The orange solid line shows the power-law, $\beta = -1.8 \pm 0.1$, fit to the distribution for $EM > 5530$ pc cm^{-6}. *Center left:* Filament FWHM in units of pc. The median FWHM of the filaments identified in this analysis is 3.6 pc. The resolution of the beam FWHM (2′, 0.9 pc) is marked (with a red dotted line) as the lower completeness limit. Widths smaller than the beam resolution result from poor fits. *Center right:* Average electron density of the filament profiles in units of cm^{-3}, computed with the best fit peak EM and width as $n_e = \sqrt{EM/\ell}$ where ℓ is the FWHM of the Plummer profile. The median electron density of the filaments identified in this analysis is 38 cm^{-3}. The electron pressure, computed as $P_e/k_{\rm B} \sim 7400$ K $\cdot\, n_e$, is shown along the top axis. *Right:* The column density of ionized hydrogen, letting $n_e \sim n_{\rm H^+}$.

We fit a power-law to the number distribution of the peak EM for bins which we consider to be complete. This results in a best fit power-law index of $\beta = -1.8 \pm 0.1$.

Filament widths

We determine the widths from the Plummer $p = 4$ fits to the mean radial profiles of the filaments. The FWHM of a Plummer $p = 4$ profile is FWHM $= 1.533 \cdot R_{\rm flat}$, see Equation 2.6 for $R_{\rm flat}$. In Figure 2.10, we plot the best fit FWHM of each mean radial profile, converted into physical units assuming an average distance of $d = 1.5$ kpc to the complex. A peak in the width distribution arises at 3.6 pc, which is well separated from the beam resolution of 0.9 pc (2′). We note that widths smaller than the image resolution likely result from poor, unreliable fits and should be interpreted as non-physical. While the `DisPerSE` algorithm does not prevent large scale features from being identified, there are regions of diffuse emission that could possibly be identified as "filaments" at lower resolution. For the EM we probe, we place a conservative upper completeness limit to the filaments widths of 7.2 pc, twice the median value. In Figure 2.11 we show the filament widths with respect to their spatial location on the EM map. Bright compact regions of emission tend to have smaller widths (in blue). Also notable is the DR 23 region which has somewhat wider filaments (in orange/red) than average.

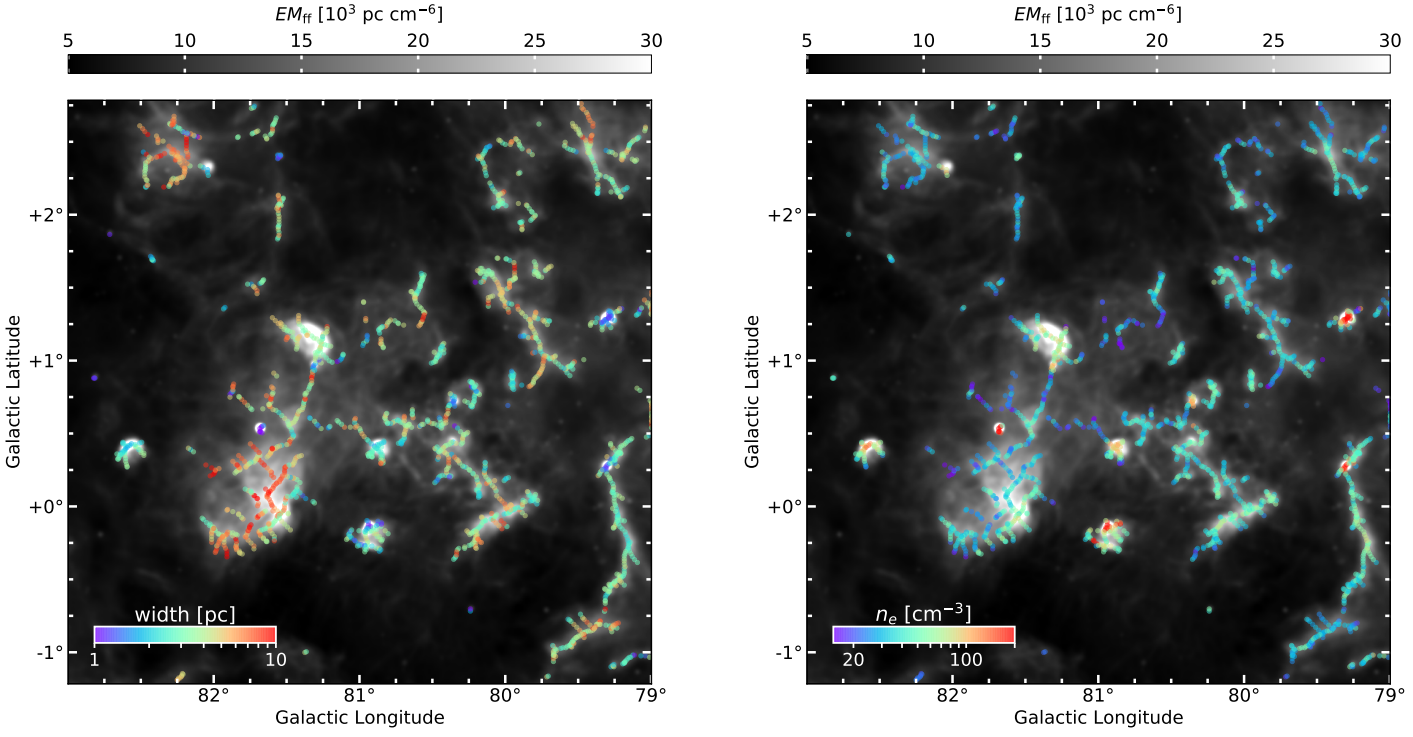


Figure 2.11: *Left:* Width of filaments' mean radial profiles are represented as colored data points and overplotted on the *EM* map. Compact bright regions, tend to have smaller widths. Another notable is the DR 23 region which has wider filaments than average. *Right:* The estimated density of each radial profile are represented as colored data points and overplotted on the *EM* map. As expected, densities are highest in known star-forming regions. Assuming the filaments are fully ionized and have a common temperature of 7400 K, the thermal pressure would be directly proportional to the density, $P_{\rm th}/k_{\rm B} \sim 2 \cdot 7400\ {\rm K} \cdot n_e$.

Electron densities of filaments

We estimate the electron density of the filament from the best fit peak EM and width of the Plummer $p = 4$ line profile, as $n_e = \sqrt{EM/\text{width}}$. In Figure 2.10, we plot the distribution in electron density of the mean radial profiles. Electron densities ranging from 10 cm^{-3} to 500 cm^{-3} are found, with a median value of 38 cm^{-3}. In Figure 2.11 we show the filament densities with respect to their spatial location on the EM map.

From the estimated completeness of emission measure and spatial scales, we estimate our completeness and sensitivity to electron densities. At the lowest possible (background subtracted) emission measure, $EM > 4.2 \times 10^3$ pc cm^{-6}, the upper bound on the width, FWHM < 7.2 pc, puts a hard limit to the possible electron densities, $n_e > 24$ cm^{-3}. Likewise, a hard upper limit to the electron density that we reliably detect, $n_e < 710$ cm^{-3}, arises from the largest possible emission measure, $EM < 4 \times 10^5$ pc cm^{-6} and the smallest width of 0.9 pc. However, our census of electron densities may not be complete for the electron densities within the range $24 < n_e$ [cm^{-3}] < 710.

Column density of ionized gas

From the density and width we compute the column density of ionized gas. We assume the gas is completely ionized, with $n_e \sim n_{\rm HII}$, and compute a column density of $N(\text{H II}) = \int n_e \mathrm{d}\ell = \sqrt{EM \cdot \text{width}}$. The ionized hydrogen column densities we derive range from $(10^{20} - 10^{21})$ cm^{-2} with a median value of 4.6×10^{20} cm^{-2}. Based on the estimated completeness of the emission measure and filament width, we can probe column densities in the range of $1.8 \times 10^{20} < N(\text{HII}) < 5.2 \times 10^{21}$. However, given how we derive the column density, we do not expect our census to be complete for any range of column densities.

Pressure of ionized gas

We calculate the electron pressure, as $P_e/k_{\rm B} \sim T_e n_e$, from the ionized gas densities and again assume $T_e = 7400$ K. We find electron pressures of 7.4×10^4 K cm^{-3} to 3.7×10^6 K cm^{-3} with a median value of $P_e/k_{\rm B} \sim 2.8 \times 10^5$ K cm^{-3}. We show these values along the upper axis of the electron density histogram in Figure 2.10. These regions are over-pressured compared with typical values in the diffuse ISM, $P/k_{\rm B} \sim 3.8 \times 10^3$ K cm^{-3} (Jenkins & Tripp 2001; Jenkins et al. 2011). Over-pressure compared to the general ISM is expected for a region where massive stars are interacting with their environment, as massive stars create regions of high pressure that will expand. Indeed the pressures we find do coincide with a small fraction ($\sim$0.05%) of nearly all gas surveyed that has a large pressure ($> 3 \times 10^5$ K cm^{-3}) and which is more prevalent at high velocities or for regions with enhanced starlight densities (Jenkins et al. 2011).

Correlation between filament properties?

In Figure 2.12 we plot the peak EM of the filaments' profiles as a function of the profile width. We color the data points by the inferred electron density. The filament peaks

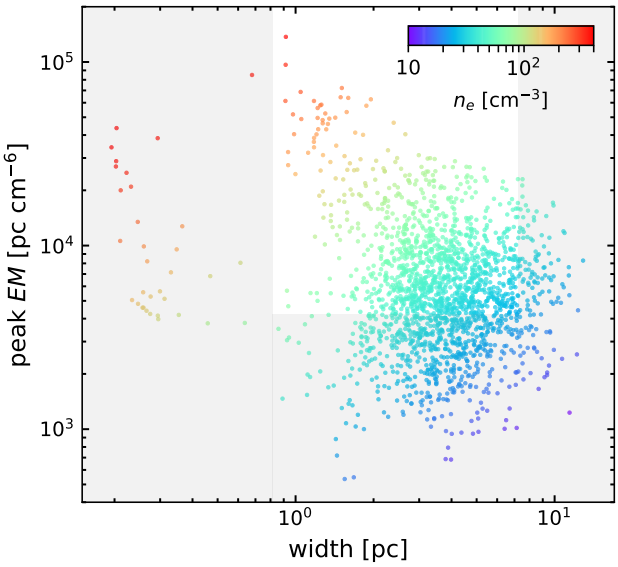


Figure 2.12: Peak EM of the best fit to the filament radial profile plotted with the best fit FWHM. The areas in gray represent values which fall outside of our completeness limits. The colors of the data points represent the electron density inferred from the filament EM and width. Profiles with FWHM smaller than the resolution beam of 0.9 pc likely have poor fits.

and widths do not appear to be correlated. A dearth of high intensity EM filaments with large widths would physically be harder to maintain and may represent a true de-populated portion of this plot.

2.6 Discussion

In this section we discuss the sources which maintain the ionization of the thermally emitting gas that we observe. We find Cyg OB2 may have a considerable influence, ionizing up to two-thirds of the emission in this region. We also discuss what is forming the filaments – likely photoevaporating surfaces of neutral material, flowing into lower-density volume-filling ionized gas. While our calculations suggest that the stellar winds of Cyg OB2 may be dissipating turbulence in the form of transitory filaments, the filament densities correlate with incident radiation and thus suggest only a minority of filaments are influenced by the stellar winds of Cyg OB2. We place our results in the context of ionized gas surveyed with [N II] FIR fine structure lines — finding remarkable agreement with properties — and we construct a framework in this region for ELD ionized gas and how it is maintained. Lastly, we inform on a bright future for LOFAR observations of diffuse emission in the Galactic plane.

2.6.1 Source of ionization of the filaments

The recombination time of warm ionized gas is generally short; $t_{\rm rec} = (\alpha_B n_e)^{-1} \sim$ 2600 yr for gas of density $n_e = 38$ cm^{-3} and temperature $T_e = 7400$ K. Thus an active source of ionizing radiation is required to maintain it. In this subsection, we describe what sources ionize the low-density gas that we observe. We ultimately find that the ionizing radiation from Cyg OB2 (Figure 2.13) can be responsible for up to two-thirds of the total ionized emission in the region, and similarly, two-thirds of ionized emission in filamentary structure. The remaining one-third of emission requires a local source of ionizing radiation, from active regions of star formation or massive field stars. Filaments that we attribute to be (externally) influenced by Cyg OB2 (e.g., see Figure 2.14 and Table 2.1) contrast with the conclusions reached by Wendker et al. (1991) that the filaments result from strings of intrinsic B stars. A general conclusion from this section is thus that massive OB stars which are no longer locally confined by dense gas play a prominent role in shaping filamentary structure of ionized gas.

Local massive stars

Signatures of the youngest, most dense regions of active star formation are not traced at these frequencies and resolution (e.g. DR 15, DR 21, and W75N). Somewhat more evolved regions which have begun bursting open have lower density and smaller emission measures, as we see, leading to a more pronounced spatial separation between the star and warm gas which it ionizes. In these regions, local ($d \lesssim 10$ pc for a single O3 star) sources supply ionizing photons. Examples of this include AFGL 2636, BD+43° 3654 and DR 16, DR 20, DR 22, and the complex environments of and between DR 17 and DR 23, with the notable distinction that all of these regions have not just one but a small cluster of massive OB stars within them. These regions are generally found at larger distances ($d \gtrsim 30$ pc) from Cyg OB2, to the east in the image. Likewise (single) massive field stars, for example to the northeast and southeast regions of the image, shape local filamentary structure.

Cyg OB2

To explore the role of Cyg OB2 in ionizing filamentary structure, we construct a map of the EUV (> 13.6 eV) ionizing photon rate per unit area from Cyg OB2, as shown in Figure 2.13, from catalogs of OB stars (Berlanas et al. 2018, and references therein). We describe how the map is created in detail in Appendix 2.B, following the procedure of Tiwari et al., in prep. Two important features of the map: (i) it is a simple 2D projection, and (ii) we do not attempt to account for any absorption or attenuation of ionizing photons. The total ionizing photon rate of the association from our model is 6×10^{50} EUV-photons s^{-1}, which we note agrees with the expectation of a zero age main sequence association with stellar mass $M_\star \approx 1.6 \times 10^4$ M$_\odot$ (Starburst99; Leitherer et al. 1999).

In Figure 2.13 we compare the ionizing photon flux from Cyg OB2 with the ionizing photon flux estimated directly from the optically thin thermal radio continuum at 1.4

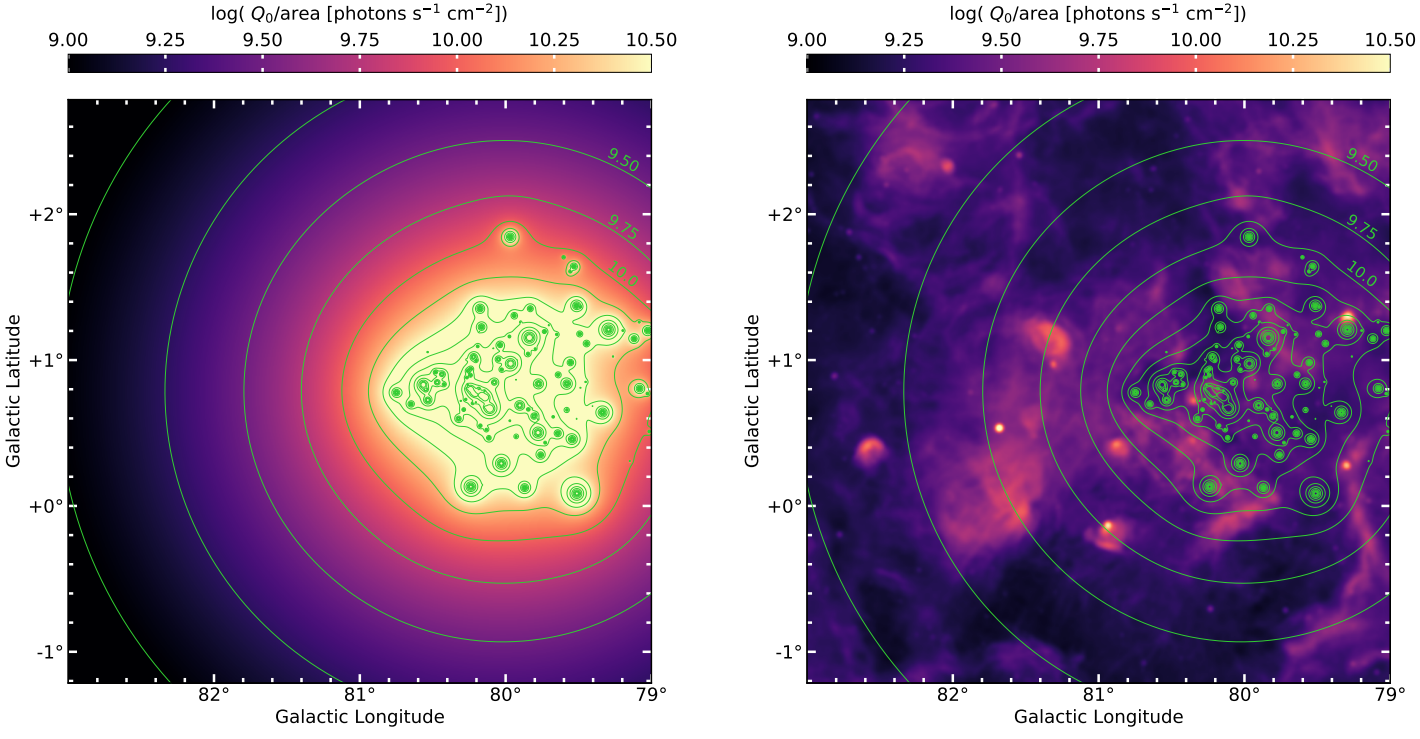


Figure 2.13: *Left:* A map of the ionizing photon (> 13.6 eV) rate per unit area constructed from the observed OB stars of the Cyg OB2 association (for details see Appendix 2.B). Green contours are shown at $\log(Q_0/\text{area [photons s}^{-1}\text{ cm}^{-2}]) = (9, 9.25, 9.5, ...12.25)$. *Right:* The ionizing photon rate per unit area as traced by thermal free-free emission. The green contours of the Cyg OB2 ionizing field are overlaid.

GHz (e.g., Emig et al. 2020b),

$$Q_0/\text{area} = \left(1.998 \times 10^8 \text{ photons s}^{-1} \text{ cm}^{-2}\right) \left(\frac{EM}{10^3 \text{ pc cm}^{-6}}\right) \times \left(\frac{T_e}{10^4 \text{ K}}\right)^{-0.833-0.034 \ln(T_e/10^4 \text{ K})} . \quad (2.7)$$

The Cyg OB2 ionizing photon flux outshines a large portion of the inferred flux from the radio emission. We quantify how much of the low-density ionized gas can be maintained by the ionizing photons from Cyg OB2. We take the ratio of the two images shown in Figure 2.13, which equates to the observed ionizing photon flux divided by Cyg OB2's output ionizing photon flux. We compute that at 67% of the pixels in this region the ionizing photon flux from Cyg OB 2 is sufficient to maintain the ionization.

We make a rough comparison of the influence of Cyg OB2's ionizing radiation on filaments. Starting from the mask of the filament spines, we extend the mask at each filament segment to a square region of 29 pixel sides centered on the segment. This results in about a third of the pixels in the image being attributed to a filament. Taking the ratio of the images and considering only the masked pixels, we again find that for 67% of the pixels the ionizing photon flux from Cyg OB 2 is sufficient to provide the ionization.

We focus on some of the features and regions which appear to be influenced by Cyg OB2 in Figure 2.14. These radio continuum features correlate extremely well with 24 μm emission (see Figure 2.5) and do not appear to be directly related to individual H II regions. In Table 2.1 we include estimated properties of identified filaments. The widths, emission measures, and densities have been computed with the filament analysis described in Section 2.5. The length has been estimated by eye using the furthest extent of a straight line along a filament, starting and ending where the emission reaches a local intensity of half maximum. In this manner, the length of the curved features such as CXR 9a,b are slightly underestimated. We also compute the ionizing photon flux and the ionizing photon rate. In these filaments, the ionizing photon flux from Cyg OB2 at their projected distances is a factor of 10 or more than required.

Internal B stars?

The ionizing photon rates of the filaments exemplified in Table 2.1 are $\sim 2\times10^{48}$ EUV-photons s^{-1}, a rate which could be supplied by early type B stars local to the filaments. Wendker et al. (1991) argued for the filaments being internally ionized by about 3 to 10 stars of type B2 to B3. They give additional reasoning that filaments ionized externally are typically smaller, strongly curved, and with shallower intensity gradients at the edges. However, searches for B stars in these (low extinction) regions have subsequently been made and can now rule out the presence of (B) stars with confidence (Comerón & Pasquali 2012; Comerón et al. 2020). Furthermore, in Section 2.5 we mention examples of externally illuminated ionized filaments of similar length, width and curvature as the filaments observed in the Cygnus X region, and

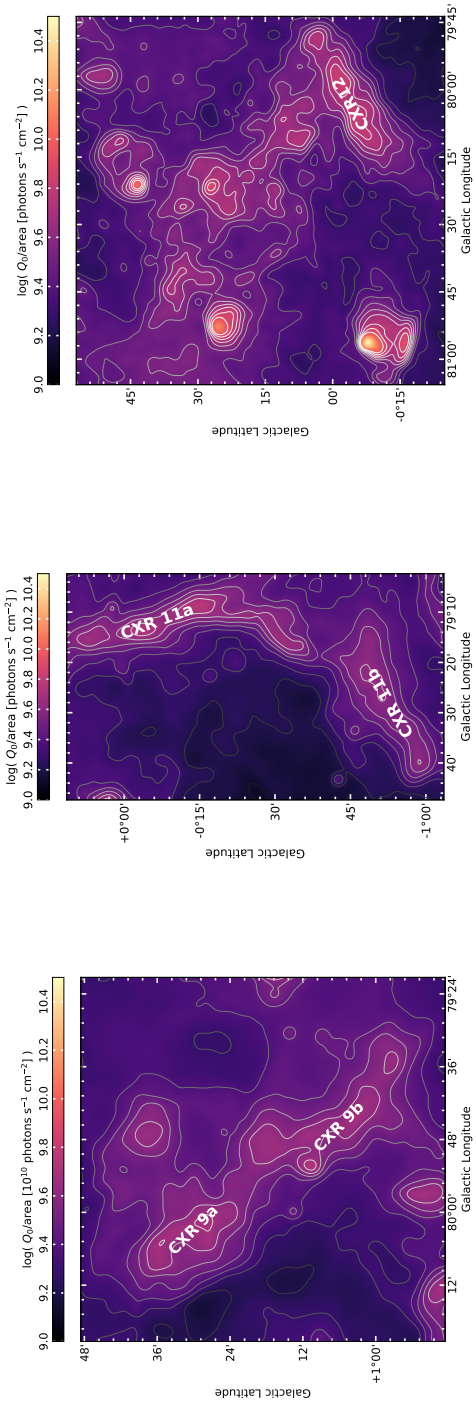


Figure 2.14: Here we show a zoom-in on filaments whose properties we discuss in more detail (see Table 2.1 and Sections 2.6.1 & 2.6.2).

Table 2.1: Properties of select filaments which are likely influenced by Cyg OB2.

Filament	Width (pc)	Length (pc)	EM (pc cm^{-6})	n_e (cm^{-3})	Q_0/area (10^9 photons s^{-1} cm^{-2})	Q_0 (10^{48} photons s^{-1})	M_+ ($M_\odot$)	D_{CygOB2} (pc)
CXR 9a	5.3	8.0	5600	33	1.4	2.8	740	19
CXR 9b	4.6	14.9	5600	35	1.4	2.1	1100	15
CXR 11a	3.6	14.7	7500	46	1.9	1.7	880	34
CXR 11b	4.6	13.8	4200	30	1.1	1.5	880	47
CXR 12	4.2	15.0	7800	43	2.0	2.4	1100	22

Width is the median FWHM width of all profiles of the filament.
Length is the by-eye estimate of the longest extent of the filament.
EM is the median peak emission measure over all profiles of the filament.
n_e is the median electron density estimated from all profiles of the filament.
Q_0/area is the median of the EUV ionizing photon flux estimated from all profiles of the filament.
Q_0 is the median of the EUV ionizing photon rate estimated from all profiles of the filament.
D_{CygOB2} is the approximate projected distance from Cyg OB2.

in the next section (Section 2.6.2), we extensively discuss evidence for the (external) mechanisms responsible for the forming of filaments.

2.6.2 Origin of the filaments

Photoevaporation

When an ionized volume of gas is not in (pressure) equilibrium with neutral gas, an ionization front propagates into the neutral medium. With increasing distance from the ionizing source, an ionization front is preceded by a layer of swept up neutral gas and somewhat further away an H_2 dissociation front (Elmegreen & Lada 1977). Since gas at the ionization front is over-pressured compared with the inner, low-density ionized gas, a flow occurs towards the ionizing source. The back-reaction created when the ionized gas pushes off the neutral material (the rocket effect) may eject even more mass and exert significant forces. Ionized gas with enhanced density (and intensity) is observed at this photoevaporating boundary. Photoevaporating surfaces are also referred to as ionized boundary layers and occur in champagne flows (Tenorio-Tagle 1979) and blister-like H II regions.

We sketch a rough portrait of the PDR like transition from ionized, atomic, to molecular gas at a photoevaporating surface (see (Hollenbach & Tielens 1999) and references therein). First we derive the width of the neutral region of the interface in order to compare the distribution and morphology of the ionized gas (3.6 pc width) and the PAH emission. Assuming the gas in the filaments is fully ionized up to the ionization front, the median thermal pressure is $P/k_{\rm B} = 2n_e T_e \approx 5.6 \times 10^5$ K cm^{-3}. The temperature of neutral material is $T \sim 100$ K at an extinction of $A_{\rm V} \sim 2$ mag. Assuming the ionized gas at this ionization front is in pressure equilibrium with neutral material, the gas would have a density $n_{\rm PDR} \sim 5600$ cm^{-3}. Adopting $\mathrm{N_{PDR}}/A_{\rm V} = 2.1 \times 10^{21}$ cm^{-2} mag^{-1} (Zhu et al. 2017), then $N_{\rm PDR}/n_{\rm PDR} \approx 0.2$ pc. Thus, for these high pressures, we expect a relatively narrow region of approximately 0.2 pc (or 30″ for Cygnus X) of neutral gas bright in PAH emission between the ionized gas filament and the dissociation front. For less intense radiation or a less dense medium, the width of the neutral region extends further.

Photoevaporating boundaries are seen at the PDR interfaces traced by 8 μm emission in Figure 2.5 in more compact regions of star-formation that have local sources of intense radiation – e.g., filaments associated with sources AFGL 2636, DR 7, DR 17, DR 22 and DR 23. Other possible examples are the filaments which fall in the regions between DR 16 and DR 17 – where in Figure 2.5, emission is found surrounding massive stars but not overlapping with them.

For gas flowing away from the ionization front approximately at the sound speed, $C_{\rm H\ II} \sim 10$ km s^{-1},[9] a filament of 3.6 pc would be visible for 0.4 Myr, but would continue eating away at the neutral clouds. An ionized mass of 940 $M_\odot$, which is the mean value of the Table 2.1 entries, implies a mass loss rate of the neutral cloud of 2400 $M_\odot$ Myr^{-1}. Each photoevaporating surface with these approximate characteristics would evaporate 1.2×10^4 $M_\odot$ of ionized gas over the lifetime, ∼5 Myr,

[9] $c_{\rm s} = \sqrt{k_{\rm B}T/\mu m_{\rm H}}$, with μ the mean weight of the particles and $m_{\rm H}$ the mass of a hydrogen atom

of Cyg OB2. This mass is comparable to the stellar mass of Cyg OB2 and about 10% of the mass of a typical molecular cloud of 10^5 $M_\odot$ (Tielens 2005).

The properties of the filaments identified in Cygnus X are comparable with those of the California Nebula and the ionization front IC 434. The ionized filaments of the California Nebula have $EM = 2200$ pc cm^{-6} and FWHM size of 1.1 pc, and they arise from a single, high-velocity O star passing by and illuminating a molecular cloud some 50 pc away (e.g., Bohnenstengel & Wendker 1976). The ionization front IC 434 arises from a photoevaporating flow ionized by σ Ori with an emission measure of $EM = 2 \times 10^4$ pc cm^{-6} measured in Hα and also is about 1 pc wide (e.g., Ochsendorf et al. 2014). The densities computed for the California Nebula and IC 434 assuming $n_e = \sqrt{EM/\ell}$ are $\sim$44 and 47 cm^{-3}, respectively.

To determine if the filaments show a relationship to ionizing radiation, we estimate the dimensionless ionization parameter, U, and plot it as a function of density in Figure 2.15. Following Pellegrini et al. (2011), we define the ionization parameter from the ionizing photon flux and density as

$$U = \frac{Q_0}{\pi(\mathrm{width})^2 c n_e}. \tag{2.8}$$

The ionization parameter is indicative of the ratio between radiation and gas pressures. The electron density is expected to depend on U when its properties depend on the incident radiation field – also including when radiation pressure has influence. The ionization parameter has typical values of $U \gtrsim (10^{-2} - 10^{-3})$ in H II regions while the WIM has $U \sim 10^{-4}$ (e.g., Tielens 2005; Kewley et al. 2019). Because there is a general trend of increasing U as the density increases in our data, this indicates that ionizing radiation influences the filaments' properties. This would argue against stellar winds shock-heating the ionized gas and regulating the ionized gas pressure and density (see Section 2.6.2).

We mention a caveat in these calculations. If the "filaments" are truly photoevaporating surfaces, then their emission is more likely to be distributed in sheets rather than filaments, and thus traversing a pathlength along the line of sight. The electron densities of the ionized gas measured in this way would be about two times smaller, if we assume that the pathlength along the line of sight is equal to the longest dimension of the filament (which seems to be at most a factor of four longer than the width).

Looking at this scenario from a different perspective, we also discuss whether filaments that do not have or are significantly displaced from PAH or CO emission can plausibly be photoevaporating flows. The largest filaments, which are far from individual star-forming regions and may be influenced by Cyg OB2, have significant offsets between the ionized filament and neutral material (as traced by CO or 3.6,8 μm emission). Examples of this include CXR 12, CXR 11a,b, and the region just south-east of the Cyg OB2 core — all of which are shown in Figure 2.14. Using CXR 11b as an example, the projected distance between the ionized ridge and the PAH emitting ridge is $d \approx 7' \sim 3$ pc. The large separation would imply that the neutral gas is rather diffuse; the estimated density is $n_{\mathrm{PDR}} = \mathrm{N}_{\mathrm{PDR}}/d \sim 450$ cm^{-3}, since $\mathrm{N}_{\mathrm{PDR}} \approx 4.2 \times 10^{21}$ cm^{-2}. Pressure equilibrium, at $T \approx 100$ K would thus be $P/k_{\mathrm{B}} \sim 4.5 \times 10^4$ K cm^{-3}. For a temperature of 7400 K, the filament's ionized density is expected to be 6 cm^{-3}. Even if we assume CXR 11b is a sheet rather than a filament and we

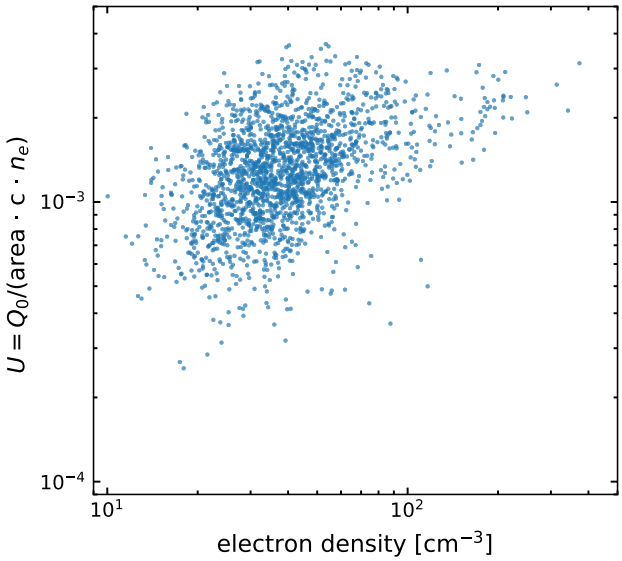


Figure 2.15: The dimensionless ionization parameter, U, estimated for each radial profile plotted against its electron density. The correlation between density and U indicates that the filament properties depend on the incident radiation field. No correlation as a function of density would indicate other (feedback) processes influence the physical properties (density) of the filaments.

let the pathlength along the line of sight equal the length along the plane of the sky, $\ell = 13.8$ pc, the electron density is still a factor of three too large, $n_e = \sqrt{EM/\ell} = 17$ cm^{-3}. In conclusion, ionized filaments that do not have neutral gas abutting the filament edge are likely not photoevaporating surfaces.

Stellar Winds

Next, we turn our attention to the strong stellar winds of Cyg OB2. Multiple interactions of the winds from massive stars produce diffuse shock-heated X-ray gas.[10] A region of hot ($T \gtrsim 10^6$ K) plasma fills the volume surrounding the star(s) at small radii. Swept-up interstellar gas is compressed at larger radii (e.g., Weaver et al. 1977; Harper-Clark & Murray 2009), forming a relatively dense shell of ionized gas and warm dust that is in pressure equilibrium with the hot gas.

A *Chandra* X-ray survey centered on Cyg OB2 has revealed a $R \approx 4.1$ pc cavity of hot ($T_X \approx 5 \times 10^6$ K) and diffuse ($n_X \approx 0.06$ cm^{-3}) gas (Albacete Colombo et al. 2018). In comparison, the core of Cyg OB2 marked with a gray circle in Figure 2.5, has a radius of $r = 34'$ ~15 pc. X-ray observations in the Cygnus X region are complicated by high levels of absorption (especially soft X-rays, which are dominated by wind-ISM interactions) in the Galactic plane; the total extent and intensity of the X-ray emission could reasonably be underestimated. Estimating the volume averaged pressure from hot gas, we find $P_X/k_B \approx 2 n_X T_X \sim 6 \times 10^5$ K cm^{-3}, where n_X and T_X

[10] Any previous (generations of) supernovae would also contribute to shock-heated X-ray gas.

are the electron number density and temperature. This matches the median thermal pressure of the ionized filaments, which is what would be expected for ionized gas in pressure equilibrium with hot shocked gas.

There are several radio filaments abutting a cavity around the Cyg OB2 stars and where dense gas counterparts as traced through PAH or CO emission are not immediately obvious. The main region under consideration is shown in Figure 2.14,*left*; it stretches ∼32 pc in projection, ending in the southwest with CXR 12. As seen most clearly in Figure 2.5, the 24 μm emission (coincident with thermal radio emission) does not appear to be bordered by PAH emission at 8 μm. CO avoids the regions of ionized gas (warm dust), however CO is present in projection in some regions at larger radii. Another region to consider is CXR 9. While faint PAH emission and traces of CO emission are seen towards the south, in CXR 9b, dense gas does not seem the be related to CXR 9a. The regions in consideration are close to Cyg OB2 in projection, falling within about a 20 pc radius.

If the filaments (which are not associated with neutral gas) are influenced by stellar winds, they may be short-lived turbulent features. A correlation between the luminosity of giant ($r > 10$ pc) H II regions (and their size) and the line-width has been established (e.g., Terlevich & Melnick 1981), though the origin of the supersonic turbulence may be gravity-driven or (wind)feedback-driven. We note the ionized gas line-widths in the (resolved) filaments of Cygnus X are broader than the purely thermally broadened expectation. For example, the median ionized gas full width half maximum (FWHM) of the zoom-in regions of Figure 2.14, determined from H110α (4.8 GHz) observations, is $\Delta v = 27$ km s^{-1} (Piepenbrink & Wendker 1988), whereas the Doppler-broadened thermal line-width[11] is $\Delta v_{th} = 16$ km s^{-1} (Brocklehurst & Seaton 1972). The rms turbulent velocity width is calculated as $\Delta v_{rms} = \sqrt{\Delta v^2 - \Delta v_{th}{}^2} \sim 22$ km s^{-1} (Salgado et al. 2017b).

We determine the thermal energy in the hot gas as $E_X = P_X(\frac{4}{3}\pi R^3) \sim 7\times10^{47}$ erg. Using Starburst99 (Leitherer et al. 1999), the energy output from stellar winds of an $M_\star = 2 \times 10^4$ M$_\odot$ population of age 4 Myr is estimated at $E_w \sim 3 \times 10^{52}$ erg. Since the mechanical energy of the wind is more than four orders of magnitude larger than the hot gas energy and theory predicts it should be roughly half (Weaver et al. 1977), the hot gas is likely venting into the medium. Judging also by the morphology of the gas, the hot gas does not appear to be entirely well contained. A possible vent may be the region to the North which seems to be devoid of gas and dust tracers – H II, CO, MIR, FIR.

Averaging the mechanical energy of the wind over the 4 Myr lifetime of the association, the rate of energy injection by the winds is estimated at $\dot{E}_w \sim 2\times10^{38}$ erg s^{-1}. Let's compare that with the calculated energy dissipation rate in the potentially turbulent-induced filament. The (kinetic) energy in the observed filaments is $E_{fil} = \frac{1}{2}M_+(\Delta v_{rms})^2 \sim \frac{1}{2}(940\ M_\odot)(22\ \text{km s}^{-1})^2 \sim 5 \times 10^{48}$ erg. Turbulent energy will be dissipated on a timescale comparable to the sound crossing timescale, $t = \text{width}/\Delta v_{rms} \sim (3.6\ \text{pc} / 22\ \text{km s}^{-1}) \sim 0.2$ Myr. We calculate the energy dissipation rate (potentially) provided by filaments by bringing the energy and timescales

[11] $\Delta v = (30.25\ \text{km s}^{-1})\left(\frac{m_p}{m}\frac{T_e}{2\times10^4\ \text{K}}\right)^{1/2}$ where Δv is the line FWHM, m_p is the proton mass and m is the nuclear mass (Brocklehurst & Seaton 1972).

together, $\dot{E}_{\rm fil} \sim 8 \times 10^{35}$ erg s^{-1}. Hence, we conclude that only $\sim 0.4\%$ of the mechanical energy in the stellar winds of the Cyg OB2 association is coupled to turbulent energy in the ionized gas.

2.6.3 Comparing filament properties with [N II] findings

In Section 2.6.2, we noted that the filament properties (EM, width, density) are consistent with filaments observed at photoevaporating surfaces of ionized boundary layers and stellar wind compressed ionized gas. Here we focus on a comparison of filament properties with ionized gas surveyed through [N II].

Ionized gas with electron densities distributed around 38 cm^{-3} appears to be consistent with ionized gas traced through the fine structure lines of [N II] at 122 and 205 μm with *Herschel* PACS (Goldsmith et al. 2015). Goldsmith et al. (2015) find a mean electron density of $n_e = 29$ cm^{-3} over 96 lines-of-sight of 16″ each in the Galactic midplane. While [N II] fine structure lines provide a reliable probe of the density of gas in the range $10 < n_e$ [cm^{-3}] $< 10^3$, only densities as large as 100 cm^{-3} were detected in their lines-of-sight. Detections of [N II] which fall just outside, $(\ell, b) = (78.1132°, 0.0°)$, of the region we analyze resulted in an electron density of $n_e = (13.0 \pm 1.7)$ cm^{-3}. Furthermore, a lower limit of $n_e \gtrsim 30$ cm^{-3} was derived from [N II] in the DR 21 region (White et al. 2010). Additionally, Pineda et al. (2019) derive the electron densities of 21 discrete spectral components using [N II] 205 μm and radio recombination lines. They find an average electron density of 41 cm^{-3} with values ranging from 8 to 170 cm^{-3}. The electron densities we find are in remarkable agreement and indicate that what we observe in the Cygnus X region with low-frequency radio emission may point to the types of environments which are traced by these [N II] surveys.

This conclusion is also supported by the H II column densities that Goldsmith et al. (2015) derives. They find [N II] column densities of $N(\mathrm{N}^+) = (1 - 20) \times 10^{16}$ cm^{-3}. Using their estimated nitrogen fractional abundance $X(\mathrm{N}+) = 2.9 \times 10^{-4}$, their column densities would translate into ionized hydrogen column densities of $N(\mathrm{H}^+) = (3 - 70) \times 10^{19}$ cm^{-3}. As shown in Figure 2.10, the column densities of ionized gas found in the [N II] survey are consistent with the column densities of filaments of ionized gas in the Cygnus X region.

2.6.4 Connection to ELD ionized gas

Extended low-density (ELD) ionized gas (Mezger 1978) has characteristic densities of $n_e = (5 - 10)$ cm^{-3} and pathlengths of (50–200) pc. In the region we investigate, diffuse thermal emission of $EM = 5300$ pc cm^{-6} is seen across the $\mathcal{O}(100$ pc) region. In addition to filling the Cygnus X region, emission may also arise along the line of sight, as we are looking down a spiral arm. Thus considering path-lengths of 0.1 – 1 kpc, the volume filling electron density is estimated at $n_e \approx (2 - 7)$ cm^{-3}. We take $n_e \approx 5$ cm^{-3} as a representative value, and note this is accurate to within a factor of two. The mass of ionized gas in the $R = 50$ pc volume we analyze is then 6×10^4 M$_\odot$ (or 2×10^5 M$_\odot$ in a Strömgren volume). Ionized gas of density $n_e \sim 5$ cm^{-3} and cloud size 100 pc are consistent with the properties of ELD ionized gas. Tying the volume-

filling ionized gas in Cygnus X with ELD ionized gas is consistent with previous results which connect ELD ionized gas to envelopes of H II regions (Shaver 1976; Anantharamaiah 1986; McKee & Williams 1997) and those which suggest ELD gas predominantly arises from the most massive regions of star formation in our Galaxy (Murray & Rahman 2010).

The thermal pressure felt by the volume filling warm ionized gas, for an electron temperature of $T_e \approx 7400$ K and a fully ionized medium, is $P/k_{\rm B} = 2n_eT_e \sim 7.4 \times 10^4$ K cm^{-3}. While this pressure is a factor of 10 lower than in the ionized filaments ($P/k_{\rm B} \sim 6 \times 10^5$ K cm^{-3}) and the X-ray emitting plasma, it is still elevated compared with the diffuse ISM ($P/k_{\rm B} \sim 4 \times 10^3$ K cm^{-3}) (Jenkins et al. 2011). Overpressurization is consistent even for the globally elevated ISM pressure expected in spiral arms and at small galactic radii (Wolfire et al. 2003).

In the Cygnus X region, ELD ionized gas is filled and replenished by photoevaporating filaments (similar in properties to [N II] gas of (Goldsmith et al. 2015; Pineda et al. 2019)) eroding neutral clouds over the lifetime of the massive stars. As we calculated in Section 2.6.2, a characteristic filament in the region photoevaporates 1.2×10^4 M$_\odot$ of cloud mass over the lifetime of the massive stars. Five of the characteristic filaments supply an equal mass of ionized gas in the region analyzed, and ten filaments would replenish the full Strömgren volume.

2.6.5 Future LOFAR observations

Our results have demonstrated new capabilities provided by LOFAR to characterize low-density ionized gas at low radio frequencies and high spatial resolution. Our analysis indicates that observations making use of the full bandwidth of data for a typical LOFAR HBA observation would reach noise levels of (25 mJy beam^{-1})/$\sqrt{20}$ = 5 mJy beam^{-1} or free-free emission measures of 160 pc cm^{-6} at a resolution of $2'$, reaching similar EM depths as the CGPS 1.4 GHz observations. Future investigations in conjunction with LOFAR's LBA at 30–80 MHz will enable ionized gas (with typical ionized gas temperatures of 7000 K) to be characterized with free-free optical depths down to ∼400 pc cm^{-6}. With the LOFAR Survey (Shimwell et al. 2019), the Galactic Plane in the Northern hemisphere will be covered to a spatial resolution of up to $6''$. Our analysis, builds on previous LOFAR analyses (e.g., Arias et al. 2019), which show the power of LOFAR to characterize the continuum emission at low radio frequencies in the Galaxy.

2.7 Conclusions

Photoionized gas probes the influence of massive stars on their environment. The Cygnus X region ($d \sim 1.5$ kpc) is one of the most massive star forming complexes in our Galaxy, in which the Cyg OB2 association (age of 3–5 Myr and stellar mass of $M_\star \approx 2 \times 10^4$ M$_\odot$) has a dominant influence. We observed the Cygnus X region at 142 MHz using LOFAR and corrected for missing short-spacing information during image deconvolution. Together with archival data from the Canadian Galactic Plane Survey, we investigate the morphology, distribution, and physical conditions of low-

density ionized gas in a $4^\circ \times 4^\circ$ ($\sim$100 pc $\times$ 100 pc) region at a resolution of $2'$ (0.9 pc). As first discussed at radio frequencies by Wendker et al. (1991), the ionized gas in this region is characterized by filamentary structure. We use the `DisPerSE` and `FilChaP` packages to characterize the radial profiles of low-density ionized filaments. Our results are as follows:

- We have demonstrated a procedure for correcting for a lack of short-spacing information in LOFAR observations. We compare our results with feathering and find them to be consistent within 6% (on average) across the region analyzed, thanks to LOFAR's excellent sensitivity to large scale emission. Future LOFAR HBA observations, especially together with the LOFAR LBA at 30–70 MHz, will characterize the low-density ISM to deep emission measures and high spatial resolution.

- Radio continuum emission in the region is largely consistent with free-free thermal emission down to our LOFAR observing frequency of 142 MHz. This agrees well with previously analyses down to 408 MHz (Wendker et al. 1991; Xu et al. 2013).

- The low density ionized gas traced by the radio continuum shows a strong correspondence with warm dust traced by 24 μm emission.

- We fit free-free optical depths at 142 MHz in diffuse, extended (2° scale) emission with values of $\tau_{\rm ff} = (0.1 - 1.5)$, corresponding to emission measures of $EM = (2 - 40) \times 10^3$ pc cm^{-6}. While in four compact, unresolved regions of active star formation (DR 7, 15, 21, 22), we measure $\tau_{\rm ff} > 4$.

- We characterize 1934 EM radial profiles from 906 filaments. We find a power-law distribution in peak EM down to our completeness limit of 4200 pc cm^{-6}. A characteristic width of 3.6 pc arises in the distribution, well separated from our spatial resolution of 0.9 pc. The median electron density within the filaments is 38 cm^{-3} — quite similar to the ionized gas probed along the Galactic plane through [N II] (e.g., Goldsmith et al. 2015; Pineda et al. 2019). We derive thermal pressures within a median value of 6×10^5 K cm^{-3}, indicating the filaments are over-pressured compared with the neutral ISM.

- We construct an ionizing photon map of the Cyg OB2 association and compare it with the ionizing photon flux measured from the thermal radio continuum. We find that the ionizing photon flux from Cyg OB2 is sufficient to maintain ionization in 67% of the region and the filaments.

- We estimate that the (high-pressure) filaments are likely photoevaporating surfaces flowing into volume-filling warm ionized gas that is relatively low in pressure (with density $n_e \sim 5$ cm^{-3}). We notice a correlation between the ionization parameter, U, and electron density, indicating filaments are primarily influenced by stellar radiation. We estimate that typical photoevaporating surfaces influenced by Cyg OB2 each convert $\sim 2 \times 10^4$ M$_\odot$ of neutral material (or 10% of a typical $\sim 10^5$ M$_\odot$ molecular cloud) into ionized gas over the lifetime of the association.

- A minority of filaments do not have or are significantly displaced from neutral gas (as traced by PAH and CO emission). We estimate the influence of stellar winds from Cyg OB2 and find that ∼0.4% of the mechanical energy in the winds may be coupled to turbulent energy in the ionized gas. In which case, some of the ionized filaments may be transitory features resulting from dissipated turbulence.

- The volume-filling ionized gas in the Cygnus X region — $M_+ \sim 2 \times 10^5$ $M_\odot$ with $n_e \sim 5$ cm^{-3} and ∼100 pc pathlength — is consistent with properties of extended low-density (ELD) ionized gas (Mezger 1978) and connects to previous findings which attribute ELD ionized gas to envelopes of H II regions and those which suggest the most massive star forming regions likely dominate ELD emission. ELD ionized gas which leaks from this largely inhomogeneous region can be replenished in mass by ∼10 (typical) photoevaporating surfaces over the ~ 5 Myr lifetime of the OB stars.

Acknowledgements

We thank Nicola Schneider for sharing data products, and Frits Sweijen and Alex Mechev for maintaining software on the grid infrastructure and in Leiden. KLE thanks the Green Bank Observatory for the hospitality as hosting her as a (remote) guest during the completion of this work.

KLE, HJAR and AGGMT acknowledge financial support from the Dutch Science Organization (NWO) through TOP grant 614.001.351. GJW gratefully acknowledges the support of an Emeritus Fellowship from the Leverhulme Trust. AGGMT acknowledges support through the Spinoza premier of the NWO. Part of this work was carried out on the Dutch national e-infrastructure with the support of the SURF Cooperative through grant e-infra 160152.

This paper is based (in part) on results obtained with International LOFAR Telescope (ILT) equipment under project codes `LC0_032`. LOFAR (van Haarlem et al. 2013) is the Low Frequency Array designed and constructed by ASTRON. It has observing, data processing, and data storage facilities in several countries, that are owned by various parties (each with their own funding sources), and that are collectively operated by the ILT foundation under a joint scientific policy. The ILT resources have benefited from the following recent major funding sources: CNRS-INSU, Observatoire de Paris and Universite d'Orleans, France; BMBF, MIWF-NRW, MPG, Germany; Science Foundation Ireland (SFI), Department of Business, Enterprise and Innovation (DBEI), Ireland; NWO, The Netherlands; The Science and Technology Facilities Council, UK; Ministry of Science and Higher Education, Poland. The research presented in this paper has used data from the Canadian Galactic Plane Survey, a Canadian project with international partners, supported by the Natural Sciences and Engineering Research Council.

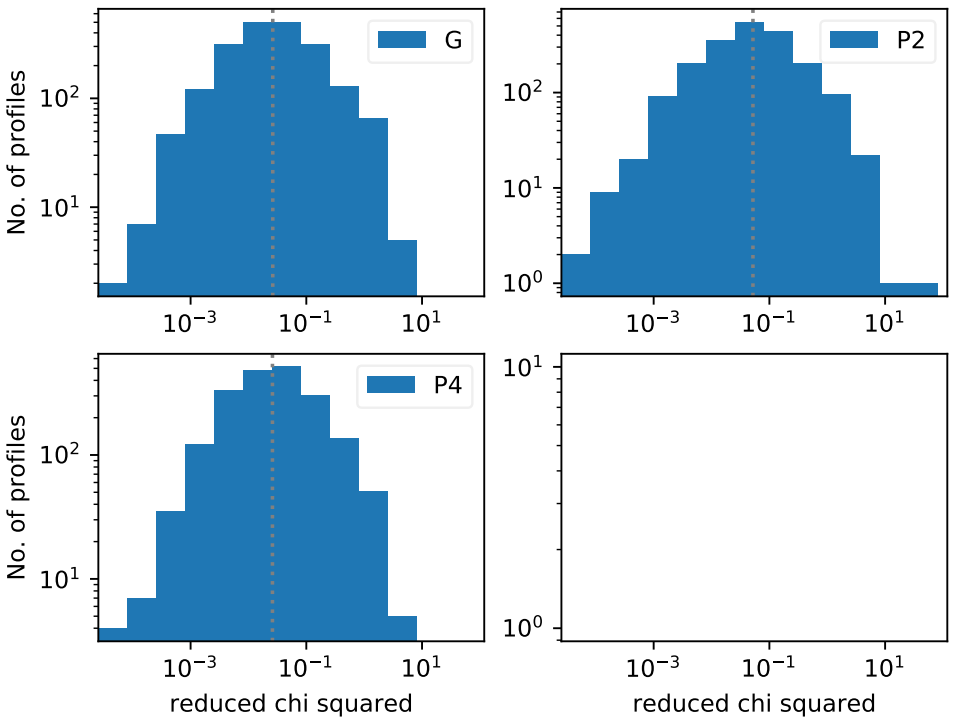


Figure 2.16: Reduced χ^2 distributions from mean radial profile fits. The four panels represent the distribution resulting from Gaussian fits (G; top left panel), Plummer $p = 2$ profile fits (P2; top right panel), and Plummer $p = 4$ profile fits (P4; bottom left panel). Broadly speaking the results between the fits are consistent, with Plummer $p = 4$ resulting in the lowest median value in the reduced χ^2 distribution.

2.A Additional fits to filament radial profile

We plot the histograms of the reduced χ^2 of fits to the filament radial profiles, for Gaussian and Plummer ($p = 2, 4$) profiles (Figure 2.16). We also show histograms with the results of those fits: the peak (Figure 2.17) and the FWHM width (Figure 2.18).

2.B Ionizing Photons from Cyg OB2

Using the procedure laid out in Tiwari et al., in preparation, we estimate the ionizing photon rate per unit area from early-type Cyg OB2 cluster members by synthesizing a list of stars from catalogs by Wright et al. (2015) and Berlanas et al. (2018, 2020), who in part used the catalog of Comerón & Pasquali (2012). All stars from these catalogs have effective temperatures and luminosities assigned from spectroscopic analyses. We take the assigned parameters from Berlanas et al. (2018, 2020) if possible, then from Wright et al. (2015), and finally Comerón & Pasquali (2012) via Berlanas et al.

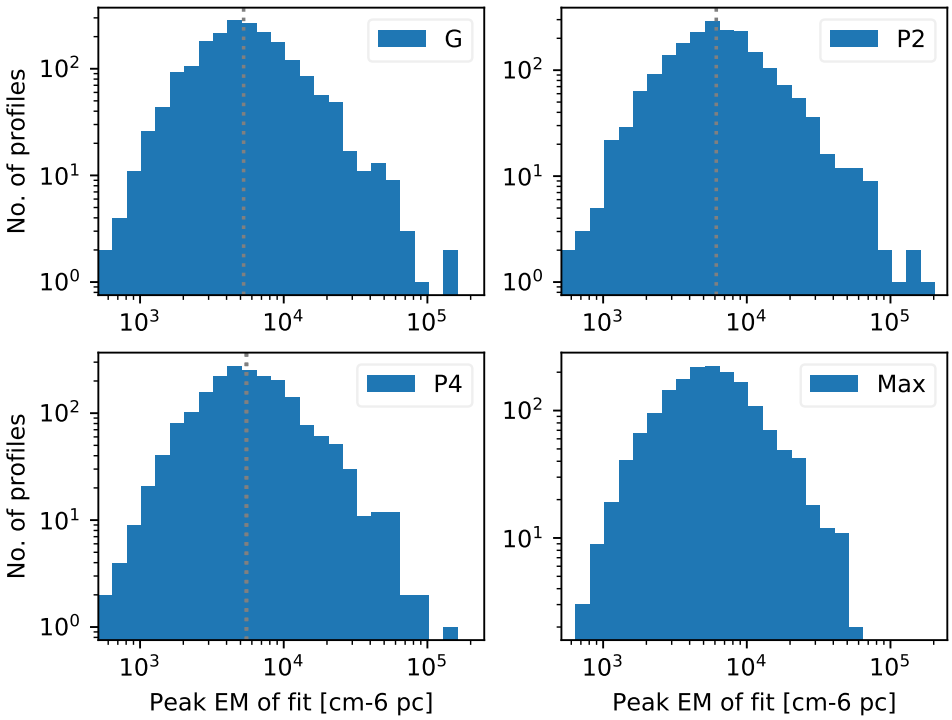


Figure 2.17: Peak EM distributions from mean radial profile fits. The four panels represent the distribution resulting from Gaussian fits (G; top left panel), Plummer $p = 2$ profile fits (P2; top right panel), Plummer $p = 4$ profile fits (P4; bottom left panel), and the maximum value of the profile (Max; bottom right panel). The median value of the EM distributions is robust to different determinations.

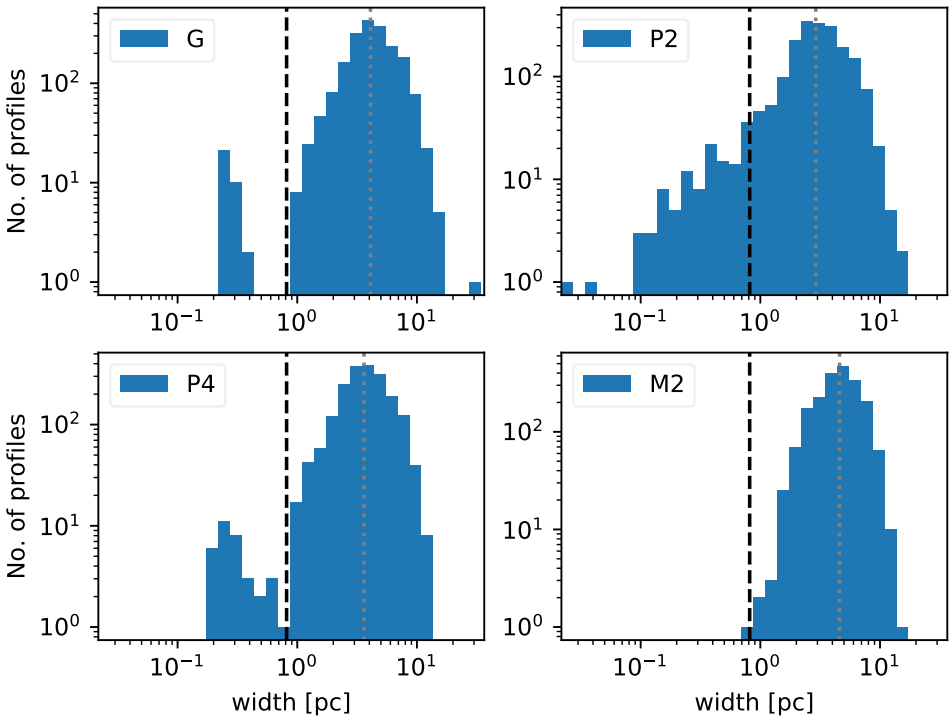


Figure 2.18: Width distributions from mean radial profile fits. The four panels represent the distribution resulting from Gaussian fits (G; top left panel), Plummer $p = 2$ profile fits (P2; top right panel), Plummer $p = 4$ profile fits (P4; bottom left panel), and the moment 2 computed width (M2; bottom right panel). A peak in the width distribution is seen across all four different width determinations.

(2018). With these effective temperatures and luminosities, we pick out models for each O or B star from the Potsdam Wolf-Rayet (PoWR) stellar atmosphere grids (Hamann & Gräfener 2004; Todt et al. 2015; Sander et al. 2015; Hainich et al. 2019). These models are gridded by effective temperature and gravity, so we first interpolate the gravity, $\log g$, from the grid's associated temperature and luminosity values. From the synthetic spectra provided by the PoWR models, we integrate the total ionizing flux above 13.6 eV. We then use the stellar coordinates and the ionizing photon flux to estimate the ionizing photon rate per unit area at any location around the H II region using projected distances at 1.5 kpc and summing over the flux from all stars.

3 | Super Star Clusters in the Central Starburst of NGC 4945

Abstract

NGC 4945 is a nearby (3.8 Mpc) galaxy hosting a nuclear starburst and Seyfert Type 2 AGN. We use the Atacama Large Millimeter/submillimeter Array (ALMA) to image the 93 GHz (3.2 mm) free-free continuum and hydrogen recombination line emission (H40α and H42α) at 2.2 pc (0.12″) resolution. Our observations reveal 27 bright, compact sources with FWHM sizes of 1.4–4.0 pc, which we identify as candidate super star clusters. Recombination line emission, tracing the ionizing photon rate of the candidate clusters, is detected in 15 sources, 6 of which have a significant synchrotron component to the 93 GHz continuum. Adopting an age of $\sim$5 Myr, the stellar masses implied by the ionizing photon luminosities are $\log_{10}(M_\star/\mathrm{M}_\odot) \approx 4.7$–6.1. We fit a slope to the cluster mass distribution and find $\beta = -1.8 \pm 0.4$. The gas masses associated with these clusters, derived from the dust continuum at 350 GHz, are typically an order of magnitude lower than the stellar mass. These candidate clusters appear to have already converted a large fraction of their dense natal material into stars and, given their small free-fall times of $\sim$0.05 Myr, are surviving an early volatile phase. We identify a point-like source in 93 GHz continuum emission which is presumed to be the AGN. We do not detect recombination line emission from the AGN and place an upper limit on the ionizing photons which leak into the starburst region of $Q_0 < 10^{52}$ s^{-1}.

K. L. Emig, A. D. Bolatto, A. K. Leroy, E. A. C. Mills, M. J. Jiménez Donaire, A. G. G. M. Tielens, A. Ginsburg, M. Gorski, N. Krieger, R. C. Levy, D. S. Meier, J. Ott, E. Rosolowsky, T. A. Thompson, S. Veilleux
The Astrophysical Journal, 903, 50 (2020)

3.1 Introduction

Many stars form in clustered environments (Lada & Lada 2003; Kruijssen 2012). Bursts of star formation with high gas surface density produce massive ($> 10^5$ $M_\odot$), compact (FWHM size of 2-3 pc; Ryon et al. 2017) clusters, referred to as super star clusters. Super star clusters likely have high star-formation efficiencies (Goddard et al. 2010; Ryon et al. 2014; Adamo et al. 2011, 2015; Chandar et al. 2017; Johnson et al. 2016; Ginsburg & Kruijssen 2018). They may represent a dominant output of star-formation during the peak epoch of star-formation ($z \sim 1 - 3$; Madau & Dickinson 2014). The process by which these massive clusters form now may also relate to the origin of globular clusters.

The earliest stages of cluster formation are the most volatile and currently, unconstrained (Dale et al. 2015; Ginsburg et al. 2016; Krause et al. 2016; Li et al. 2019; Krause et al. 2020). Characterizing properties of young (<10 Myr) clusters is a key step towards understanding their formation, identifying the dominant feedback processes at each stage of cluster evolution, determining which clusters survive as gravitationally bound objects, and linking all of these processes to the galactic environment.

While young clusters of mass $\sim 10^4$ $M_\odot$ are found within our Galaxy (Bressert et al. 2012; Longmore et al. 2014; Ginsburg et al. 2018), the most massive, young clusters in the local universe are often found in starbursting regions and merging galaxies (e.g., Zhang & Fall 1999; Whitmore et al. 2010; Linden et al. 2017). Direct optical and even near-infrared observations of forming clusters are complicated by large amounts of extinction. Analyses of optically thin free-free emission and long wavelength hydrogen recombination lines of star clusters offer an alternative, extinction-free probe of the ionizing gas surrounding young star clusters (Condon 1992; Roelfsema & Goss 1992; Murphy et al. 2018). However, achieving a spatial resolution matched to the size of young clusters $\mathcal{O}(1$ pc) (Ryon et al. 2017) in galaxies at the necessary frequencies and sensitivities has only recently become possible thanks to the Atacama Large Millimeter/submillimeter Array (ALMA).

We have recently analyzed forming super star clusters in the central starburst of the nearby (3.5 Mpc) galaxy NGC 253 at $\sim$2 pc resolution (Leroy et al. 2018; Mills et al. 2020). NGC 4945 is the second object we target in a campaign to characterize massive star clusters in local starbursts with ALMA.

NGC 4945 is unique in that it is one of the closest galaxies (3.8 ± 0.3 Mpc; Karachentsev et al. 2007) where a detected AGN and central starburst coexist. In the central $\sim$200 pc, the starburst dominates the infrared luminosity and ionizing radiation (Spoon et al. 2000; Marconi et al. 2000), and an outflow of warm ionized gas has been observed (Heckman et al. 1990; Moorwood et al. 1996; Mingozzi et al. 2019). Individual star clusters have not previously been observed in NGC 4945, due to the high extinction at visible and short IR wavelengths (e.g., $A_V \gtrsim 36$ mag; Spoon et al. 2000). Evidence for a Seyfert AGN comes from strong, variable X-ray emission, as NGC 4945 is one of the brightest sources in the X-ray sky and has a Compton thick column density of 3.8×10^{24} cm^{-2} (Marchesi et al. 2018). A kinematic analysis of H_2O maser emission yields a black hole mass of 1.4×10^6 $M_\odot$ (Greenhill et al. 1997).

In this article, we use ALMA to image the 93 GHz free-free continuum and hy-

drogen recombination line emission (H40α and H42α) at 2.2 pc (0.12″) resolution. This emission allows us to probe photo-ionized gas on star cluster scales and thereby trace ionizing photon luminosities. We identify candidate star clusters and estimate properties relating to their size, ionizing photon luminosity, stellar mass, and gas mass.

Throughout this article, we plot spectra in velocity units with respect to a systemic velocity of v_{sys} = 580 km s^{-1} in the local standard of rest frame; estimates of the systemic velocity vary by ±25 km s^{-1} (e.g., Henkel et al. 2018; Chou et al. 2007; Roy et al. 2010). At the distance of 3.8 Mpc, 0.1″ corresponds to 1.84 pc.

3.2 Observations

We used the ALMA Band 3 receivers to observe NGC 4945 as part of the project 2018.1.01236.S (PI: A. Leroy). We observed NGC 4945 with the main 12 m array telescopes in intermediate and extended configurations. Four spectral windows in Band 3 – centered at 86.2, 88.4, 98.4, and 100.1 GHz – capture the millimeter continuum primarily from free-free emission and cover the hydrogen recombination lines of principal quantum number (to the lower state) $\mathsf{n} = 40$ and $\mathsf{n} = 42$ from the α ($\Delta\mathsf{n} = 1$) transitions. The rest frequency of H40α is 99.0230 GHz and of H42α is 85.6884 GHz.

In this article, we focus on the 93 GHz ($\lambda \sim 3.2$ mm) continuum emission and the recombination line emission arising from compact sources in the starbursting region. We image the data from an 8 km extended configuration, which are sensitive to spatial scales of 0.07″–6″ (2–100 pc), in order to focus on the compact structures associated with candidate clusters. We analyze the observatory-provided calibrated visibilities using version 5.4.0 of the Common Astronomy Software Application (CASA; McMullin et al. 2007).

When imaging the continuum, we flag channels with strong spectral lines. Then we create a continuum image using the full bandwidth of the line-free channels. We also make continuum images for each spectral window. For all images, we use Briggs weighting with a robust parameter of $r = 0.5$.

When imaging the two spectral lines of interest, we first subtract the continuum in uv space through a first order polynomial fit. Then, we image by applying a CLEAN mask (to all channels) derived from the full-bandwidth continuum image. Again, we use Briggs weighting with a robust parameter of $r = 0.5$, which represents a good compromise between resolution and surface brightness sensitivity.

After imaging, we convolved the continuum and line images to convert from an elliptical to a round beam shape. For the full-bandwidth continuum image presented in this article, the fiducial frequency is $\nu = 93.2$ GHz and the final full-width half maximum (FWHM) beam size is $\theta = 0.12''$. The rms noise away from the source is ≈0.017 mJy beam^{-1}, equivalent to 0.2 K in Rayleigh-Jeans brightness temperature units. Before convolution to a round beam, the beam had a major and minor FWHM of 0.097″ × 0.071″.

For the H40α and H42α spectral cubes, the final FWHM beam size is 0.20″, convolved from 0.097″ × 0.072″ and from 0.11″ × 0.083″, respectively. The slightly

lower resolution resulted in more sources with significantly detected line emission. We boxcar smoothed the spectral cubes from the native 0.488 MHz channel width to 2.93 MHz. The typical rms in the H40α cube is 0.50 mJy beam^{-1} per 8.9 km s^{-1} channel. The typical rms in the H42α cube is 0.48 mJy beam^{-1} per 10.3 km s^{-1} channel.

As part of the analysis, we compare the high resolution data with observations taken in a 1 km intermediate configuration as part of the same observing project. We use the intermediate configuration data to trace the total recombination line emission of the starburst. We use the continuum image provided by the observatory pipeline, which we convolve to have a circular beam FWHM of 0.7″; the rms noise in the full-bandwidth image is 0.15 mJy beam^{-1}. The spectral cubes have typical rms per channel of 0.24 mJy beam^{-1} with the same channel widths as the extended configuration cubes. We do not jointly image the configurations because our main science goals are focused on compact, point-like objects. The extended configuration data on their own are well suited to study these objects and any spatial filtering of extended emission will not affect the analysis.

We compare the continuum emission at 3 mm with archival ALMA imaging of the $\nu = 350$ GHz ($\lambda \sim 850$ μm) continuum (project 2016.1.01135.S, PI: N. Nagar). At this frequency, dust emission dominates the continuum. We imaged the calibrated visibilities with a Briggs robust parameter of $r = -2$ (towards uniform weighting), up-weighting the extended baselines to produce a higher resolution image, suitable for comparison to our new Band 3 data. We then convolve the images to produce a circularized beam, resulting in a FWHM resolution of 0.12″ (from an initial beam size of 0.10″ × 0.064″), exactly matched to our 93 GHz continuum image. These data have an rms noise of 0.7 mJy beam^{-1} (0.2 K).

We compare the ALMA data with Australian Long Baseline Array (LBA) imaging of $\nu = 2.3$ GHz continuum emission (Lenc & Tingay 2009). At this frequency and resolution, the radio continuum is predominantly synchrotron emission. We use the Epoch 2 images (courtesy of E. Lenc) that have a native angular resolution slightly higher than the 3 mm ALMA data, with a beam FWHM of 0.080″ × 0.032″ and an rms noise of 0.082 mJy beam^{-1}.

3.3 Continuum Emission

The whole disk of NGC 4945, as traced by *Spitzer* IRAC 8 μm emission (Program 40410, PI: G. Rieke), is shown in Figure 3.1. The 8 μm emission predominantly arises from UV-heated polycyclic aromatic hydrocarbons (PAHs), thus tracing the interstellar medium and areas of active star formation. The square box indicates the 8″ × 8″ (150 pc × 150 pc) starburst region that is of interest in this article.

Figure 3.2 shows the 93 GHz ($\lambda \sim 3$ mm) continuum emission in the central starburst of NGC 4945. Our image reveals ∼30 peaks of compact, localized emission with peak flux densities 0.6–8 mJy (see Section 3.3.1). On average, the continuum emission from NGC 4945 at this frequency is dominated by thermal, free-free (bremsstrahlung) radiation (Bendo et al. 2016). Free-free emission from bright, compact regions may trace photo-ionized gas in the immediate surroundings of massive stars. We take into consideration the point-like sources detected at 93 GHz as candidate massive star

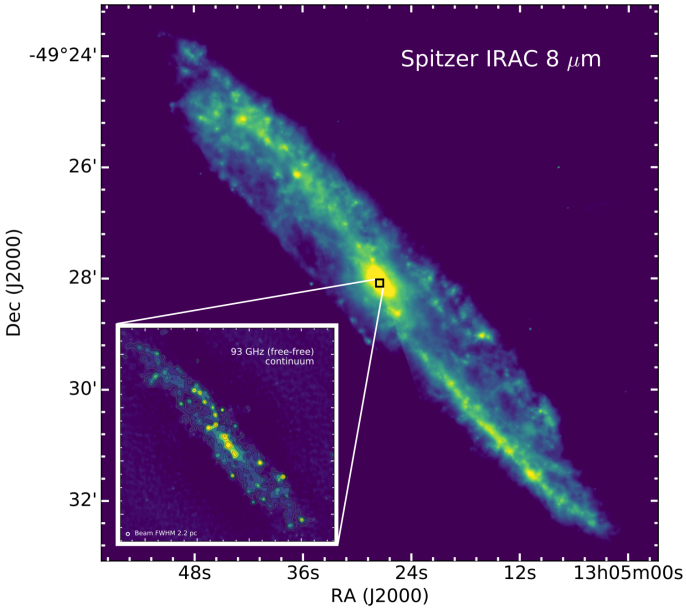


Figure 3.1: *Spitzer* IRAC 8 μm emission from UV-heated PAHs over the full galactic disk of NGC 4945. The black square box indicates the $8'' \times 8''$ (150 pc $\times$ 150 pc) central starburst region of interest in this article; the inset shows the ALMA 93 GHz continuum emission.

clusters, though some contamination by synchrotron-dominated supernova remnants or dusty protoclusters may still be possible. The morphology of the 93 GHz emission and clustering of the peaks indicate possible ridges of star formation and shells. The extended, faint negative bowls flanking the main disk likely reflect the short spacing data missing from this image. We do not expect that they affect our analysis of the point source-like cluster candidates.

The large amount of extinction present in this high inclination central region ($i \sim 72°$; Henkel et al. 2018) has previously impeded the direct observation of its star clusters. Paschen-α (Pa-α) emission (Marconi et al. 2000) of the $\mathsf{n} = 3$ hydrogen recombination line at 1.87 μm, shown in Figure 3.3, reveals faint emission above and below the star-forming plane. Corrected for extinction, the clumps of ionized emission traced by Pa-α would give rise to free-free emission below our ALMA detection limit. Pa-α along with mid-infrared spectral lines give support for dust extinction of $A_V >$ 160 mag surrounding the AGN core and more generally $A_V \gtrsim 36$ mag in the star-forming region (Spoon et al. 2000).

A large fraction (18/29) of the 93 GHz sources coincide with peaks in dust emission at 350 GHz, as shown in Figure 3.3. Overall there is a good correspondence between the two tracers. This indicates that candidate clusters are relatively young and may still harbor reservoirs of gas, though in Section 3.5.7 we find that the fraction of the mass still in gas tends to be relatively small.

In Figure 3.3, the 93 GHz peaks without dust counterparts tend to be strong sources of emission at 2.3 GHz (Lenc & Tingay 2009), a frequency where synchrotron

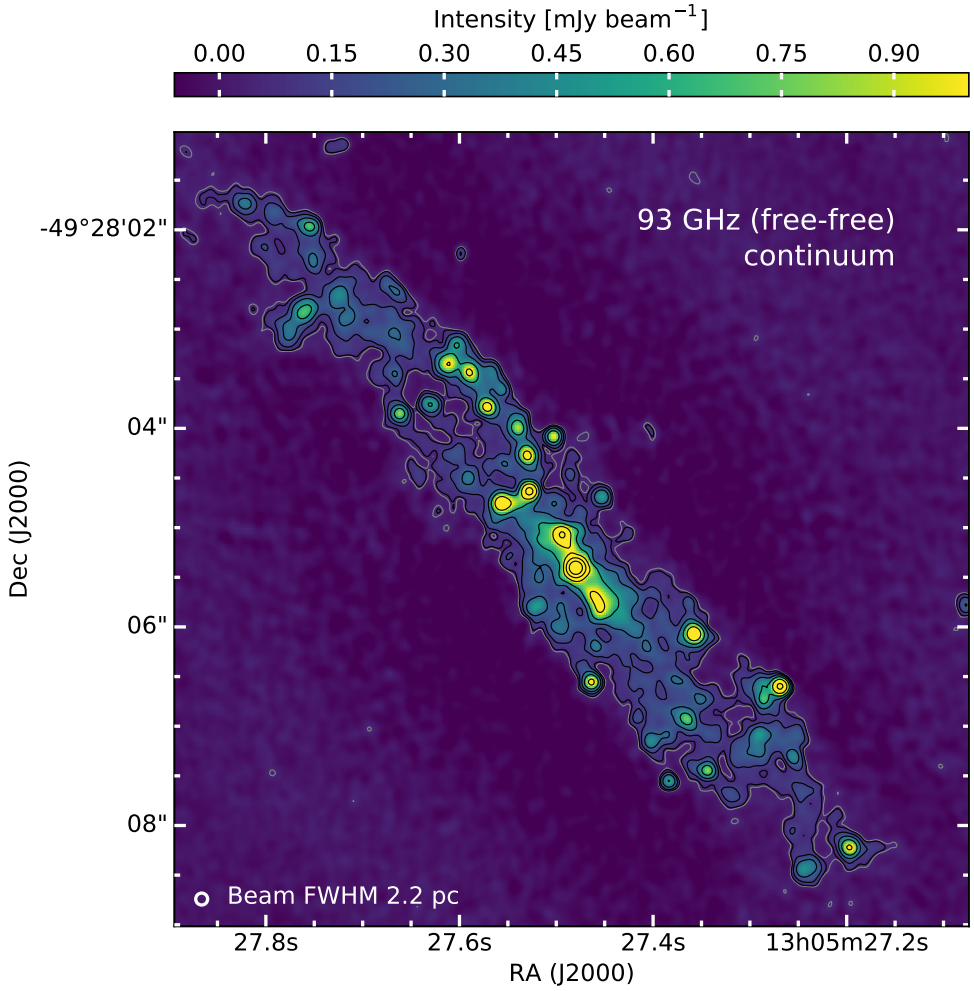


Figure 3.2: ALMA 93 GHz ($\lambda \sim 3.2$ mm) continuum emission in the central starburst of NGC 4945. The continuum at this frequency is dominated by ionized, free-free emitting plasma. In this paper, we show that the point-like sources are primarily candidate, massive star clusters. The brightest point source of emission at the center is presumably the Seyfert AGN. The rms noise away from the source is $\sigma \approx 0.017$ mJy beam^{-1} and the circularized beam FWHM is $0.12''$ (or 2.2 pc at the distance of NGC 4945). Contours of the continuum image show 3σ emission (gray) and $[4, 8, 16, ...256]\sigma$ emission (black).

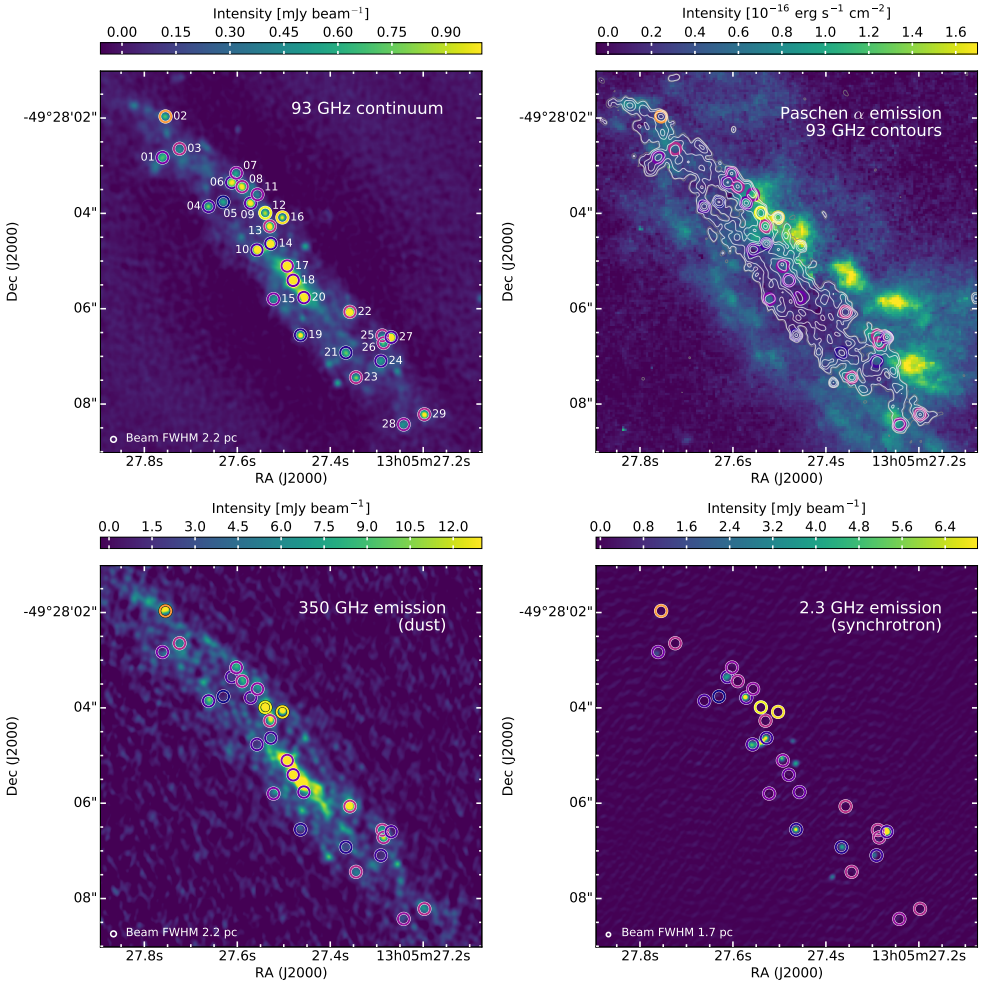


Figure 3.3: *Top left*: 93 GHz continuum emission with sources identified, also see Table 3.1. Circles show apertures (diameter of 0.24″) used for continuum extraction. Their colors indicate the measured in-band spectral index, as in Figure 3.4 where dark purple indicates synchrotron dominated emission and yellow indicates dust dominated emission. *Top right*: HST Paschen-α emission – hydrogen recombination line, $\mathsf{n} = 3$, at 1.87 μm – (courtesy P. van der Werf) tracing ionized gas at $\approx 0.2''$ resolution (Marconi et al. 2000). Dust extinction, of $A_V > 36$ mag, obscures the Pa-α recombination emission at shorter wavelengths from the starburst region. Contours trace 93 GHz continuum, as described in Figure 3.2. *Bottom left*: ALMA 350 GHz continuum emission tracing dust. *Bottom right*: Australian LBA 2.3 GHz continuum imaging of synchrotron emission primarily from supernova remnants (Lenc & Tingay 2009).

emission typically dominates. As discussed in Lenc & Tingay (2009), the sources at this frequency are predominantly supernova remnants. The presence of 13 possible supernova remnants – four of which are resolved into shell-like structures with 1.1 to 2.1 pc in diameter – indicates that a burst of star-formation activity started at least a few Myr ago. Lenc & Tingay (2009) modeled SEDs of the sources spanning 2.3–23 GHz and found significant opacity at 2.3 GHz ($\tau = 5 - 22$), implying the presence of dense, free-free plasma in the vicinity of the supernova remnants.

At 93 GHz, the very center of the starburst shows an elongated region of enhanced emission (about 20 pc in projected length, or $\sim 1''$) that is also bright in 350 GHz emission. This region is connected to the areas of highest extinction. Higher column densities of ionized plasma are also present in the region; Lenc & Tingay (2009) observations reveal large free-free opacities at least up to 23 GHz. The brightest peak at 93 GHz, centered at $(\alpha, \delta)_{93} = (13\mathrm{h}\,05\mathrm{m}\,27.4798\mathrm{s} \pm 0.004\mathrm{s}, -49^\circ\,28'\,05.404'' \pm 0.06'')$, is co-located with the kinematic center as determined from H_2O maser observations $(\alpha, \delta)_{H_2O} = (13\mathrm{h}\,05\mathrm{m}\,27.279\mathrm{s} \pm 0.02\mathrm{s}, -49^\circ\,28'\,04.44'' \pm 0.1'')$ (Greenhill et al. 1997) and presumably harbors the AGN core. We refer to the elongated region of enhanced emission surrounding the AGN core as the circumnuclear disk. The morphological similarities between 93 GHz and 350 GHz, together with the detection of a synchrotron point source (likely supernova remnant; see Section 3.3.2 and Source 17) in the circumnuclear disk, indicate star-formation is likely present there.

3.3.1 Point Source Identification

We identify candidate star clusters via point-like sources of emission in the 93 GHz continuum image. Sources are found using `PyBDSF` (Mohan & Rafferty 2015) in the following way. Islands are defined as contiguous pixels (of nine pixels or more) above a threshold of seven times the global rms value of $\sigma \approx 0.017$ mJy beam^{-1}. Within each island, multiple Gaussians may be fit, each with a peak amplitude greater than the peak threshold, a threshold of ten times the global rms. We chose this peak threshold to ensure that significant emission can also be identified in the continuum images made from individual spectral windows. The number of Gaussians is determined from the number of distinct peaks of emission higher than the peak threshold and which have a negative gradient in all eight evaluated directions. Starting with the brightest peak, Gaussians are fit and cleaned (i.e., subtracted). A source is identified with a Gaussian as long as subtracting its fit does not increase the island rms.

Applying this algorithm to our 93 GHz image yielded 50 Gaussian sources. We remove 5 sources that fall outside of the star-forming region. We also removed 3 sources that appeared blended, with an offset $< 0.12''$ from another source. Finally we remove 13 sources that do not have a flux density above ten times the global rms after extracting the 93 GHz continuum flux density through aperture photometry (see Section 3.3.2). As a result, we analyze 29 sources as candidate star clusters. In Figure 3.3, we show the location of each source with the apertures used for flux extraction. The sources match well with what we would identify by eye.

3.3.2 Point Source Flux Extraction

For each source, we extract the continuum flux density at 2.3 GHz, 93 GHz, and 350 GHz through aperture photometry. Before extracting the continuum flux at 2.3 GHz, we convolve the image to the common resolution of $0.12''$. We extract the flux density at the location of the peak source within an aperture diameter of $0.24''$. Then we subtract the extended background continuum that is local to the source by taking the median flux density within an annulus of inner diameter $0.24''$ and outer diameter $0.30''$; using the median suppresses the influence of nearby peaks and the bright surrounding filamentary features. The flux density of each source at each frequency is listed in Table 3.1. When the extracted flux density within an aperture is less than three times the global rms noise (in the 2.3 GHz and 350 GHz images), we assign a three sigma upper limit to that flux measurement.

In Figure 3.4, we plot the ratio of the flux densities extracted at 350 GHz and 93 GHz (S_{350}/S_{93}) against the ratio of the flux densities extracted at 2.3 GHz and 93 GHz ($S_{2.3}/S_{93}$). Synchrotron dominated sources, which fall to the bottom right of the plot, separate from the free-free (and dust) dominated sources, which lie in the middle of the plot. One exception is the AGN (Source 18) which, due to self-absorption at frequencies greater than 23 GHz, is bright at 93 GHz but not at 2.3 GHz and therefore has the lowest $S_{2.3}/S_{93}$ ratio.

From the continuum measurements we construct simple SEDs for each source. These SEDs are used for illustrative purposes and do not affect the analysis in this paper. Examples of the SEDs of three sources are included in Figure 3.5. We show an example of a free-free dominated source (Source 22), which represents the majority of sources, as well as a dust (Source 12) and a synchrotron (Source 14) dominated source. Of the sources with extracted emission of $>3\sigma$ at 2.3 GHz, nine also have the free-free absorption of their synchrotron spectrum modeled. We plot this information whenever possible. When a source identified by Lenc & Tingay (2009) lies within $0.06''$ (half the beam FWHM) of the 93 GHz source, we associate the low-frequency modeling with the 93 GHz source. We take the model fit by Lenc & Tingay (2009) and normalize it to the 2.3 GHz flux that we extract — as an example, see the solid purple curve in the middle panel of Figure 3.5. The SEDs of all sources are shown in Figure 3.13 in Appendix 3.C.

Table 3.1: Properties of the continuum emission from candidate star clusters.

Source	RA	Dec	S_{93} (mJy)	α_{93}	$S_{2.3}$ [a] (mJy)	S_{350} [b] (mJy)	$f_{\rm ff}$	$f_{\rm syn}$ [c]	$f_{\rm d}$ [c]
01	13:05:27.761	−49:28:02.83	1.28 ± 0.13	-0.80 ± 0.11	2.3 ± 0.4	...	0.51 ± 0.08	0.49	...
02	13:05:27.755	−49:28:01.97	1.01 ± 0.10	0.80 ± 0.22	...	26.6	0.78 ± 0.05	...	0.22
03	13:05:27.724	−49:28:02.64	0.82 ± 0.08	-0.27 ± 0.23	...	14.8	0.89 ± 0.16	0.11	...
04	13:05:27.662	−49:28:03.85	0.95 ± 0.09	-1.12 ± 0.42	...	16.0	0.27 ± 0.31	0.73	...
05	13:05:27.630	−49:28:03.76	0.77 ± 0.08	-1.57 ± 0.63	...	...	0.00 ± 0.40	1.00	...
06	13:05:27.612	−49:28:03.35	1.73 ± 0.17	-1.11 ± 0.22	7.1 ± 0.7	...	0.28 ± 0.16	0.72	...
07	13:05:27.602	−49:28:03.15	0.98 ± 0.10	-0.61 ± 0.19	...	15.5	0.65 ± 0.14	0.35	...
08	13:05:27.590	−49:28:03.44	1.89 ± 0.19	-0.40 ± 0.20	...	14.9	0.80 ± 0.15	0.20	...
09	13:05:27.571	−49:28:03.78	1.76 ± 0.18	-1.01 ± 0.22	12.2 ± 1.2	...	0.36 ± 0.16	0.64	...
10	13:05:27.558	−49:28:04.77	2.41 ± 0.24	-1.10 ± 0.20	9.5 ± 0.9	...	0.29 ± 0.14	0.71	...
11	13:05:27.557	−49:28:03.60	0.76 ± 0.08	-0.59 ± 0.38	...	13.9	0.66 ± 0.28	0.34	...
12	13:05:27.540	−49:28:03.99	1.28 ± 0.13	1.47 ± 0.37	...	37.4	0.62 ± 0.09	...	0.38
13	13:05:27.530	−49:28:04.27	1.78 ± 0.18	-0.22 ± 0.15	...	20.8	0.93 ± 0.11	0.07	...
14	13:05:27.528	−49:28:04.63	3.06 ± 0.31	-1.22 ± 0.19	14.3 ± 1.4	11.2	0.20 ± 0.14	0.80	...
15	13:05:27.522	−49:28:05.80	0.83 ± 0.08	-0.71 ± 0.28	...	...	0.57 ± 0.21	0.42	...
16	13:05:27.503	−49:28:04.08	0.98 ± 0.10	1.34 ± 0.37	...	23.9	0.65 ± 0.09	...	0.35
17	13:05:27.493	−49:28:05.10	3.69 ± 0.37	-0.61 ± 0.13	5.3 ± 0.5	47.1	0.65 ± 0.10	0.35	...
18	13:05:27.480	−49:28:05.40	9.74 ± 0.97	-0.85 ± 0.05	...	39.0	0.47 ± 0.03	0.53	...
19	13:05:27.464	−49:28:06.55	1.30 ± 0.13	-1.36 ± 0.38	7.8 ± 0.8	...	0.10 ± 0.28	0.90	...
20	13:05:27.457	−49:28:05.76	2.51 ± 0.25	-0.88 ± 0.13	...	22.4	0.45 ± 0.09	0.55	...
21	13:05:27.366	−49:28:06.92	1.19 ± 0.12	-1.38 ± 0.41	4.7 ± 0.5	...	0.09 ± 0.30	0.91	...
22	13:05:27.358	−49:28:06.07	2.57 ± 0.26	-0.19 ± 0.18	...	34.8	0.95 ± 0.13	0.05	...
23	13:05:27.345	−49:28:07.44	0.95 ± 0.09	-0.28 ± 0.25	...	9.7	0.88 ± 0.18	0.12	...
24	13:05:27.291	−49:28:07.10	0.90 ± 0.09	-1.22 ± 0.36	3.2 ± 0.4	...	0.20 ± 0.26	0.80	...
25	13:05:27.288	−49:28:06.56	0.91 ± 0.09	-0.27 ± 0.35	2.1 ± 0.4	15.2	0.89 ± 0.26	0.11	...
26	13:05:27.285	−49:28:06.72	1.13 ± 0.11	-0.33 ± 0.38	1.4 ± 0.4	18.5	0.85 ± 0.28	0.15	...
27	13:05:27.269	−49:28:06.60	2.75 ± 0.28	-1.08 ± 0.14	31.2 ± 3.1	...	0.30 ± 0.10	0.70	...

28	13:05:27.242	−49:28:08.43	0.80 ± 0.08	-0.54 ± 0.12	...	...	0.70 ± 0.09	0.30	...
29	13:05:27.198	−49:28:08.22	1.43 ± 0.14	-0.29 ± 0.20	...	12.6	0.88 ± 0.14	0.12	...

RA and Dec refer to the center location of a Gaussian source identified in the 93 GHz continuum image, in units of hour angle and degrees, respectively. S_{93} is the flux density in the 93 GHz full bandwidth continuum image. α_{93} is the spectral index at 93 GHz ($S \propto \nu^{\alpha}$), as determined from the best fit slope to the 85–101 GHz continuum emission. $S_{2.3}$ is the flux density extracted in the 2.3 GHz continuum image. S_{350} is the flux density extracted in the 350 GHz continuum image. $f_{\rm ff}$, $f_{\rm syn}$, $f_{\rm d}$ are the free-free, synchrotron, and dust fractional contribution to the 93 GHz continuum emission, respectively, as determined from the spectral index. See Section 3.3.3 for a description of how the errors on these fractional estimates are determined.

[a] A 3σ upper limit to the sources undetected in 2.3 GHz continuum emission is 1.2 mJy.

[b] The error on the 350 GHz flux density measurement is 3.2 mJy. A 3σ upper limit to the undetected sources is 9.6 mJy.

[c] The error on the fractional contributions are the same as for $f_{\rm ff}$ unless otherwise noted.

3.3.3 Free-free Fraction at 93 GHz

The flux density of optically thin, free-free emission at millimeter wavelengths (see Appendix 3.A.2; Draine 2011) arises as

$$S_{\rm ff} = (2.08\ {\rm mJy}) \left(\frac{n_e n_+ V}{5\times10^8\ {\rm cm^{-6}\ pc^3}}\right) \left(\frac{T_e}{10^4\ {\rm K}}\right)^{-0.32} \times \left(\frac{\nu}{100\ {\rm GHz}}\right)^{-0.12} \left(\frac{D}{3.8\ {\rm Mpc}}\right)^{-2} \tag{3.1}$$

where $EM_{\rm C} = n_e n_+ V$ is the volumetric emission measure of the ionized gas, D is the distance to the source, and T_e is the electron temperature of the medium. Properties of massive star clusters (i.e., ionizing photon rate) can thus be derived through an accurate measurement of the free-free flux density and an inference of the volumetric emission measure. We will determine a free-free fraction, $f_{\rm ff}$, and let $S_{\rm ff} = f_{\rm ff} S_{93}$.

In this section we focus on determining the portion of free-free emission that is present in the candidate stars clusters at 93 GHz. To do this, we need to estimate and remove contributions from synchrotron emission and dust continuum. We determine an in-band spectral index across the 15 GHz bandwidth of the ALMA Band 3 observations. Using the spectral index[1], we constrain the free-free fraction as well as the fractional contributions of synchrotron and dust (see Figure 3.4 and Table 3.1). We found the in-band index to give stronger constraints and, for 2.3 GHz, to be more reliable than extrapolating (assuming indices of -0.8,-1.5) because of the two decades difference in frequency coupled with the large optical depths already present at 2.3 GHz.

Band 3 Spectral Index

We estimate an in-band spectral index at 93 GHz, α_{93}, from a fit to the flux densities in the spectral window continuum images of the ALMA Band 3 data. The spectral windows span 15 GHz, which we set up to have a large fractional bandwidth. We extract the continuum flux density of each source in each spectral window using aperture photometry with the same aperture sizes as described in Section 3.3.2. We fit a first order polynomial to the five continuum measurements – four from each spectral window, one from the full bandwidth image. The fit to the in-band spectral index is listed in Table 3.1 with the one sigma uncertainty to the fit. The median uncertainty of the spectral indices is 0.13. The spectral indices measured in this way were consistent for the brightest sources with the index determined by `CASA tclean`; however, we found our method to be more reliable for the fainter sources. Nonetheless, the errors are large for faint sources.

Decomposing the fractional contributions of emission type

From the in-band spectral index fit, we estimate the fractional contribution of each emission mechanism — free-free, synchrotron, and (thermal) dust — to the 93 GHz

[1]similar methods have been used by e.g., Linden et al. (2020)

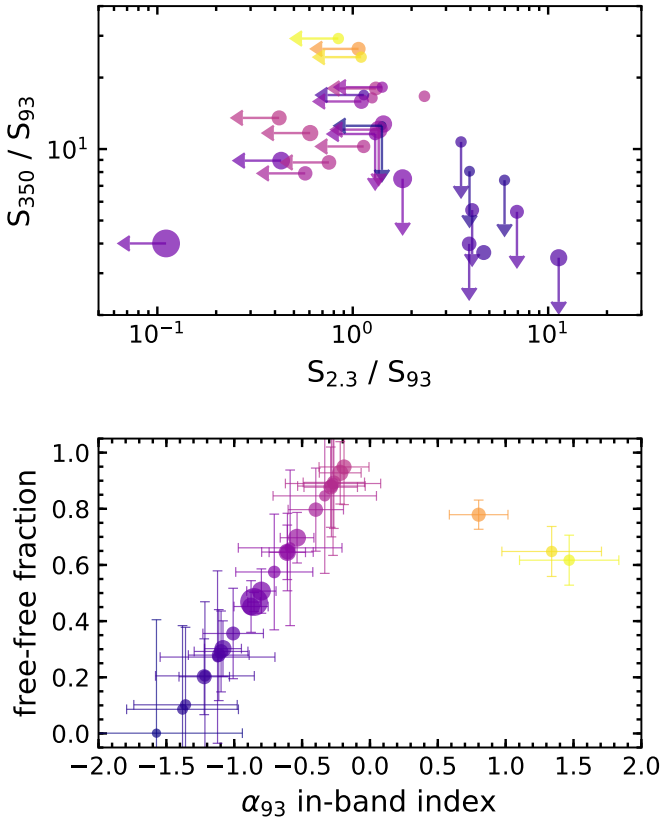


Figure 3.4: *Top:* The ratio of the flux densities extracted at 350 GHz and 93 GHz (S_{350}/S_{93}) plotted against the ratio of the flux densities extracted at 2.3 GHz and 93 GHz ($S_{2.3}/S_{93}$), and *bottom:* the relation we use to determine the free-free fraction from the in-band index, α_{93}, at 93 GHz. Data points in both plots are colored by the in-band index derived only from our ALMA data. Yellow indicates dust dominated sources, whereas purple indicates synchrotron dominated sources. Candidate star clusters in which free-free emission dominates at 93 GHz appear ∼pink. The diameter of each data point is proportional to the flux density at 93 GHz and correspondingly inversely proportional to the error of the in-band spectral index.

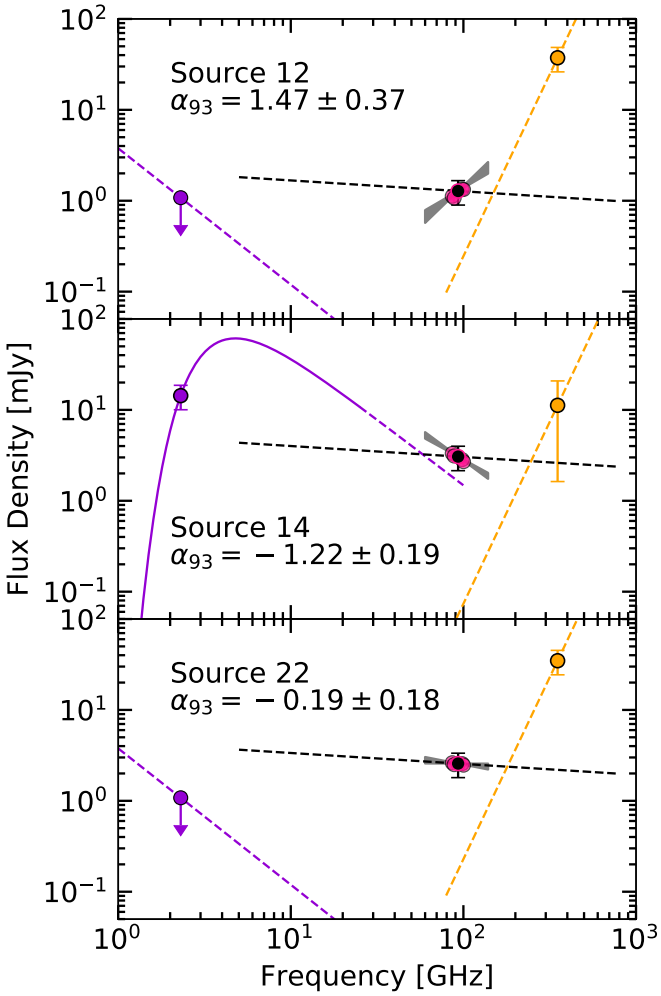


Figure 3.5: Example SEDs constructed for each source. *Top:* Dust dominated, Source 12. *Middle:* Synchrotron dominated, Source 14. *Bottom:* Free-free dominated, Source 22. The dashed orange line represents a dust spectral index of $\alpha = 4.0$, normalized to the flux density we extract at 350 GHz (orange data point). The dashed black line represents a free-free spectral index of $\alpha = -0.12$, normalized to the flux density we extract at 93 GHz (black data point). The pink data points show the flux densities extracted from the band 3 spectral windows. The gray shaded region is the 1σ error range of the band 3 spectral index fit, except we have extended the fit in frequency for displaying purposes. The purple line represents a synchrotron spectral index of $\alpha = -1.5$, normalized to the flux density we extract at 2.3 GHz (purple data point); except for Source 14 where the solid purple line represents the normalized 2.3–23 GHz fit from Lenc & Tingay (2009). Error bars on the flux density data points are 3σ.

continuum. To do this, we simulate how mixtures of synchrotron, free-free, and dust emission could combine to create the observed in-band index. We assume fixed spectral indices for each component and adjust their fractions to reproduce the observations (see below).

We consider that a source is dominated by two types of emission (a caveat which we discuss in detail in Section 3.6): free-free and dust, or free-free and synchrotron. With synthetic data points, we first set the flux density at 93 GHz, S_{93}, to a fixed value and vary the contributions of dust and free-free continua, such that $S_{93} = S_d + S_{ff}$. For each model, we determine the continuum flux at each frequency across the Band 3 frequency coverage as the sum of the two components. We assume that the frequency dependence of the free-free component is $\alpha_{\rm ff} = -0.12$ and the frequency dependence of the dust component is $\alpha_{\rm d} = 4.0$. We fit the (noise-less) continuum of the synthetic data across the Band 3 frequency coverage with a power-law, determining the slope as α_{93}, the in-band index. We express the results in terms of a free-free fraction and dust fraction, rather than absolute flux. In this respect, we explore free-free fractions to the 93 GHz continuum ranging from $f_{\rm ff} = (0.001, 0.999)$ in steps of 0.001. We let the results of this process constrain our free-free fraction when the in-band index measured in the actual (observed) data is $\alpha_{93} \geq -0.12$.

Next, we repeat the exercise, but we let synchrotron and free-free dominate the contribution to the continuum at 93 GHz. We take the frequency dependence of the free-free component as $\alpha_{\rm ff} = -0.12$ across the Band 3 frequency coverage, and we set the synchrotron component to $\alpha_{\rm syn} = -1.5$. We let the results of this process constrain the free-free fraction of the candidate star clusters when the fit to their in-band index is $\alpha_{93} \leq -0.12$.

Our choices for the spectral indices of the three emission types are motivated as follows. In letting, $\alpha_{\rm ff} = -0.12$ we assume that the free-free emission is optically thin (e.g., see Appendix 3.A.2). We do not expect significant free-free opacity at 93 GHz given the somewhat-evolved age of the candidate clusters in the starburst (see Section 3.5.2). In letting $\alpha_{\rm d} = 4.0$, we assume the dust emission is optically thin with a wavelength-dependent emissivity so that $\tau \propto \lambda^{-2}$ (e.g., see Draine 2011). The low optical dust optical depths estimated in Section 3.5.6 imply a dust spectral index steeper than 2, though the exact value might not be 4, e.g., if our assumed emissivity power law index is not applicable. The generally faint dust emission indicates that our assumptions about dust do not have a large effect on our results. For the synchrotron frequency dependence, we assume $\alpha_{\rm syn} = -1.5$. This value is consistent with the best fit slope of $\alpha_{\rm syn} = -1.4$ found by Bendo et al. (2016) averaged over the central $30''$ of NGC 4945. Furthermore, the median slope of synchrotron-dominated sources modeled at 2.3–23 GHz is -1.11 (Lenc & Tingay 2009), indicating that even at 23 GHz the synchrotron spectra already show losses due to aging, i.e., are steeper than a canonical initial injection of $\alpha_{\rm syn} \approx -0.8$. While our assumed value of $\alpha_{\rm syn} = -1.5$ is well-motivated on average, variations from source to source are likely present.

Through this method of decomposition, the approximate relation between the in-band index and the free-free fraction is

$$f_{\rm ff} = \begin{cases} 0.72\,\alpha_{93} + 1.09, & -1.5 \leq \alpha_{93} \leq -0.12 \\ -0.24\,\alpha_{93} + 0.97, & 4.0 \geq \alpha_{93} \geq -0.12 \end{cases} \tag{3.2}$$

for which $\alpha_{93} \leq -0.12$, the synchrotron fraction is found to be $f_{\rm syn} = 1 - f_{\rm ff}$ and we set $f_{\rm d} = 0$, and for which $\alpha_{93} \geq -0.12$ the dust fraction is found to be $f_{\rm d} = 1 - f_{\rm ff}$ and $f_{\rm syn} = 0$. This relation is depicted in the bottom panel of Figure 3.4 using the values that have been determined for each source.

Estimated free-free, synchrotron, and dust fractions to the 93 GHz continuum

Table 3.1 and Figure 3.4 summarize our estimated fractional contribution of each emission mechanism to the 93 GHz continuum of each source. We use the same relation above to translate the range of uncertainty on the spectral index to an uncertainty in the emission fraction estimates. We find, at this $0.12''$ resolution, the median free-free fraction of sources is $f_{\rm ff} = 0.62$ with median absolute deviation of 0.29. Most of the spectral indices are negative, and as a result, we find synchrotron emission can have a non-trivial contribution with a median fraction of $f_{\rm syn} = 0.36$ and median absolute deviation of 0.32. On the other hand, three of the measured spectral indices are positive. The median dust fraction of sources is $f_{\rm d} = 0$ and median absolute deviation of 0.10. A 1σ limit on the fractional contribution of dust does not exceed $f_{\rm d} = 0.47$ for any single source.

Additional continuum observations of comparable resolution at frequencies between 2 GHz and 350 GHz would improve the estimates of the fractional contribution of free-free, dust and synchrotron to the 93 GHz emission.

3.4 Recombination Line Emission

Hydrogen recombination lines at these frequencies trace ionizing radiation ($E > 13.6$ eV); this recombination line emission is unaffected by dust extinction. The integrated emission from a radio recombination line transition to quantum number n, which we derive for millimeter wavelength transitions in Appendix 3.A.1, is described by

$$\int S_{\mathsf{n}}\, {\rm dv} = \left(65.13\ {\rm mJy\ km\ s^{-1}}\right) b_{\mathsf{n}+1} \left(\frac{n_e n_p V}{5\times10^8\ {\rm cm^{-6}\ pc^3}}\right) \left(\frac{D}{3.8\ {\rm Mpc}}\right)^{-2} \times \left(\frac{T_e}{10^4\ {\rm K}}\right)^{-1.5} \left(\frac{\nu}{100\ {\rm GHz}}\right) \tag{3.3}$$

where $b_{\mathsf{n}+1}$ is the LTE departure coefficient, $EM_{\rm L} = n_e n_p V$ is the volumetric emission measure of ionized hydrogen, D is the distance to the source, T_e is the electron temperature of the ionized gas, and ν is the rest frequency of the spectral line.

Figure 3.6 shows spectra of the 15 sources with detected radio recombination line emission. We extract H40α and H42α spectra at the location of each source with an aperture diameter of $0.4''$, or twice the beam FWHM. We average the spectra of the two transitions together to enhance the signal-to-noise ratio of the recombination line emission. To synthesize an effective H41α profile, we interpolate the two spectra of each source to a fixed velocity grid with a channel width of 10.3 km s^{-1}, weight each spectrum by $\sigma_{\rm rms}^{-2}$ where $\sigma_{\rm rms}$ is the spectrum standard deviation, and average

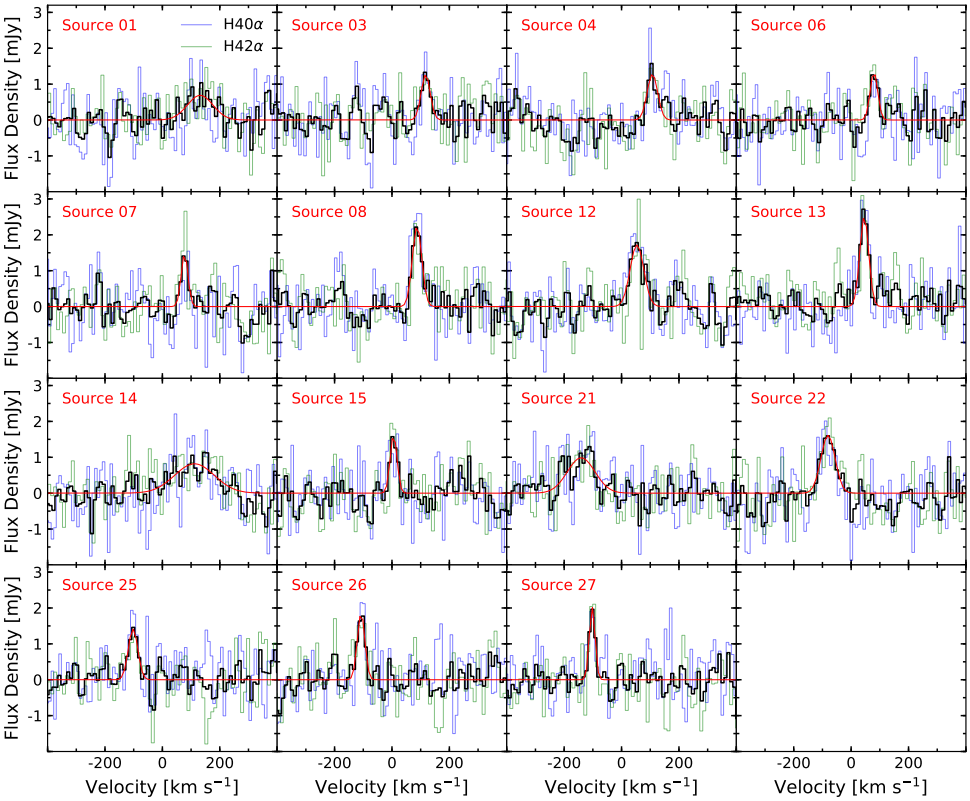


Figure 3.6: Radio recombination line spectra for sources with significantly detected emission. The thin blue line is the H40α spectrum. The thin green line is the H42α. These spectra have been regridded from their native velocity resolution to the common resolution of 10.3 km s^{-1}. The thick black line is the weighted average spectrum of H40α and H42α, effectively H41α. In red is the best fit to the effective H41α radio recombination line feature.

the spectra together lowering the final noise. The averaged spectrum has an effective transition of H41α at $\nu_{\rm eff} = 92.034$ GHz.

We fit spectral features with a Gaussian profile. We calculate an integrated signal-to-noise ratio for each line by integrating the spectrum across the Gaussian width of the fit (i.e., $\pm\sigma_{\rm Gaus}$) and then dividing by the noise over the same region, $\sqrt{N}\sigma_{\rm rms}$, where N is the number of channels covered by the region. We report on detections with an integrated signal of $> 5\sigma_{\rm rms}$. Table 3.2 summarizes the properties of the line profiles derived from the best-fit Gaussian. The median rms of the spectra is $\sigma_{\rm rms} = 0.34$ mJy.

In Figure 3.12 of Appendix 3.B, we show that the central velocities of our detected recombination lines are in good agreement with the kinematic velocity expected of the disk rotation. To do this, we overlay our spectra on H40α spectra extracted from the intermediate configuration observations (0.7″ resolution).

In 12 of the 15 sources, we detect relatively narrow features of FWHM $\sim$ (24 –

Table 3.2: Average Line Profiles nominally located near H41α.

Source	V_{cen} km s^{-1}	Peak (mJy)	FWHM km s^{-1}	$\sigma_{\rm rms}$ (mJy)
01	131.6 ± 12	0.69 ± 0.16	105.2 ± 28	0.35
03	117.4 ± 4.0	1.26 ± 0.28	37.2 ± 9.4	0.36
04	107.5 ± 4.3	1.28 ± 0.26	42.9 ± 10	0.36
06	79.4 ± 3.1	1.33 ± 0.27	31.1 ± 7.3	0.33
07	77.2 ± 3.0	1.46 ± 0.32	28.1 ± 7.0	0.36
08	87.5 ± 2.1	2.24 ± 0.24	38.7 ± 4.9	0.33
12	53.6 ± 3.9	1.72 ± 0.24	55.9 ± 9.1	0.39
13	45.7 ± 1.7	2.68 ± 0.26	34.8 ± 3.9	0.31
14	111.8 ± 12	0.82 ± 0.12	163.1 ± 28	0.33
15	5.6 ± 2.8	1.62 ± 0.33	28.2 ± 6.6	0.37
21	-140.7 ± 8.3	0.99 ± 0.15	113.3 ± 20	0.34
22	-80.6 ± 4.0	1.61 ± 0.22	58.4 ± 9.4	0.34
25	-102.3 ± 3.3	1.42 ± 0.24	39.0 ± 7.7	0.32
26	-107.6 ± 2.2	1.86 ± 0.27	30.9 ± 5.3	0.33
27	-102.3 ± 1.6	2.02 ± 0.27	23.6 ± 3.6	0.28

$V_{\rm cen}$ is the central velocity of the best fit Gaussian. Peak is the peak amplitude of the Gaussian fit. FWHM is the full-width half maximum of the Gaussian fit. $\sigma_{\rm rms}$ is the standard deviation of the fit-subtracted spectrum.

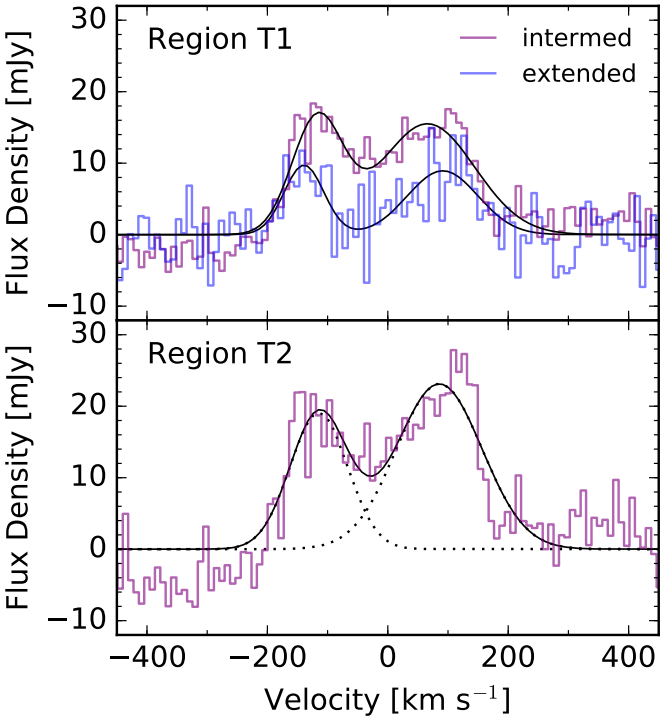


Figure 3.7: H40α line spectra extracted from the aperture regions T1 (top) and T2 (bottom; see Figure 3.8). In blue is the extended-configuration ("extended") spectrum extracted from the high-resolution 0.2″ data; this spectrum shows the maximum total integrated line flux extracted. In purple, the intermediate-configuration ("intermed") spectra extracted from low-resolution, native 0.7″ data; the spectrum from the T2 region is the total maximum integrated line flux from this data. The solid black line represents the sum total of two Gaussian fits. The dotted, black line represents the single Gaussian fits.

58) km s^{-1}. Larger line-widths of FWHM $\sim (105 - 163)$ km s^{-1} are observed from bright sources which also have high synchrotron fractions, indicating that multiple components, unresolved motions (e.g., from expanding shells or galactic rotation), or additional turbulence may be present. Six sources with detected recombination line emission have considerable ($f_{\rm syn} \gtrsim 0.50$) synchrotron emission (i.e. Sources 1, 4, 6, 14, 21, 27) at 93 GHz.

In the top panel of Figure 3.7, we show the total recombination line emission extracted from the starburst region in the 0.2″ resolution, "extended" configuration observations. The aperture we use, designated as region T1, is shown in Figure 3.8. Details of the aperture selection are described in Section 3.4.1. The spectrum consists of two peaks reminiscent of a double horn profile representing a rotating ring. We fit the spectrum using the sum of two Gaussian components. In Table 3.3, we include the properties of the best fit line profiles. The sum total area of the fits is (2.1±0.6) Jy km s^{-1}.

Table 3.3: H40α line profiles from the regions of total flux.

Region	Config	$v_{cen,1}$ (km s^{-1})	$Peak_1$ (mJy)	$FWHM_1$ (km s^{-1})	$v_{cen,2}$ (km s^{-1})	$Peak_2$ (mJy)	$FWHM_2$ (km s^{-1})	H40α flux (mJy km s^{-1})
T1	extended	-139 ± 9	9.7 ± 2	78 ± 22	91 ± 13	8.9 ± 1.8	138 ± 14	2100 ± 600
T1	intermed	-117 ± 9	16 ± 3	99 ± 21	66 ± 2	16 ± 2	186 ± 15	4800 ± 900
T2	intermed	-114 ± 13	19 ± 4	113 ± 31	86 ± 13	23 ± 3	167 ± 14	6400 ± 1000

3.4.1 Line Emission from 0.7 ″ resolution, Intermediate configuration Observations

Figure 3.7 also shows integrated spectra derived from intermediate-resolution (0.7″) data. We use these data as a tracer of the total ionizing photons of the starburst region. We expect that the intermediate-resolution data includes emission from both discrete, point-like sources and diffuse emission from any smooth component.

The total integrated emission in the intermediate-configuration data is about three times larger than the integrated emission in the extended-configuration data. Spectra representing the total integrated line flux are shown in Figure 3.7. In Figure 3.8, we show the integrated intensity map of H40α emission from the intermediate-configuration (0.7″) data. In Table 3.3, we include the best fit line profiles.

We also compare the line profiles of H40α and H42α in the intermediate-configuration (0.7″) data, see Table 3.4 and Figures 3.9 & 3.10. We find that the integrated line emission of H42α is enhanced compared with H40α, reaching a factor of 2 greater when integrated over the entire starburst region. Yet we see good agreement between the two lines at the scale of individual cluster candidates. Spectral lines (possibly arising from c-C_3H_2) likely contaminate the H42α line flux in broad, typically spatially unresolved, line profiles.

Total Emission from the Starburst Region

In Figure 3.8 we show the integrated intensity map of H40α emission integrated between $v_{systemic} \pm 170$ km s^{-1}, as calculated from the 0.7″ intermediate-configuration data. Diffuse emission is detected throughout the starburst region and up to 30 pc in apparent size beyond the region where we detect the bright point sources at high resolution.

Also shown in Figure 3.8 are the apertures used to extract spectra in Figure 3.7. We fit a two dimensional Gaussian to the continuum emission in the 0.7″ resolution observations (see Figure 3.9). This results in a best fit centered at $(\alpha, \delta) = (13\mathrm{h}\,05\mathrm{m}\,27.4896\mathrm{s}, -49^\circ\,28'\,05.159'')$, with major and minor Gaussian widths of $\sigma_{\rm maj} = 2.4''$ and $\sigma_{\rm min} = 0.58''$, and an angle of $\theta = 49.5^\circ$; we use this fit as a template for the aperture location, position angle, width and height. We independently vary the major and minor axes (in multiples of $0.5\sigma_{\rm maj}$ and $0.5\sigma_{\rm min}$, respectively) in order to determine the aperture which maximizes the total integrated signal in channels within ± 170 km s^{-1}. With the extended-configuration cube, we find the largest integrated line emission with an aperture of $8.4'' \times 1.5''$, which we refer to as T1. With the intermediate-configuration cube, the largest integrated line emission arises with an aperture of $12.0'' \times 2.9''$, which we refer to as T2.

We extract the total H40α line flux from the intermediate-configuration (0.7″ resolution) cube using the T2 aperture (see bottom panel of Figure 3.7). The spectrum shows a double horn profile, indicating ordered disk-like rotation. We fit the features with two Gaussians. The sum total of their integrated line flux is (6.4 ± 1.0) Jy km s^{-1}.

We also extract H40α line flux from the intermediate-configuration (0.7″ resolution) cube within the T1 aperture in order to directly compare the integrated line flux in the two different data sets using the same aperture regions. We find more

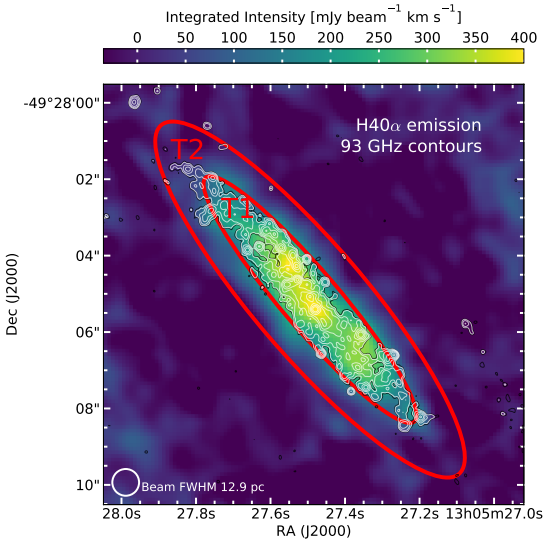


Figure 3.8: Integrated intensity (moment 0) map of H40α emission integrated between $V_{\rm systemic} \pm$ 170 km s^{-1} and observed with the intermediate telescope configuration at native 0.7″ resolution. Overlaid are contours of the 93 GHz continuum from extended-configuration, high-resolution (0.12″) data – as described in Figure 1. Red ellipses mark the apertures used to extract the total line emission from regions T1 and T2.

emission in the intermediate-configuration data, a factor of ∼2.3 greater than the extended-configuration data. This indicates that some recombination line emission originates on large scales (>100 pc) to which the high-resolution, long baselines are not sensitive.

H42α Contamination

In this section we compare the line profiles from H40α and H42α extracted from the 0.7″ intermediate configuration data. In principle, we expect the spectra to be virtually identical, which is why we average them to improve the signal-to-noise at high-resolution. Here we test that assumption at low-resolution. To summarize, we find evidence that a spectral line may contaminate the H42α measured line flux in broad (typically spatially unresolved) line profiles. Yet we see good agreement between the two lines at the scale of individual cluster candidates.

We extracted spectra in three apertures to demonstrate the constant velocity offset of the contaminants. We approximately matched the locations of these apertures to those defined in Bendo et al. (2016), in which H42α was analyzed at 2.3″ resolution; in this way we are able to confirm the flux and line profiles we extract at 0.7″ resolution with those at 2.3″. The non-overlapping circular apertures with diameters of 4″ designated as North (N), Center (C), and South (S) are shown in Figure 3.9.

We used our intermediate-configuration data to extract an H40α and an H42α spectrum in each of the regions. We overplot the spectra of each region in Figure 3.10. We fit a single Gaussian profile to the line emission, except for H42α emission in region

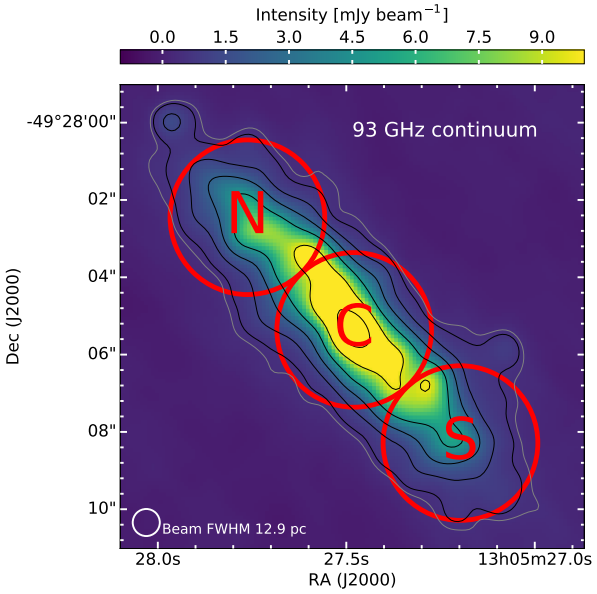


Figure 3.9: Continuum emission at 93 GHz observed with an intermediate configuration with native resolution FWHM = 0.7″ (or 12.9 pc at the distance of NGC 4945). The rms noise away from the source is $\sigma \approx 0.15$ mJy beam^{-1}. Contours of the continuum image show 3σ emission (gray) and $[4, 8, 16, ...]\sigma$ emission (black). Apertures (red) with a diameter of 4″ mark the regions N, C, and S.

Table 3.4: Comparison of integrated recombination line flux.

Region	H42α flux (Jy km s^{-1})	H40α flux (Jy km s^{-1})	Ratio H42α/H40α
N	2.1 ± 0.2	1.5 ± 0.2	1.4 ± 0.2
C	5.9 ± 0.3	3.7 ± 0.3	1.6 ± 0.2
S	1.9 ± 0.2	0.96 ± 0.1	2.0 ± 0.3

N where two Gaussian components better minimized the fit. The total area of the fits are presented in Table 3.4 as the integrated line flux.

Our line profiles of H42α are similar in shape and velocity structure as those analyzed in Bendo et al. (2016) and the integrated line emission is also consistent (within 2σ). This indicates that we are recovering the H42α total line flux and properties with our data.

On the other hand, the H40α flux we extract is about a factor of $\sim$1.6 lower than the H42α fluxes in these apertures (see Figure 3.10 and Table 3.4). The discrepancy grows to a factor of 2 in the profile extracted from the total region.

The additional emission seen in the H42α spectrum at velocities +100 km s^{-1} to +250 km s^{-1} with respect to the bright, presumably hydrogen recombination line peak, is absent in the H40α profile. It is not likely to be a maser-like component of hydrogen recombination emission since the relative flux does not greatly vary in

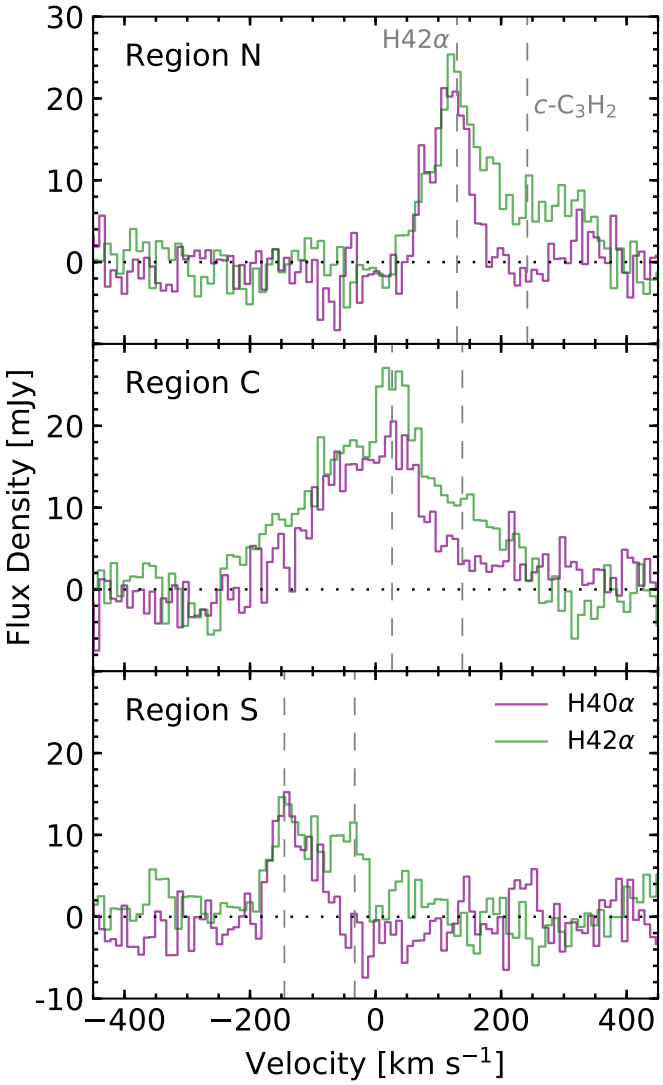


Figure 3.10: Comparison of our H40α (purple) and H42α (green) from intermediate configuration, low-resolution data from the regions defined in Figure 3.9 as N (top), C (middle), and S (bottom). We find H42α to be contaminated by spectral lines which may include c-C_3H_2 $4_{32} - 4_{23}$ — shown as a dashed line in the panels at expected velocities with respect to H42α. When contaminant lines are included, the integrated line flux of H42α is over estimated by a factor of 1.5 in these apertures; this grows to a factor of 2 when integrating over the total starburst emission.

different extraction regions, and densities outside of the circumnuclear disk would not approach the emission measures necessary (e.g., $EM_v \gtrsim 10^{10}$ cm^{-6} pc^3) for stimulated line emission.

We searched for spectral lines in the frequency range $\nu_{\rm rest} \sim 85.617$ – 85.660 GHz, corresponding to these velocities and find several plausible candidates, though we were not able to confirm any species with additional transitions in the frequency coverage of these observations. A likely candidate may be the $4_{32} - 4_{23}$ transition of c-C_3H_2. c-C_3H_2 has a widespread presence in the diffuse ISM of the Galaxy (e.g., Lucas & Liszt 2000) and the $2_{20} - 2_{11}$ transition has been detected in NGC 4945 (Eisner et al. 2019). As an example we plot the velocity of c-C_3H_2 $4_{32} - 4_{23}$ relative to H42α in Figure 3.10.

3.5 Physical Properties of the Candidate Star Clusters

In this section, we estimate properties of the candidate star clusters, summarized in Table 3.6. We discuss their size and approximate age. Properties of the ionized gas content, such as temperature (see Table 3.5), metallicity, density and mass are derived from the continuum and recombination line emission. We estimate the ionizing photon rate of the candidate stars clusters and use it to infer the stellar mass (see Figure 3.11). From the dust emission at 350 GHz, we estimate gas masses of the candidate star clusters. With a combined total mass from gas and stars, we estimate current mass surface densities and free-fall times.

We exclude Source 5 from the analysis since the free-free fraction is $f_{\rm ff} < 0.01$. We also remove the presumed AGN core (Source 18) from the analysis.

3.5.1 Size

The sources identified through `PyBDSF` in the 93 GHz continuum image are fit with two dimensional Gaussians. The average of the major and minor (convolved) FWHM is listed in Table 3.6 as the FWHM size of the source in units of pc. The Gaussian fits are all consistent with circular profiles within error. FWHM sizes of (1.4–4.0) pc are observed, consistent with typical sizes of young, massive star clusters (Ryon et al. 2017; Leroy et al. 2018). However, the lower end may reflect the resolution limit of our beam, with a FWHM size of 2.2 pc. The uncertainties we report reflect the errors of the Gaussian fit.

Based on high-resolution imaging of embedded clusters in the nucleus of the Milky Way and NGC 253, some of these clusters might break apart at higher resolution (Ginsburg et al. 2018, Levy et al., in prep). If they follow the same pattern seen in these other galaxies, each source would have one or two main components potentially with several associated fainter components.

3.5.2 Age

Throughout our analysis, we assume that the candidate star clusters formed in an instantaneous burst of star-formation roughly 5 Myr ago (with a likely uncertainty of ∼1 Myr). This (approximately uniform) age is supported through the coincident detection of RRLs and supernovae remnants, previous analyses of the global population of the burst (e.g., Marconi et al. 2000; Spoon et al. 2000) and an orbital timescale of ≈3 Myr for the starburst region. We elaborate on this supporting evidence below.

As we discuss in Section 3.3.3, dust does not significantly contribute to 93 GHz emission (with a median fraction of $f_{\rm d} = 0$), but synchrotron emission does through supernova remnants. Supernova explosions begin from ∼3 Myr in the lifetime of a cluster and cease around ∼40 Myr when the most massive stars have died out; this puts the loosest bounds on the age of the candidate clusters we observe. The coincident detection of supernovae remnants in a third (6/15) of our recombination line detected sources implies that the burst is likely not at the earliest stage of the supernovae phase. However the ionizing photon rate changes dramatically over 3 Myr to 10 Myr, dropping by about two orders of magnitude (Leitherer et al. 1999). As a result clusters are significantly harder to detect in radio recombination lines or free-free continuum emission after ∼5 Myr.

Properties of star-forming activity in the central starburst have been estimated by combining far-infrared (FIR) and optical/IR tracers. Marconi et al. (2000) discerned an age of 6 Myr and mass of 4×10^7 $M_\odot$ by using Paα and Brγ to trace the energy distribution of the photon output of the population. However, the dust extinction was underestimated, complicated by the uncertainty in the AGN contribution. Mid-infrared (MIR) observations with the Infrared Space Observatory (ISO; Kessler et al. 1996) of line ratios further constrained this scenario. Spoon et al. (2000) estimated an extinction of $A_{\rm V} = 36^{+18}_{-11}$, determined that the AGN is not dominating the ionizing radiation field, and found that the star-forming population is consistent with a burst of age ≥5 Myr.

As a sanity check on whether a synchronized burst might be expected, we calculate the orbital timescale associated with the the burst region. Taking the rotation velocity ∼ 170 km s^{-1} from the integrated spectrum and the radius ∼ 80 pc associated with region T1, we estimate an orbital timescale of ∼ 3 Myr. If we take this as roughly the timescale for the nuclear disk to react to changing conditions, a burst shutting off or turning on in a ∼ 5 Myr timescale is reasonable.

3.5.3 Temperature and Metallicity

The ratio of the integrated recombination line flux (Equation 3.3) to the free-free continuum flux density (Equation 3.1) allows the electron temperature to be determined. Dependencies on the distance, emission measure, and (possible) beam-filling effects cancel out under the assumption that the two tracers arise in the same volume of gas. We show in Appendix 3.A.3, that when taking the ratio of the integrated line to continuum, $R_{\rm LC}$, and solving for the temperature, T_e, we arrive at

$$T_e = 10^4 \text{ K} \left[b_{\mathfrak{n}+1} \, (1+y)^{-1} \left(\frac{R_{\rm LC}}{31.31 \text{ km s}^{-1}} \right)^{-1} \left(\frac{\nu}{100 \text{ GHz}} \right)^{1.12} \right]^{0.85} \tag{3.4}$$

Table 3.5: Temperature analysis.

Source	$\int S_{\rm L}\,{\rm dV}$ (mJy km s^{-1})	S_{93} (mJy)	$f_{\rm ff}$	T_e (K)
08	92 ± 15	3.5 ± 0.4	0.74 ± 0.20	6000 ± 1700
13	99 ± 14	2.9 ± 0.3	0.91 ± 0.22	5600 ± 1400
22	100 ± 21	3.4 ± 0.3	0.78 ± 0.20	5600 ± 1700
26	61 ± 13	2.6 ± 0.3	0.72 ± 0.29	6400 ± 2600
27	50 ± 10	3.6 ± 0.4	0.45 ± 0.16	6500 ± 2300

$\int S_{\rm L}\,{\rm dV}$ refers to the integrated line emission. S_{93} is the continuum flux density extracted at 93 GHz in the $0.2''$ resolution image. $f_{\rm ff}$ is the estimated free-free fraction at $0.2''$ resolution. T_e is the electron temperature derived using Equation 3.4.

where $b_{\mathsf{n}+1}$ is the non-LTE departure coefficient, and y is the abundance ratio of ionized helium to hydrogen number density, $y = n_{He+}/n_p$, which we fix as $y = 0.10$ (de Pree et al. 1996; Mills et al. 2020).

Table 3.5 lists the temperatures we derive in the region. We focus on the 5 sources with bright (peak S/N $> 4.7\sigma$) and well-fit recombination line emission. Most of these sources have higher free-free fractions than the median. To derive the temperatures, we re-evaluate the continuum (fraction of) free-free emission at the resolution of $0.2''$, since the free-free fraction may change with resolution. Therefore, we convolve the Band 3 continuum images to $0.2''$ resolution. We extract the continuum from the full-bandwidth image through aperture photometry, using an aperture diameter of $0.4''$. In order to exactly match the processing of the spectral line data, we do not subtract background continuum emission within an outer annulus. Then, by extracting the continuum in each spectral window (using the same aperture diameters just described), we fit for the in-band spectral index. We use the procedure described in Section 3.3.3 to constrain the free-free fraction from the spectral index fit.

With the free-free fraction and measured fluxes, we plug in the line to continuum ratio into Equation 3.4 and take $b_{\mathsf{n}} = 0.73$ (Storey & Hummer 1995) to arrive at the temperature. The departure coefficient at $\mathsf{n} = 41$ is loosely ($< 15\%$ variation) dependent on the temperature. We iterate (once) on the input b_{n} and output temperature. $b_{\mathsf{n}} = 0.73$ is the modeled value for this temperature and for typical densities of $n_e = (10^3 - 10^4)$ cm^{-3} of ionized gas surrounding young, massive stars and consistent with the ionized gas densities we derive in Section 3.5.4.

The uncertainties in the electron temperatures we derive in Table 3.5 are dominated by the uncertainties in the free-free fraction. We take the mean and standard deviation values of $T_e = (6000 \pm 400)$ K as a representative electron temperature of the ionized plasma in the candidate star clusters. This temperature is consistent with the temperature derived from a lower-resolution analysis of NGC 4945 at $2.3'' \times 2.6''$ resolution, which finds $T_e = (5400 \pm 600)$ K (Bendo et al. 2016).

Our estimated temperature implies a thermal line width of (16 ± 4) km s^{-1} (Brocklehurst & Seaton 1972). Given that this is smaller than our observed line widths, non-thermal motions from bulk velocities (such as turbulence, inflow or outflow) must

contribute to broadening the spectral line profiles.

The electron temperature of free-free plasma surrounding massive stars is related to the metallicity of the plasma, as the metals contribute to gas cooling. Shaver et al. (1983) established a relation,

$$12 + \log_{10}(\mathrm{O/H}) = (9.82 \pm 0.02) - (1.49 \pm 0.11)\frac{T_e}{10^4\,\mathrm{K}}, \tag{3.5}$$

with the temperatures and metallicities derived with (auroral) collisionally excited lines at optical wavelengths. Furthermore, they showed that these temperatures are consistent with electron temperatures derived from radio recombination lines. We find a representative O/H metallicity of $12 + \log_{10}(\mathrm{O/H}) = 8.9 \pm 0.1$. This value is in approximate agreement (within 2σ) with the average metallicity and standard deviation of $12 + \log_{10}(\mathrm{O/H}) = 8.5 \pm 0.1$ (Stanghellini et al. 2015) determined in 15 star-forming regions in the galactic plane of NGC 4945 (and which is consistent with no radial gradient) using strong-line abundance ratios of oxygen, sulfur, and nitrogen spectral lines.

3.5.4 Ionized Gas: Emission Measure, Density and Mass

We determine the volumetric emission measure of gas ionized in candidate stars clusters using Equations 3.1 and 3.3 together with the mean temperature derived in Section 3.5.3. In Table 3.6, we list the results for each candidate star cluster. Emission measures that we determine from the free-free continuum range from $\log_{10}(EM_{\mathrm{C}}/\mathrm{cm}^{-6}\,\mathrm{pc}^3) \sim 7.3$ – 8.7, with a median value of 8.4. We also calculate the volumetric emission measure of ionized hydrogen as determined by the effective H41α recombination line when applicable, noting that $EM_{\mathrm{C}} = (1 + y)\,EM_{\mathrm{L}}$. The line emission measures range from $\log_{10}(EM_{\mathrm{L}}/\mathrm{cm}^{-6}\,\mathrm{pc}^3) \sim 8.4$ – 8.9, with a median value of 8.5. The uncertainty in the emission measures is $\sim$0.4 dex and is dominated by the errors of the free-free fraction.

Next, we solve for the electron density. We use the emission measure determined from the free-free continuum, assume $n_e = n_+$, and consider a spherical volume with $r = \mathrm{FWHM_{size}}/2$. We arrive at densities between $\log_{10}(n_e/\mathrm{cm}^{-3}) = 3.1$–$3.9$ with a median value of 3.5.

We matched (see Section 3.3.2) five of the candidate star clusters that have recombination line emission detected – Sources 1, 6, 14, 21, 27 – with the 2.3 GHz objects of Lenc & Tingay (2009) which have the free-free optical depth modeled through their low-frequency turnovers. Although the 2.3 GHz objects have non-thermal indices, it is their radio emission which is opaque to free-free plasma. Using the optical depths derived in Lenc & Tingay (2009) and our fiducial electron temperature, we solve for the density through the relation, $\tau \approx 3.28 \times 10^{-7} \left(\frac{T_e}{10^4\ \mathrm{K}}\right)^{-1.35} \left(\frac{\nu}{\mathrm{GHz}}\right)^{-2.1} \left(\frac{EM_\ell}{\mathrm{cm}^{-6}\ \mathrm{pc}}\right)$ (Condon & Ransom 2016), where $EM_\ell = n_e n_+ \ell$ and for which a spherical region the pathlength ℓ translates as $\ell = \frac{3}{4}r$. We find densities in the range $\log_{10}(n_e/\mathrm{cm}^{-3}) = 3.3$ – 3.6. This agrees well with the values we separately derive.

We convert the ionized gas density and source sizes to an ionized gas mass through,

$$M_+ = 1.36 m_{\mathrm{H}}\, n_+ \frac{4}{3}\pi r^3 \tag{3.6}$$

where we have assumed a 1.36 contribution of helium by mass and we let $r = \mathrm{FWHM_{size}}/2$. The ionized gas masses of the candidate star clusters range from $\log_{10}(M_+ \ / \ \mathrm{M_\odot}) = 2.7$ – 3.5 with a median value of 3.1. The ionized gas mass is a small fraction ($\lesssim 1\%$) of the stellar mass (see Section 3.5.5).

3.5.5 Ionizing Photon Production and Stellar Mass

We estimate the number of the ionizing photons needed per second to maintain the total free-free emitting content (see Table 3.6). From the emission measure of ionized gas and the temperature-dependent recombination coefficient for case B recombination, the rate of ionizing photons (see Appendix 3.A.3) with $E > 13.6$ eV is

$$Q_0 = \left(3.8 \times 10^{51} \ \mathrm{s^{-1}}\right) \left(\frac{n_e n_+ V}{5 \times 10^8 \ \mathrm{cm^{-6}\ pc^3}}\right) \times \left(\frac{T_e}{10^4 \ \mathrm{K}}\right)^{-0.83}. \tag{3.7}$$

where $EM_\mathrm{C} = n_e n_+ V$ is the volumetric emission measure of the total ionized gas which we take from the continuum derived emission measure, and T_e is the electron temperature of the ionized gas. Our candidate star clusters have ionizing photon rates in the range $\log_{10}(Q_0/\mathrm{s^{-1}}) \sim 50.4$ – 51.8. The sum of the ionizing photon rate over all candidate, massive star clusters is 5.3×10^{52} $\mathrm{s^{-1}}$. In the top panel of Figure 3.11, the ionizing photon rates of the candidate clusters are plotted as complementary cumulative fractions.

We use Starburst99 calculations (Leitherer et al. 1999) to infer the stellar mass from the ionizing photon output of a 5 Myr old stellar population, via

$$M_\star \approx \frac{Q_0}{4.7 \times 10^{45}} \mathrm{M_\odot}. \tag{3.8}$$

We arrive at this value by simulating a single 10^6 $\mathrm{M_\odot}$ stellar population, with the initial mass function (IMF) of Kroupa (2001), a maximum stellar mass of 100 $\mathrm{M_\odot}$, and the default stellar evolution tracks and tuning parameters. Then we divide the ionizing photon output at 5 Myr by the initial mass of the stellar population. We note that this is a rough approximation which has not accounted for the amount of ionizing photons absorbed by dust, mass ejected from the system, and/or enhanced emission from stellar binaries.

Our candidate star clusters have stellar masses in the range $\log_{10}(M_\star/\mathrm{M_\odot}) \sim 4.7$–6.1 (see Table 3.6) with a median of 5.5. The error on the mass estimate is ∼0.4 dex. The sum of the stellar masses of the candidate stars clusters is $\approx 1.1 \times 10^7$ $\mathrm{M_\odot}$. In the bottom panel of Figure 3.11, the estimated stellar masses of the candidate clusters are plotted as cumulative fractions.

Table 3.6: Physical properties of candidate star clusters.
Source 5 is not included since its free-free fraction is $f_{\rm ff} < 0.01$; Source 18 is not included since it is the AGN core. FWHM is the source size as best fit from a Gaussian (to flux that has not been deconvolved); the errors reflect the fit of the Gaussian. $EM_{\rm C}$ is the free-free emission measure derived from the continuum as in Equation 3.1. $EM_{\rm L}$ is the hydrogen free-free emission measure derived from effective H41α as in Equation 3.3; we note $EM_{\rm C} = (1+y)\,EM_{\rm L}$. Q_0 is the ionizing photon rate derived from $EM_{\rm C}$ as in Equation 3.7. $M_\odot$ is the stellar mass derived from Q_0 as in Equation 3.8.
[a] The error on these quantities is $\sim$0.4 dex.

Source	FWHM (pc)	$\log(EM_{\rm C})^a$ ($\rm cm^{-6}\ pc^3$)	$\log(EM_{\rm L})^a$ ($\rm cm^{-6}\ pc^3$)	$\log(Q_0)^a$ ($\rm s^{-1}$)	$\log(M_\star)^a$ ($M_\odot$)	$\log(M_{\rm gas})^a$ ($M_\odot$)	$\log(\Sigma_{\rm Tot})^a$ ($M_\odot\ \rm pc^{-2}$)	$\log(t_{\rm ff})^a$ (yr)
01	3.0 ± 0.1	8.1	8.6	51.2	5.1	< 4.5	4.0	4.9
02	2.6 ± 0.1	8.2	...	51.3	5.2	4.8	4.3	4.7
03	3.1 ± 0.1	8.2	8.4	51.2	5.2	4.8	4.1	4.9
04	2.5 ± 0.1	7.7	8.5	50.8	4.7	4.6	4.0	4.9
06	2.9 ± 0.1	8.0	8.4	51.1	5.0	< 4.5	3.8	5.0
07	2.7 ± 0.1	8.1	8.4	51.2	5.1	4.6	4.2	4.8
08	2.4 ± 0.1	8.5	8.7	51.5	5.5	4.5	4.5	4.6
09	2.4 ± 0.1	8.1	...	51.2	5.1	< 4.3	4.1	4.8
10	2.9 ± 0.1	8.2	...	51.2	5.1	< 4.5	4.0	4.9
11	1.4 ± 0.1	8.0	...	51.1	5.0	4.0	4.6	4.5
12	2.7 ± 0.1	8.2	8.7	51.3	5.2	5.0	4.4	4.7
13	2.5 ± 0.1	8.5	8.7	51.6	5.5	4.7	4.6	4.6
14	2.2 ± 0.1	8.1	8.9	51.2	5.1	4.4	4.3	4.7
15	2.5 ± 0.1	8.0	8.4	51.0	5.0	< 4.4	4.0	4.9
16	2.3 ± 0.1	8.1	...	51.2	5.1	4.7	4.3	4.7
17	3.1 ± 0.1	8.7	...	51.7	5.7	5.2	4.6	4.6
19	2.4 ± 0.1	7.4	...	50.5	4.4	< 4.3	3.5	5.1
20	3.7 ± 0.1	8.4	...	51.4	5.3	5.1	4.2	4.9
21	3.1 ± 0.1	7.3	8.8	50.4	4.3	< 4.6	3.1	5.4
22	2.4 ± 0.1	8.7	8.7	51.8	5.7	4.9	4.8	4.5
23	2.6 ± 0.1	8.2	...	51.3	5.2	4.4	4.3	4.7
24	3.1 ± 0.1	7.6	...	50.6	4.6	< 4.6	3.4	5.2
25	3.9 ± 0.2	8.2	8.5	51.3	5.2	5.0	4.0	5.0

26	3.4 ± 0.1	8.3	8.5	51.3	5.3	4.9	4.2	4.9
27	2.2 ± 0.1	8.2	8.4	51.3	5.2	< 4.3	4.3	4.7
28	4.0 ± 0.1	8.0	...	51.1	5.0	< 4.8	3.6	5.1
29	2.9 ± 0.1	8.4	...	51.5	5.4	4.6	4.3	4.7

3.5.6 Gas Mass from Dust

We estimate the mass of gas associated with each candidate star cluster (see Table 3.6) from dust emission at 350 GHz. We determine the dust optical depth by comparing the measured intensity with that expected from an estimate of the true dust temperature. Assuming a mass absorption coefficient, we convert the optical depth to a dust column density. We arrive at a gas mass by multiplying the dust column density with the measured source size and an assumed dust-to-gas mass ratio.

We assume a dust temperature of $T_{\rm dust} = 130$ K, as has been determined for the gas kinetic temperature in the forming super star clusters in NGC 253 (Gorski et al. 2017). This is an approximation, though the uncertainty is linear. Then we convert the 350 GHz flux density into an intensity (I_{350}), and solve for the optical depth through $I_{350} \approx \tau_{350} B_\nu(T_{\rm dust})$ where $B_\nu(T_{\rm dust})$ is the Planck function evaluated at 350 GHz. We measure optical depths in the range $\tau_{350} \sim 0.02$ – 0.10, with a median value of $\tau_{350} \sim 0.04$, justifying our optically thin assumption. We note that the 3σ upper limit of the sources which have not been detected at 350 GHz corresponds to $\tau_{350} < 0.02$.

Next, we convert the optical depth to a dust column density using an assumed mass absorption coefficient (κ). We adopt $\kappa = 1.9$ cm^2 g^{-1} which should be appropriate for $\nu \sim 350$ GHz and dust mixed with gas at a density of $\sim 10^5 - 10^6$ cm^{-3} (Ossenkopf & Henning 1994), but we do note the large (factor of 2) uncertainties on this value. Finally, we combine the dust surface density with an adopted dust-to-gas mass ratio (DGR) of 1-to-100, approximately the Milky Way value and similar to the value found for starburst galaxies by Wilson et al. (2008). Our estimate for the gas surface density is determined with:

$$\Sigma_{\rm gas} = \frac{\tau_{350}}{\kappa\,{\rm DGR}}. \tag{3.9}$$

We determine the gas mass by multiplying the gas surface density by the two dimensional area of the source size, $M_{\rm gas} = A\,\Sigma_{\rm gas}$.

The gas masses we estimate are included in Table 3.6. We find values in the range of $\log_{10}(M_{\rm gas}\ /\ {\rm M}_\odot) = 4.4$ – 5.1 with a median value of 4.7. Upper limits for the sources which have not been detected in 350 GHz emission are included in the Table.

3.5.7 Total Mass from Gas and Stars

We estimate the current total mass of each candidate star cluster as the sum of the gas and star masses, where $M_{\rm Tot} = M_{\rm gas} + M_\star$. The total mass is dominated by the stellar mass, as we find low gas mass fractions of $M_{\rm gas}/M_{\rm Tot} = 0.04$ – 0.22 and a median of 0.13. When calculating the total mass of the 10 sources which are not detected at 350 GHz, we do not include the lower limit of the dust mass; we only consider the stellar mass.

We express the total mass of each source in terms of a surface density (in Table 3.6). This is calculated within the FWHM of the region; we thereby divide the total mass by 2 and divide by the 2D area of the FWHM. The values range from $\log(\Sigma_{\rm Tot}) = 3.1$ – 4.8 M$_\odot$ pc^{-2}, or alternatively, $\Sigma_{\rm Tot} = 0.3$ –13 g cm^{-2}.

Using the total mass, we estimate the gravitational free fall time of the clusters, $t_{\rm ff} = \left(\frac{\pi r^3}{8GM_{\rm Tot}}\right)^{1/2}$. The values we derive are included in Table 3.6, and range from

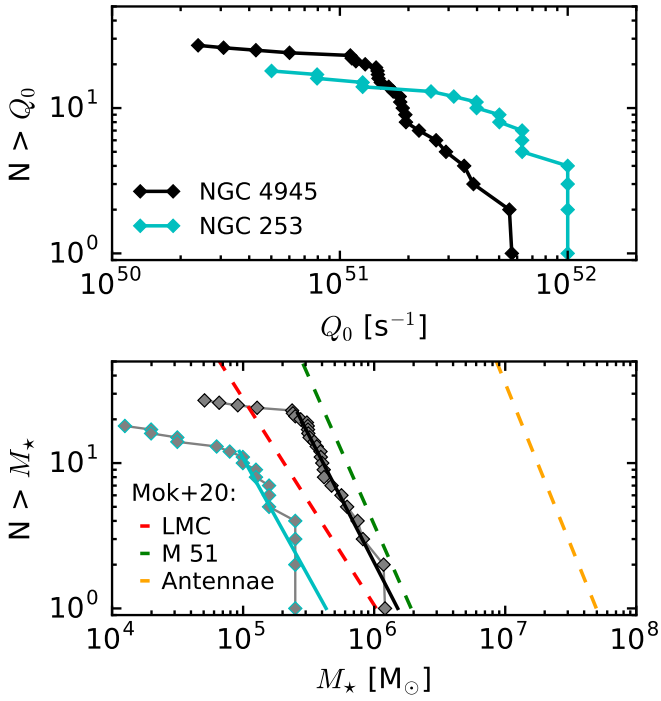


Figure 3.11: *Top:* The complementary cumulative distribution of the ionizing photon rate, Q_0, of candidate star clusters in NGC 4945 (see Table 3.6) and in NGC 253 (Mills et al. 2020). The turn-off at lower values likely reflects completeness limits. *Bottom:* The stellar mass, $M_\star$, of our candidate star clusters inferred from the ionizing photon rates of a cluster with an age of 5 Myr, plotted as a complementary cumulative distribution. We also include the stellar masses of clusters in the starburst of NGC 253 (Mills et al. 2020) and in the galaxies LMC, M 51, and the Antennae (Mok et al. 2020). Since the clusters in NGC 253 are likely close to a zero age main sequence, they would produce more ionizing photons per unit mass as compared with the slightly older stellar population in the clusters of NGC 4945.

$\log_{10}(t_{\rm ff}$ / yr) = 4.3 – 5.2 with a median of 4.6. This is the gravitational free fall time that would be experienced by gas with no support if all of the cluster mass were gas. The fact that the gas mass fractions are low and that the age appears longer than the free fall time adds support to the idea that these clusters have mostly already formed.

Summing all sources, we find a total mass within the candidate clusters of $M_{\rm Tot} \approx 1.5 \times 10^7$ M$_\odot$.

3.6 Discussion

Having estimated properties of the candidate star clusters, we explore implications of the results and the role of the candidate clusters with respect to the starburst. Before elaborating on that, we discuss the two main sources of uncertainty in the properties of the candidate star clusters: the free-free fractions at 93 GHz and the age of the burst.

3.6.1 Discussion of uncertainties

Both the free-free fractions at 93 GHz and the age of the burst could be better constrained using future observations, and this would result in more accurate estimates of almost all properties of the cluster candidates.

The first source of uncertainty that we discuss is the estimate of the free-free flux, by which we estimate a free-free fraction to the 93 GHz continuum.

In estimating the free-free fractions, we enforce that only two contributions of emission type, free-free and synchrotron, or, free-free and dust, compose the flux at 93 GHz. If we consider three emission types, we would obtain non-unique solutions in estimating the fractions of emission from the spectral index. Assuming two types is the best that we have assessed with the currently available data. Small dust opacities and a majority of negative slopes measured at 93 GHz independently indicate that dust does not substantially contribute in the majority of sources. Therefore assuming two dominant components to the emission at 93 GHz appears to be valid in the majority of sources. In order to decompose the emission at 93 GHz through SED fitting, additional observations at intermediate frequencies (within 2.3 GHz and 350 GHz) of comparable resolution are needed.

To estimate the free-free fraction, we also assume fixed energy distribution slopes for emission from free-free, synchrotron and dust. However, variations from source to source may be expected. If we re-derive the free-free fractions, letting $\alpha_{\rm d} = 2$ without changing the free-free and synchrotron indices, we find the median dust fraction of sources remains at $f_{\rm d} = 0$. This is another indication that the spectral index assumed for dust does not have a considerable impact on the majority of sources. The range of reasonable values to consider for the frequency dependence of non-thermal emission is less constrained, where single-injection indices of $\alpha_{\rm syn} = -0.5$ to -0.8 can be expected all the way up to the dramatic exponential cutoff[2] in a single-injection scenario where the highest energy electrons (at the high frequency end) completely depopulate after a characteristic energy-loss time (e.g., Klein et al. 2018a). As we note in Section 3.3.3,

[2]where $S \propto {\rm e}^{-\nu/\nu_b}$ and ν_b is the break frequency

our assumed index of $\alpha_{\rm syn} - 1.5$ agrees well with the measured value of -1.4 at lower resolution (Bendo et al. 2016). If we re-derive the free-free fractions assuming the canonical value of $\alpha_{\rm syn} = -0.8$ determined at 10 GHz (Niklas et al. 1997), the median free-free fraction of sources would be $f_{\rm ff} = 0.28$ and the synchrotron fraction would be $f_{\rm syn} = 0.72$.

On average the fractional contributions to the 93 GHz emission that we obtain assuming indices of $\alpha_{\rm ff} = -0.12$, $\alpha_{\rm syn} = -1.5$, and $\alpha_{\rm d} = 4$ appear to be reasonable and consistent. Our median free-free fraction of $f_{\rm ff} = 0.62$ with median absolute deviation of 0.29 agrees within error to the free-free fraction derived by (Bendo et al. 2016) of $f_{\rm ff} = 0.84 \pm 0.10$ at 86 GHz. Our median free-free fraction is also consistent with that derived for NGC 253, $f_{\rm ff} = 0.70 \pm 0.10$, at our same frequency but averaged over $30''$ (Bendo et al. 2015). Similarly, pilot survey results of the MUSTANG Galactic Plane Survey at 90 GHz and with parsec-scale resolution indicate that $> 80\%$ of the emission in (candidate) clusters is composed of synchotron or free-free emission (Ginsburg et al. 2020).

We can also perform another self-consistent check on the free-free emission at 93 GHz, by using the recombination line emission and assuming the candidate star clusters have uniform temperatures. A constant temperature implies a constant integrated recombination line to continuum ratio (see Equation 3.4). Letting $R_{\rm LC} \approx 35$ km s^{-1} (for a temperature of $T_e = 6000$ K), we plug in our measured line and continuum fluxes and solve for $f_{\rm ff}$. A median free-free fraction of $f_{\rm ff} = 0.85$ with median absolute deviation of 0.33 is found when considering all sources with detected recombination line emission. In comparison, the relation assumed for the in-band spectral index, also at $0.2''$ resolution, results in a median value of $f_{\rm ff} = 0.66$ and a median absolute deviation of 0.24. These values are in reasonable agreement given the uncertainties in the two methods.

Overall, the assumptions made in decomposing the emission at 93 GHz, especially regarding synchrotron emissions, likely affect individual clusters at the $\sim 30\%$ level. This is not enough to bias our overall results, but follow up observations at other frequencies would be extremely helpful.

The second potentially major source of uncertainty is the assumed age of the burst. We discuss in Section 3.5.2 how we arrive at an adopted age of the clusters of $\sim$5 Myr. The assumed age has a large impact on the stellar mass inferred from the ionizing photon rate. The ionizing photon output changes substantially (by a factor of 40 from a zero age main sequence to an age of 5 Myr) as the most massive stars explode as supernova. An uncertainty of $\sim$1 Myr about an age of 5 Myr of a star cluster results in an uncertainty in the inferred stellar mass by a factor of four. This is roughly included in the 0.4 dex uncertainty, though it would represent a systematic offset.

3.6.2 Super Star Clusters

The estimated properties (mass, size, age; Table 3.6) that we derive for these candidate star clusters meet the criteria for young, massive clusters (e.g., Portegies Zwart et al. 2010), and these sources can be considered super star clusters (SSCs). Stars clusters forming in high gas surface density environments may be able to acquire significant

amounts of mass before feedback effects (likely radiation pressure) can disrupt and/or disperse the cluster (e.g., Adamo & Bastian 2016). Super star clusters with stellar masses of $M_\star \gtrsim 10^5$ M$_\odot$ typically have high star formation efficiencies ($\varepsilon > 0.5$) and therefore remain bound.

With the mass surface densities that we estimate and the age of the burst equaling many multiples of the free-fall times, these star clusters will likely remain bound, at least initially. They are being born into a violent environment and clusters often still experience significant mortality after forming. Estimates of the virial mass, escape velocity and momentum driving determined through molecular line observations will help quantity their mass and initial dynamical state.

3.6.3 Cluster Mass Function

In Figure 3.11, we plot the cluster stellar-mass function of candidate SSCs in NGC 4945 and compare it with the cluster mass distributions in additional galaxies with SSC populations.

A power-law fit to the cluster mass distribution, of 22 sources down to $M_\star = 2.4 \times 10^5$ M$_\odot$, in NGC 4945 results in a slope of $\beta = -1.8 \pm 0.4$. Note that the fit to the data results in $\beta = -1.76 \pm 0.07$, but given the uncertainty in our mass estimates (0.4 dex), we adopt the more conservative figure. We include the SSCs of NGC 253, with stellar mass properties determined on similar spatial scales (∼2 pc resolution) and through H40α recombination lines for a zero-age main sequence population (Leroy et al. 2018; Mills et al. 2020). A power-law fit to the cluster mass distribution of NGC 253, including 12 star clusters down to $M_\star = 7.9 \times 10^4$ M$_\odot$, results in a slope of $\beta = -1.6 \pm 0.3$. We estimated the completeness limits by eye. The turn-off from power-law distributions likely includes non-physical effects, resulting from the depth/sensitivity of the observations as well as source confusion due to the high inclination in which we view the starburst regions (and which appears to be higher in NGC 253).

We also include recent results from a homogeneous analysis by Mok et al. (2020) of the cluster mass distributions of young ($\tau < 10$ Myr), massive clusters in six galaxies. As representative examples, we include the best fit power-laws from three systems: the Large Magellanic Clouds (LMC), M 51, and the Antennae System. A main result from Mok et al. (2020) is that the cluster mass functions across the six galaxies are consistent with slopes of $\beta = -2.0 \pm 0.3$. Our measurements for NGC 4945 and NGC 253 are consistent with these findings.

As we discuss in Section 3.6.6, we estimate the total stellar mass in the burst to be 8.5×10^7 M$_\odot$. Extending the best fit of our cluster mass function down to a cluster mass of $\sim 2 \times 10^4$ M$_\odot$ would account for the additional mass and correspondingly the additional ionizing photon luminosity. As we discuss in Section 3.6.4, if these lower mass clusters are present they would need to be located in a more-extended region than the super star clusters. Therefore the cluster mass distribution may not reach down to 2×10^4 M$_\odot$ in the region where the SSCs are located, but perhaps $\sim 4 \times 10^4$ M$_\odot$ if a third of the additional recombination line emission results from diffuse ionized gas.

3.6.4 Ionizing Photons and Diffuse Ionized Gas

Total ionizing photon rate of the starburst. We measure an H40α recombination line flux of (6.4 ± 1.0) Jy km s^{-1} integrated within the T2 aperture (see Figure 3.8) in the intermediate-configuration ($0.7''$) data. Assuming the temperature of $T_e = (6000 \pm 400)$ K from Section 3.5.3, the total ionized content, which accounts for the full star-formation rate of the starburst, yields an ionizing photon rate of $Q_{\rm T2} = (3.9 \pm 0.3) \times 10^{53}$ s^{-1}.

Ionizing photon rate from candidate super star clusters. The sum of the ionizing photon rate over all massive star clusters is $\approx 5.3 \times 10^{52}$ s^{-1}. The total integrated line flux of H40α in the extended-configuration ($0.2''$) data is $\int S\,{\rm dv} = (2.1 \pm 0.6)$ Jy km s^{-1} within an aperture (T1) which covers the clusters in the starburst region. With the temperature of $T_e = (6000 \pm 400)$ K, the integrated line flux corresponds to an ionizing photon rate of $Q_{\rm T1} = (1.2 \pm 0.4) \times 10^{53}$ s^{-1}. Given the large uncertainties derived for individual clusters, we consider this value consistent with the sum over individual clusters. Therefore, we conclude that 20% to 44% of the total ionizing photons in the starburst can be attributed to the candidate super star clusters identified in NGC 4945.

Low mass clusters. The intermediate-configuration ($0.7''$) data are sensitive to emission on larger physical scales, but they also reach deeper sensitivities per unit area. Consequently, emission from a distribution of many compact, low mass clusters could be traced in these data, but not in the extended-configuration data.

We do not find it likely that the deficit of line emission in the extended-configuration data can be attributed primarily to low mass clusters. Focusing only on the low resolution ($0.7''$) observations, the integrated line flux is greater in the larger aperture (T2) than in the aperture covering only the star clusters (T1), indicating that additional line emission originates outside of the area where massive clusters are forming. In order for low mass clusters to account for the difference, these would need to form in a more extended region than the bright SSCs that we see.

AGN and the circumnuclear disk. The circumnuclear disk, including the AGN, could be an origin for ionizing photons observed only on larger scales. The strongest recombination line emission is seen at low-resolution in this region yet no significant line detections are obtained towards Sources 17, 18, and 20. We test this scenario by extracting spectra from the $0.7''$ intermediate-configuration data and the $0.2''$ extended-configuration data in 7 non-overlapping apertures of $3''$ diameter consecutively distributed along the major axis of the starburst region. The ratios of the line flux in the intermediate- to extended-configuration data are all consistent within error; the ratios in the seven regions have a mean and standard deviation of 2.3 ± 0.5. This indicates that a deficit of line emission on the physical scales probed in the high resolution data is ubiquitous (and not unique to the region surrounding the AGN), and at most $\sim 20\%$ may originate from the circumnuclear disk and AGN.

Diffuse ionized gas on large-scales. Summarizing the information above, we find that up to $\sim$70% of the total radio recombination line flux may originate on scales larger than $\sim$100 pc; this emission cannot be directly attributed to the AGN. Ionizing radiation may be escaping the immediate surroundings of massive stars and reaching larger scales. The approximate age of the burst indicates clusters have had time to

shed their natal material, dust is not a significant contribution of their emission, low gas mass fractions have been determined, and the current free-fall times of the clusters indicate abundant time to expel gas. Despite these indicators, extinctions may still be considerable and patchy regions may help to leak ionizing photons. Note that if radiation is escaping from star clusters, this would also imply that the stellar masses (derived from the ionizing photon rates) are underestimated.

However, a missing 70% of line flux in the $0.2''$ resolution data should be considered as an upper limit. Given the low signal-to-noise ratio of these lines, artifacts in both the image and spectral domains can impact the properties of the line profiles. For example, incomplete *uv* coverage on the scales of the emission can result in incorrect deconvolution and thus systematic underestimates. Deeper observations at high resolution ($\sim 0.1''$) would allow radio recombination lines to be mapped out with adequate signal-to-noise ratio and obtain sensitivities approaching that of the low-resolution ($0.7''$) observations.

3.6.5 Role of the AGN

Although the Seyfert 2 AGN in NGC 4945 is one of the brightest in our sky at X-ray energies, obscuration by e.g., $A_{\rm V} \geq 160$ (Spoon et al. 2000) has prevented its observation at virtually all other wavelengths (also given the spatial resolutions observed). Our observations at 93 GHz provide another piece of evidence for its existence.

We detect a point source in the 93 GHz image with a deconvolved size determined by `PyBDSF` of FWHM $= 0.84 \pm 0.01$ pc. The flux of 9.7 ± 1.0 mJy that we extract from the source is $\sim 10\%$ of the total continuum ($\sim$120 mJy) flux in the $0.12''$ resolution data. The Band 3 spectral index we derive is $\alpha_{93} = -0.85 \pm 0.05$, which is consistent with freshly accelerated electrons emitting synchrotron radiation. However, we follow the procedure outlined in Section 3.3.3, assuming free-free emission may also contribute, and place a limit on the ionizing photon rate. The spectral index at 93 GHz corresponds to $f_{\rm ff} \geq 0.47 \pm 0.03$. If we assume a temperature of $T_e = 6000$ (30 000) K this would correspond to $Q_0 = 1.1\,(0.5) \times 10^{52}$ s^{-1}. The low level of escaping ionizing radiation is consistent with previous work from MIR line ratios (Spoon et al. 2000).

We place a limit on the ionizing photons that could be leaking into medium. The bolometric luminosity estimated from the X-ray luminosity is $L_{\rm bol} \approx 6 \times 10^{43}$ erg s^{-1}, which we determined using the relation for Seyfert AGN of $L_{\rm bol} \sim 20 L_{\rm X}$ (Hopkins et al. 2007). Given the bolometric luminosity, we can now estimate the total expected ionizing photon luminosity. We use the standard relation of Elvis et al. (1994), which posits $L_{\rm bol}/L_{\rm ion} \sim 3$ with an average ionizing photon energy of 113 eV. Therefore the expected ionizing photon luminosity of the AGN is $L_{\rm ion} \approx 2 \times 10^{43}$ erg s^{-1} $\approx 5 \times 10^9$ L$_\odot$, while our limit corresponds to $\approx 2 \times 10^{42}$ erg s^{-1}. This would imply that $< 10\%$ of the ionizing photon luminosity of the AGN is escaping into the surrounding nuclear starburst and creating ionized gas.

3.6.6 Total Burst of Star-formation

We can convert the total ionizing photon rate of the burst $Q_{\rm T2}$ (see Section 3.6.4), as measured in the T2 region, to a luminosity. Assuming an average ionizing photon energy of $\langle h\nu \rangle \approx 17$ eV, the ionizing luminosity is $L_{\rm T2} \approx 1.1 \times 10^{43}$ erg s^{-1} $= 2.8 \times 10^9$ L$_\odot$. A ratio of bolometric luminosity to ionizing photon luminosity $L_{\rm bol}/L_0 \sim 14$ is predicted for a 5 Myr old population, using Starburst99 (Leitherer et al. 1999). Thus, we can expect a bolometric luminosity of 4×10^{10} L$_\odot$ which agrees within error to the bolometric luminosity derived from FIR observations, $L_{\rm bol} = (2.0 \pm 0.2) \times 10^{10}$ L$_\odot$ (Bendo et al. 2016).

We note that, if we had adopted a lifetime for the burst of 4 Myr rather than 5 Myr, Starburst99 calculations expect a bolometric luminosity of $L_{\rm bol} \approx 2 \times 10^{10}$ L$_\odot$. Also, this calculation assumes the source of the additional recombination line emission in the intermediate-configuration data is also characterized by a 5 Myr old stellar population.

The ratio of bolometric luminosity to mass of a 5 Myr old burst is estimated at $\Psi \sim 470$ L$_\odot$/M$_\odot$. With an $L_{\rm bol} = 4 \times 10^{10}$ L$_\odot$, the inferred total stellar mass of the population is $\sim 8.5 \times 10^7$ M$_\odot$. The total stellar mass in our candidate clusters is $\approx 1.1 \times 10^7$ M$_\odot$, or ∼13% of the expected total stellar mass.

3.6.7 Star Clusters and the Central Wind

An outflow of warm ionized gas in NGC 4945 has been modeled as a biconical outflow with a deprojected velocity of v ≈ 525 km s^{-1} with emission out to 1.8 kpc (Heckman et al. 1990). The outflow has been traced in [NII], [SII] and Hα in multiple analyses (Heckman et al. 1990; Moorwood et al. 1996; Mingozzi et al. 2019). Heckman et al. (1990) estimated a total energy of 2×10^{55} erg and momentum flux of 9×10^{33} dyne $=$ 1400 M$_\odot$ km s^{-1} yr^{-1} (after correcting for the updated distance). These values do not include possible contributions by colder phases, which are likely associated with the NGC 4945 outflow (Bolatto et al., in prep.). While these values are uncertain, they are useful for an order of magnitude comparison with expected properties of the star clusters.

Gas reaching 1.8 kpc and moving at 525 km s^{-1} would have been launched ∼3 Myr ago assuming a constant velocity, which is within the time-frame of the burst. A ∼3 Myr time-scale and a velocity of 525 km s^{-1} would put 8×10^6 M$_\odot$ of warm ionized gas mass into the outflow. Using Starburst99 (Leitherer et al. 1999), an estimated mechanical luminosity of $\sim 3 \times 10^{41}$ erg s^{-1} from the clusters, fairly constant over the 5 Myr age, equates to an injected energy of $\approx 6 \times 10^{55}$ erg. It would require 30% of the expected mechanical energy output of the clusters to drive the ionized outflow. From simulations, the total momentum supplied to the ISM per supernova is expected to be 2.8×10^5 M$_\odot$ km s^{-1} (Kim & Ostriker 2015, and references therein), with stellar winds contributing an additional 50% to the momentum. Starburst99 predicts the supernova rate to be fairly constant at ∼0.009 yr^{-1} for these clusters. From these values, we estimate that the momentum supplied to the ISM by the candidate star clusters would be $\sim$ 3800 M$_\odot$ km s^{-1} yr^{-1}. It would require about 40% of the expected momentum output of the clusters to drive the ionized outflow. Therefore, neither the energetics nor the inferred momentum of the outflow prevent it from being

driven solely by the star clusters in NGC 4945, although there are very considerable uncertainties associated with this calculation.

3.6.8 Comparison with NGC 253

NGC 253 is a nearly edge-on galaxy located nearby (3.5 ± 0.2 Mpc ; Rekola et al. 2005) with similar properties as NGC 4945. It hosts a central starburst spanning ∼200 pc with young, massive clusters (Leroy et al. 2018; Mills et al. 2020). A major difference between these two galaxies is that NGC 4945 shows unambiguous signatures of harboring an active super-massive black hole. Since NGC 4945 is just the second analysis we have undertaken with millimeter wavelength ALMA observations on scales which resolve the clusters, it is relevant to compare the properties of their super stars clusters.

Analyses of the cluster population of NGC 253 at ∼2 pc resolution with ALMA (Leroy et al. 2018; Mills et al. 2020) revealed 18 sources as (proto-)super star clusters which are in the process of forming or close to a zero age main sequence (∼1 Myr) (see also Rico-Villas et al. 2020). Overall the properties of the clusters are strikingly similar as those in NGC 4945; they have a median stellar mass of 1.4×10^5 $M_\odot$ and FWHM sizes of 2.5–4 pc. The star clusters of NGC 253 boast slightly higher ionizing photon luminosities and larger gas fractions $M_{\rm gas}/M_{\rm tot} \sim 0.5$, reflecting a slightly younger age. The slopes of the cluster mass function we derive are also consistent (see Section 3.6.3), just offset by a factor of ∼3.5 in mass. As in NGC 4945, at least 30% of the of the nuclear starburst of NGC 253 originates in a clustered mode of star-formation.

While we have not estimated some of the dynamical properties of NGC 4945, the presence of broad recombination line emission in four sources in NGC 253 indicates the star clusters are operating under similar processes. In NGC 253, the sources are young enough that feedback has not managed to unbind a large fraction of the gas from the clusters. The slightly higher total mass surface densities and smaller free-fall times in NGC 4945 indicate its clusters might be surviving a young, disruptive stage.

The concerted feedback of the young (< 10 Myr), forming SSCs identified in NGC 253 (Leroy et al. 2018) are likely not responsible for the starburst driven outflow, as they are expected to impart momentum that is a factor of 10-100 lower than the outflow momentum estimated in CO (Bolatto et al. 2013b; Krieger et al. 2019). Some (global) event may be initiating the active formation of star clusters. On the other hand, the star clusters in NGC 4945 could potentially influence the outflow of warm ionized gas; some event appears to be inhibiting the formation of new star clusters.

3.7 Summary

Massive, clustered star-formation is an efficient and possibly common mode of star-formation in high gas density environments. Nearby galaxies with bursts of star-formation in the central $\mathcal{O}(100$ pc) are local laboratories to study this mode, and for NGC 4945, in the presence of a Seyfert AGN. High levels of dust extinction ($A_V \sim 40$) in the nearly edge-on ($i \sim 72°$) central region of NGC 4945 have previously prevented

the direct observation and characterization of its massive star clusters.

We identify 27 super star cluster candidates in the central starburst of NGC 4945. We derive properties of the candidate clusters through ALMA observations at 2.2 pc resolution of the 93 GHz (3 mm) free-free emission and hydrogen recombination line emission (H40α and H42α) arising in photo-ionized gas. We also use and present high-resolution 350 GHz imaging of the dust continuum observed with ALMA, and supplement our analysis with 2.3 GHz continuum imaging which primarily traces synchrotron emission (Lenc & Tingay 2009).

Our results are as follows:

- The 27 point sources identified ($\geq 10\sigma$) in 93 GHz continuum emission as candidate super star clusters have FWHM sizes of 1.4–4.0 pc. The 93 GHz emission in these bright, compact regions is dominated by free-free emission, with a median free-free fraction of $f_{\rm ff} = 0.62$. Synchrotron emission from recent supernova remnants contributes to the 93 GHz emission with a median fraction of $f_{\rm syn} = 0.36$. Substantial dust emission is found in three sources.

- We average the spectra of the H40α and H42α recombination lines to synthesize an effective H41α profile. Recombination line emission is detected in 15 candidate clusters, generally with narrow (FWHM ~ 36 km s^{-1}) line widths. Six of the detected sources have significant ($f_{\rm syn} \gtrsim 0.5$) synchrotron emission; three of those have broad line widths with FWHM > 105 km s^{-1}.

- We estimate an electron temperature of $T_e = (6000 \pm 400)$ K of the ionized gas using the flux ratio of the integrated line to free-free continuum. This electron temperature implies an average metallicity of $12 + \log_{10}({\rm O/H}) = 8.9 \pm 0.1$ surrounding these young massive stars. The ionized gas densities of $\log_{10}(n_e/{\rm cm}^{-3}) = 3.1$–$3.9$, that we derive are typical of classic H II regions. The ionized gas masses of the clusters are typically $< 1\%$ of the estimated stellar mass.

- We determine ionizing photon rates of the candidate SSCs in the range $\log_{10}(Q_0/{\rm s}^{-1}) \sim 50.4$ – 51.8. Adopting an age of $\sim$5 Myr, the stellar masses implied by the ionizing photon rates are $\log_{10}(M_\star/{\rm M}_\odot) \sim 4.7$–$6.1$. The sum of the stellar masses of the candidate SSCs is $\approx 1 \times 10^7$ M$_\odot$. The uncertainties on these measurements are 0.4 dex. We discuss the age estimate in Section 3.5.2.

- We fit the cluster stellar-mass distribution and find a slope of $\beta = -1.8 \pm 0.4$. The slope of the fit is consistent with our fit to the candidate SSCs in the central starburst of NGC 253 (Mills et al. 2020) and to recent findings in the LMC, M 51, and the Antennae system (Mok et al. 2020).

- We estimate the gas mass of the candidate clusters from dust emission. The total mass, $M_{\rm Tot}$, is evaluated by the combined stellar and gas masses. Gas mass fractions range from $M_{\rm gas}/M_{\rm Tot} = 0.04$–$0.22$, with a median value of 0.13. We calculate the total-mass surface density of the clusters and find a median value of $\Sigma_{\rm Tot} = 3 \times 10^4$ M$_\odot$ pc^{-2}. The median free-fall timescale is 0.04 Myr.

- With low-resolution ($0.7''$) observations of the H40α recombination line, we measure a total ionizing photon rate of the burst of $Q_0 = (3.9 \pm 0.3) \times 10^{53}$ s^{-1}. The candidate star clusters that we analyze contribute 20–44% of the total ionizing photon rate. Additional recombination line emission present in the low-resolution data appears to be ubiquitous throughout the starburst region, cannot be directly attributed to the AGN, and is also found outside of the area where the SSCs are located. Diffuse ionized gas may be responsible for some of the additional emission, although low mass clusters and distributed star formation are expected to also contribute. This indicates that ionizing radiation may be escaping the immediate surroundings of massive stars and reaching the larger (> 100 pc) scales traced at lower resolution.

- We compare the candidate SSCs in NGC 4945 with the (proto-)SSCs recently identified in the central starburst of NGC 253 (Leroy et al. 2018; Mills et al. 2020). The age of the burst (1 Myr) is a bit younger in NGC 253, the gas mass fractions ($\sim$0.5) are a bit higher, and the stellar mass of the clusters are slightly smaller (factor of 3.5). While the actively forming clusters are not a major contributor to the starburst driven outflow in NGC 253, the slightly older population of the star clusters in NGC 4945 may contribute to driving a nuclear outflow of warm ionized gas.

- Strong, variable X-ray emission, which is Compton thick, provides evidence for a Seyfert AGN in NGC 4945. The bright, 93 GHz source which we presume to be the AGN is point-like in our 2.2 pc beam. We measure its in-band spectral index at 93 GHz to be $\alpha_{93} = -0.85 \pm 0.05$, likely dominated by synchrotron emission. We do not detect recombination line emission from this point source. Our observations support previous findings in which UV ionizing radiation from the AGN is heavily obscured in all directions. We estimate an upper limit for its escaping ionizing photon rate of $Q_0 < 1 \times 10^{52}$ s^{-1}, which is $\lesssim 10\%$ of the expected luminosity of ionizing photons for a typical Seyfert AGN.

- Lastly, we report on a shortcoming of H42α as a "low-resolution" tracer of ionizing radiation. When observed with low spatial resolution and/or from broad line components, we find this spectral line is likely contaminated by line emission from other species, resulting in the recombination line flux (and star-formation rates derived from it) being overestimated by a factor of $\sim$2, when compared to a similar analysis of H40α.

Acknowledgments

The authors thank the anonymous referee for the careful review of the article and helpful input. We also thank Emil Lenc for providing the Australian LBA 2.3 GHz data and Paul van der Werf for the HST Pa-α data.

KLE acknowledges financial support from the Netherlands Organization for Scientific Research (NWO) through TOP grant 614.001.351. KLE thanks the Laboratory for Millimeter Wave Astronomy and the Department of Astronomy at the University

of Maryland and the Green Bank Observatory for hosting her during completion of this work. ADB and RCL acknowledge support from NSF through grants AST-1412419 and AST-1615960. AKL acknowledges support by the National Science Foundation (NSF) under Grants No.1615105, 1615109, and 1653300, as well as by the National Aeronautics and Space Administration (NASA) under ADAP grants NNX16AF48G and NNX17AF39G. EACM gratefully acknowledges support by the National Science Foundation under grant No. AST-1813765. ER acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), funding reference number RGPIN-2017-03987.

This article makes use of the following ALMA data: ADS/JAO.ALMA#2018.1.01236.S, and ADS/JAO.ALMA#2016.1.01135.S. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada), NSC and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

Facilities: ALMA, Australian LBA, HST

Software: APLpy (Robitaille & Bressert 2012), Astropy (The Astropy Collaboration 2018), CASA (McMullin et al. 2007), CRRLpy (Salas et al. 2016), Matplotlib (Hunter 2007), PyBDSF (Mohan & Rafferty 2015)

3.A Millimeter Wavelength Emission from Free-free Continuum and Recombination Lines

In this section we present the relations for the recombination line intensity and free-free continuum applicable for millimeter wavelength emission, largely by bringing together information derived in Gordon & Sorochenko (2002) and Draine (2011). We show how these relations can be used to estimate the temperature and ionizing photon rate of the emitting plasma.

At frequencies of $\mathcal{O}(10$ GHz), the electron temperature T_e of the thermal emission from ionized gas accurately characterizes the relative populations of electrons in bound atomic levels, and the system is well approximated by local thermodynamic equilibrium (LTE). Collisions dominate and line emission is described by the Boltzmann distribution.

As the collisional cross section of a Bohr atom rapidly decreases ($\propto \mathsf{n}^4$) towards smaller principal quantum numbers, or higher frequencies of $\mathcal{O}(100$ GHz), collisions become less important, and the observed line intensity is not exactly set by the kinetic motion of the electrons. Radiative processes, which dominate the smallest principal quantum numbers, influence the level populations; smaller n are underpopulated as compared with a Boltzmann distribution. The population rates into an energy level are not exactly balanced by rates out of the level, and a correction coefficient (b_n) for a departure from LTE is necessary.

3.A.1 Recombination Line Intensity

Considering Kirchoff's law of thermodynamics with no net change in intensity through the medium, the emission of a radio recombination line in LTE, $S_{\mathsf{n}}(\mathrm{LTE})$, is related to κ_{n}, the fractional absorption per unit pathlength ℓ, for small line optical depths by

$$S_{\mathsf{n}}(\mathrm{LTE}) \approx \kappa_{\mathsf{n}}\, \ell\, B_\nu(T_e)\, \Omega \tag{3.10}$$

for a transition to final principal quantum number n, where B_ν is the Planck function and Ω is the solid angle on the sky. For small optical depths of the line and the free-free continuum τ_{c}, the non-LTE line flux density is, by definition,

$$S_{\mathsf{n}} \approx S_{\mathsf{n}}(LTE) b_{\mathsf{n}+1} \left(1 + \frac{\tau_C}{2}\beta\right). \tag{3.11}$$

Here, $b_{\mathsf{n}+1}$ is the departure coefficient, which is defined as $b_{\mathsf{n}} = n/n_{\mathrm{LTE}}$, the ratio of the actual number density of atoms with an electron in level n to the number which would be there if the population were in LTE at the temperature of the ionized gas, such that for LTE, $b_{\mathsf{n}} = 1$. β is the departure coefficient which accounts for stimulated emission. The second term in parenthesis is usually negligible at millimeter wavelengths, such that $(1 + 0.5\tau_c\beta) \approx 1$. In Eq. 3.14 of Section 3.A.2, the absorption coefficient of the free-free continuum is given and the optical depth can be derived to show that typical parameters (e.g. $EM_l = 10^7$ cm^{-6} pc and $T_e = 10^4$ K) of ionized regions around massive stars result in $\tau_c \lesssim 3 \times 10^{-6}$ at 100 GHz. Typical values of β for $\mathsf{n} \gtrsim 40$ are $|\beta| \lesssim 40$ (e.g., Storey & Hummer 1995).

In its expanded form, the LTE absorption coefficient of a radio recombination line can be expressed in terms of the Saha-Boltzmann distribution, oscillator strength, and line frequency as

$$\kappa_{\mathsf{n}} \approx n_e n_+ \frac{2}{\sqrt{\pi}} \frac{e^2}{m_e} \left(\frac{h^2}{2 m_e k_B}\right)^{3/2} \frac{h}{k_B}\, T_e^{-2.5} R_H Z^2 \Delta\mathsf{n} \left(1 - \frac{3}{2}\frac{\Delta\mathsf{n}}{\mathsf{n}}\right) \exp\left(\frac{\chi_{\mathsf{n}}}{k_B T_e}\right) \times M(\Delta\mathsf{n}) \left(1 + \frac{3}{2}\frac{\Delta\mathsf{n}}{\mathsf{n}}\right) \phi_\nu \tag{3.12}$$

where n_e and n_+ are the number densities of electrons and ions respectively; R_H is the Rydberg constant for hydrogen; Z is the effective nuclear charge; $\Delta\mathsf{n}$ is the change in energy levels of the given transition; χ_{n} is the energy required to ionized the atom from state n, but $\exp(\chi_{\mathsf{n}}/k_B T_e)$ is small (< 1.02) for $\mathsf{n} \geq 40$ and typical ionized gas temperatures; $M(\Delta\mathsf{n} = 1, 2) = 0.190775, 0.026332$ is an approximation factor for the oscillator strength; and, ϕ_ν is the line profile (normalization; in SI units, Hz^{-1}) such that $\int_{-\infty}^{\infty} \phi_\nu \, \mathrm{d}\nu = 1$.

Bringing the above equations together, using the Rayleigh-Jeans approximation of the Planck function, and integrating over the line profile, we arrive at the expression for the integrated line flux density,

$$\int S_{\mathsf{n}} \, \mathrm{dv} = \left(65.13 \text{ mJy km s}^{-1}\right) b_{\mathsf{n}+1} \left(\frac{n_e n_p V}{5 \times 10^8 \text{ cm}^{-6} \text{ pc}^3}\right) \left(\frac{D}{3.8 \text{ Mpc}}\right)^{-2} \times \left(\frac{T_e}{10^4 \text{ K}}\right)^{-1.5} \left(\frac{\nu}{100 \text{ GHz}}\right) \tag{3.13}$$

where we take $n_+ = n_p$, $Z = 1$, and $\Delta \mathsf{n} = 1$. For convenience and self consistency, we express the emission measure in terms of source volume, V, and the solid angle in terms of distance to the source, D. Let $\Omega = \frac{\pi r^2}{D^2}$, where r is the radius of the region, and if we let $\ell = \frac{4}{3} r$, the volumetric emission measure is given as $EM_V = n_e n_p V$ where $V = \frac{4}{3}\pi r^3$.

3.A.2 Continuum Intensity

The absorption coefficient for free-free continuum (Oster 1961) in the Rayleigh-Jeans limit is, as a function of frequency,

$$\kappa_{\rm c} = \frac{n_e n_+}{\nu^2} \frac{8 Z^2 e^6}{3\sqrt{3} m_e^3 c} \left(\frac{\pi}{2}\right)^{1/2} \left(\frac{m_e}{k_B T_e}\right)^{3/2} g_{\rm ff} \tag{3.14}$$

where the gaunt free-free factor is (Draine 2011)

$$g_{\rm ff} \approx 13.91 \left(Z \frac{\nu}{\rm Hz}\right)^{-0.118} \left(\frac{T_e}{\rm K}\right)^{0.177} \tag{3.15}$$

valid for $\nu_p \ll \nu \ll k T_e / h$ where the plasma frequency is $\nu_p = 8.98 (n_e / {\rm cm}^{-3})^{1/2}$ kHz, and to within 10% when $1.4 \times 10^{-4} < Z \nu T_e^{-1.5} < 0.25$.

The continuum intensity for an optically thin medium is

$$S_{\rm c} \approx B_\nu \, \kappa_{\rm c} \, \ell \, \Omega. \tag{3.16}$$

As we did for the line intensity, we can also express the above relation in terms of distance D and the volumetric emission measure of the region $EM_V = n_e n_p V$ where $V = \frac{4}{3}\pi r^3$:

$$S_c = (2.080 \text{ mJy}) \, Z^{1.882} \left(\frac{n_e n_+ V}{5 \times 10^8 \text{ cm}^{-6} \text{ pc}^3}\right) \left(\frac{D}{3.8 \text{ Mpc}}\right)^{-2} \left(\frac{T_e}{10^4 \text{ K}}\right)^{-0.323} \times \left(\frac{\nu}{100 \text{ GHz}}\right)^{-0.118} \tag{3.17}$$

evaluated in the Rayleigh-Jeans limit.

3.A.3 Physical Properties

Line to Continuum Ratio and Temperature

Using the previous expressions, the ratio between the RRL and continuum flux density is:

$$\frac{\int S_{\mathsf{n}} \, {\rm dv}}{S_{\rm c}} = \left(31.31 \text{ km s}^{-1}\right) b_{\mathsf{n}+1} \left(\frac{T_e}{10^4 \text{ K}}\right)^{-1.177} (1+y)^{-1} \left(\frac{\nu}{100 \text{ GHz}}\right)^{1.118} \tag{3.18}$$

where $y = n_{\rm He^+} / n_p$ is the ratio of singly ionized helium to hydrogen by number and $n_{\rm He^+}$ is the singly ionized helium number density. If we assume the emission region

is composed of only hydrogen and singly ionized helium, then

$$\frac{n_p}{n_+} = \frac{n_p}{n_p + n_{\mathrm{He^+}}} = \left(1 + \frac{n_{\mathrm{He^+}}}{n_p}\right)^{-1} = (1+y)^{-1}\,. \tag{3.19}$$

We can rearrange the integrated RRL line to continuum ratio and solve for the electron temperature T_e of the emission region,

$$T_e = 10^4\ \mathrm{K}\left[b_{\mathrm{n+1}}\,(1+y)^{-1}\left(\frac{R_{\mathrm{lc}}}{31.31\ \mathrm{km\ s^{-1}}}\right)^{-1}\left(\frac{\nu}{100\ \mathrm{GHz}}\right)^{1.118}\right]^{0.85} \tag{3.20}$$

where we denote the integrated RRL line to continuum ratio as $R_{\mathrm{lc}} = \frac{\int S_{\mathrm{n}}\ \mathrm{dv}}{S_{\mathrm{c}}}$.

Ionizing Photon Rate

The rate of ionizing photons ($E > 13.6eV$) is given by,

$$Q_0 = n_e n_+ V \alpha_B \tag{3.21}$$

where α_B is the case B recombination coefficient (Draine 2011),

$$\alpha_B = 2.59 \times 10^{-13}\ \mathrm{cm^3\ s^{-1}}\left(\frac{T_e}{10^4\ \mathrm{K}}\right)^{-0.833-0.034\ln\left(T_e\ /\ 10^4\ \mathrm{K}\right)} \tag{3.22}$$

which is valid for 3000 K $< T_e <$ 30,000 K. Thus we have,

$$Q_0 = \left(3.805 \times 10^{51}\ \mathrm{s^{-1}}\right)\left(\frac{n_e n_+ V}{5 \times 10^8\ \mathrm{cm^{-6}\ pc^3}}\right)\left(\frac{T_e}{10^4\ \mathrm{K}}\right)^{-0.833-0.034\ln\left(T_e\ /\ 10^4\ \mathrm{K}\right)}. \tag{3.23}$$

3.B Recombination Line Spectra of all Sources

In Figure 3.12, we show that the central velocities of our detected recombination lines are in good agreement with the kinematic velocity expected of the disk rotation, though we include the recombination line spectra for all 29 sources (including those that are not significantly detected). The recombination line spectra from our sources are overlaid on H40α spectra extracted from the intermediate configuration observations (0.7″ resolution). High-resolution recombination line emission is coincident with emission from the intermediate configuration data.

3.C Spectral Energy distribution of all Sources

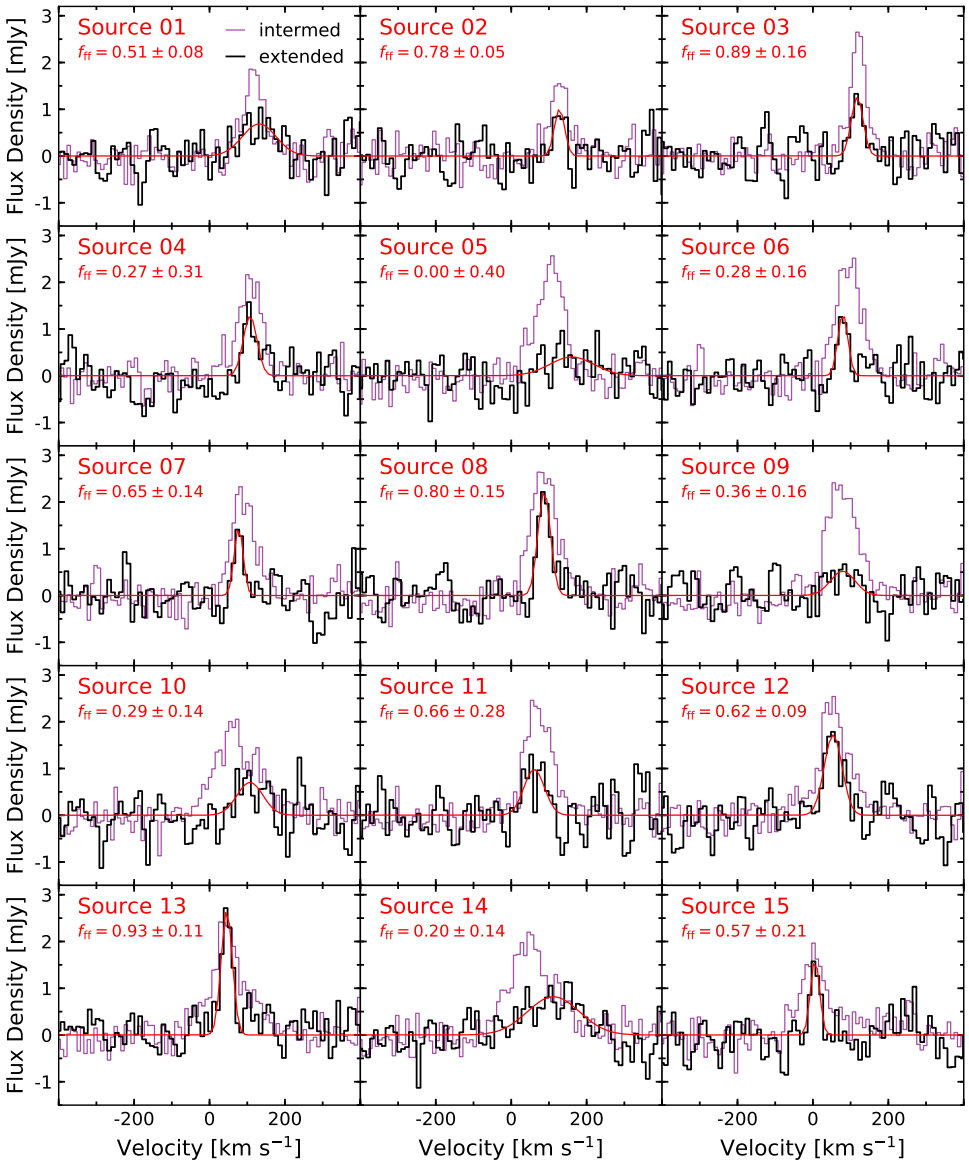


Figure 3.12: *(cont.)*

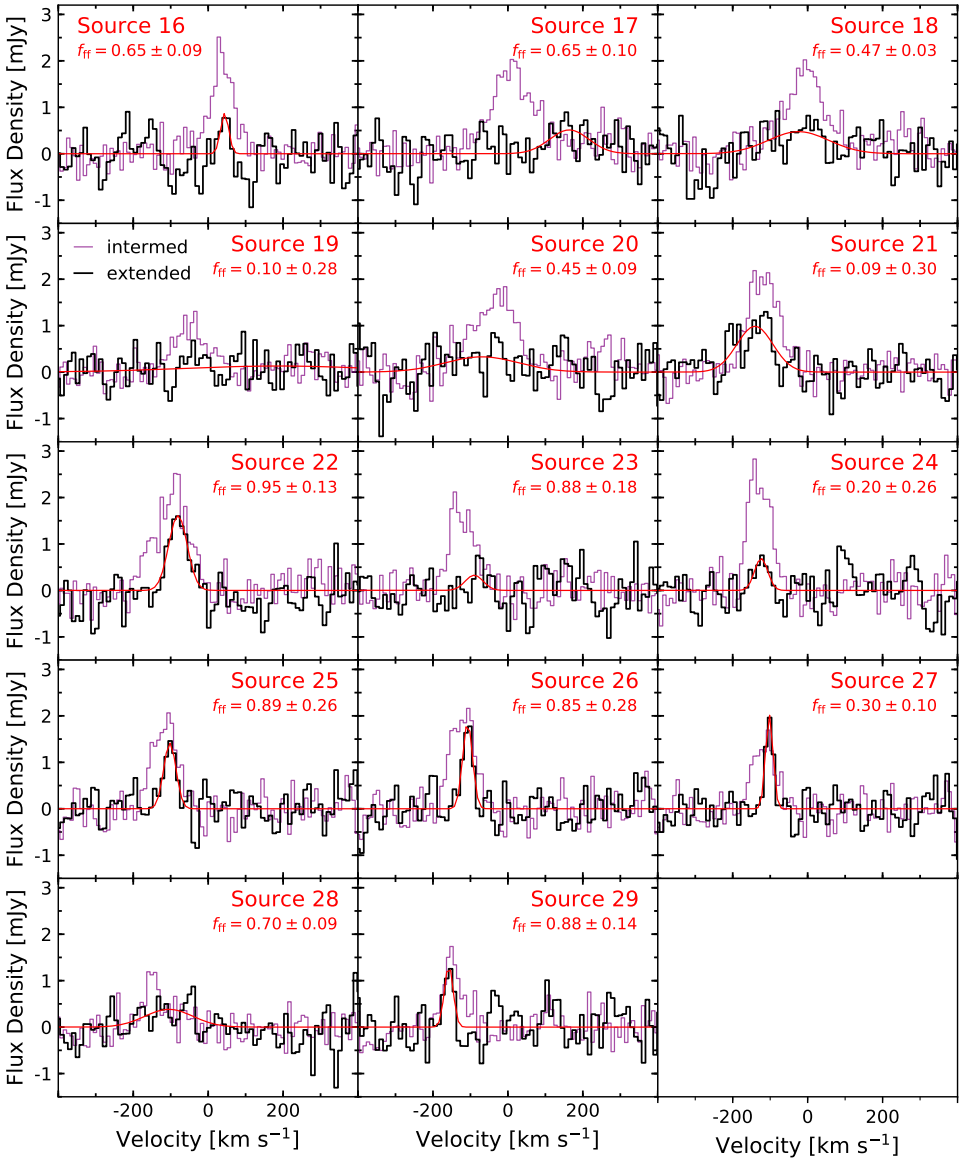


Figure 3.12: Effective H41α recombination line spectra (thick, black line) for sources with significantly detected emission and the best fit line profile (red) – same as in Figure 3.6. H40α line emission tracing the kinematics is overlaid in thin purple; it represents the H40α spectrum centered at the same source locations and extracted from low resolution ($\sim$0.7″) observations.

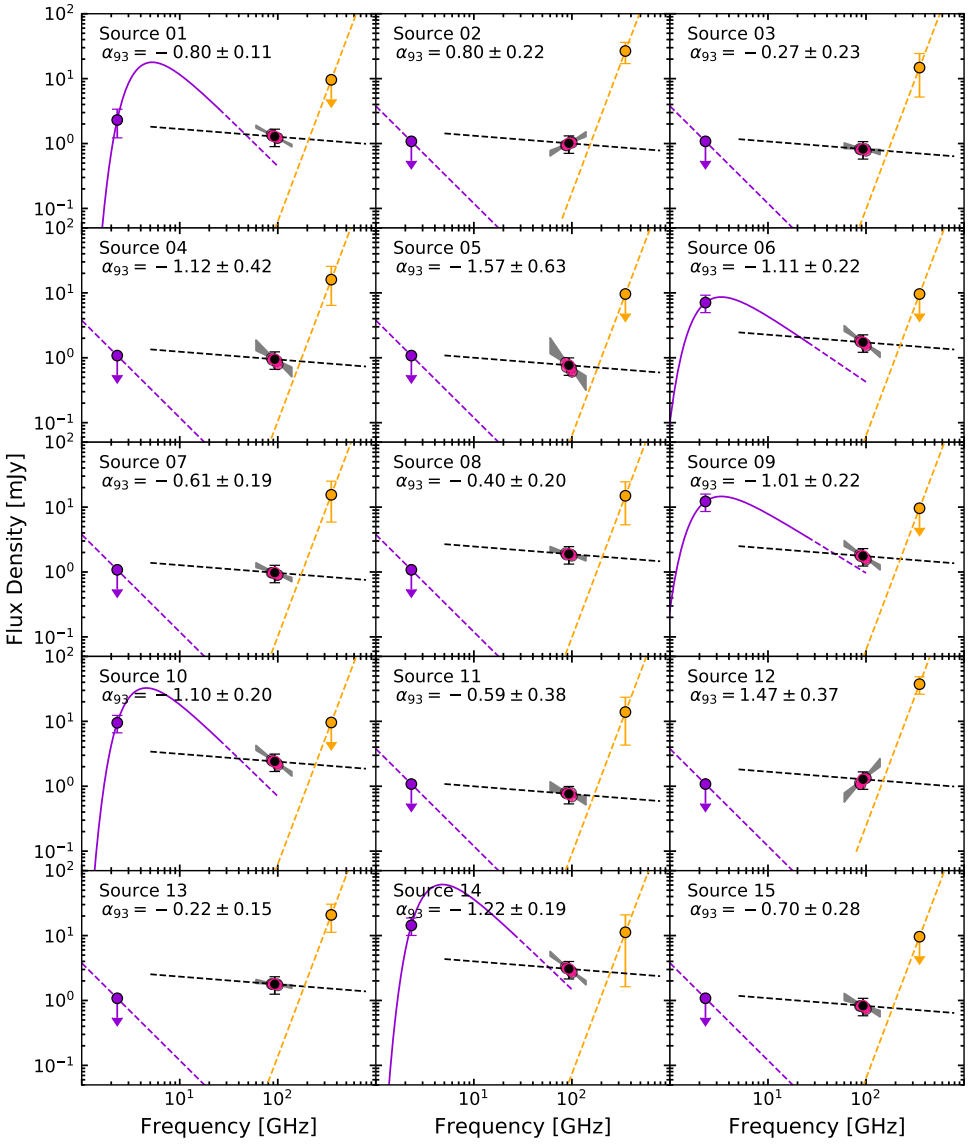


Figure 3.13: *(cont.)*

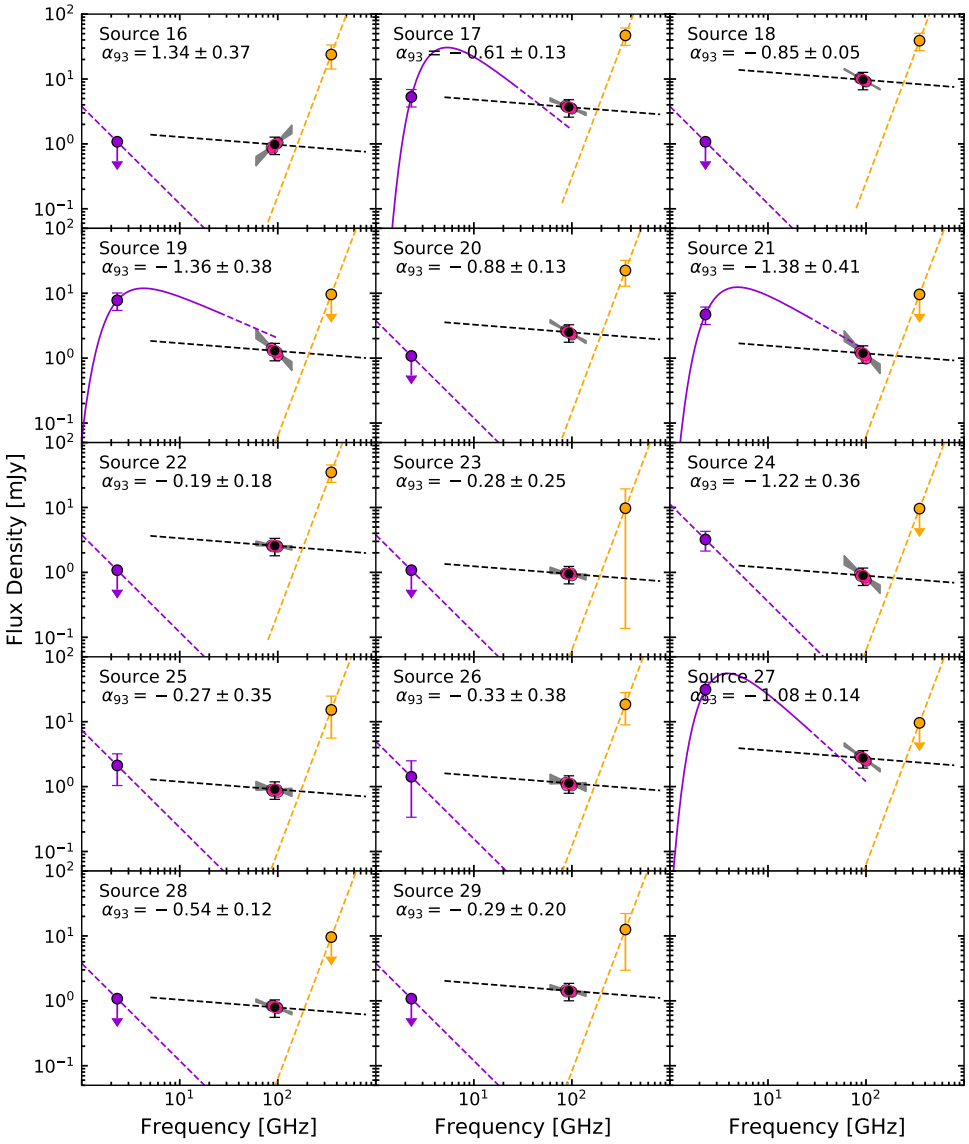


Figure 3.13: SEDs constructed for each source, same as in Figure 3.5. The dashed green line represents a dust spectral index of $\alpha = 4.0$, normalized to the flux density we extract at 350 GHz (orange data point). The dashed black line represents a free-free spectral index of $\alpha = -0.12$, normalized to the flux density we extract at 93 GHz (black data point). The white data points show the flux densities extracted from the band 3 spectral windows. The gray shaded region is the 1σ error range of the band 3 spectral index fit, except we have extended the fit in frequency for displaying purposes. The purple dashed line represents a synchrotron spectral index of $\alpha = -1.5$, normalized to the flux density we extract at 2.3 GHz (purple data point); except for Source 14 where the solid purple line represents the matched 2.3–23 GHz fit. Error bars on the flux density data points are 3σ.

4 | The first detection of radio recombination lines at cosmological distances

Abstract

Context. Recombination lines involving high principal quantum numbers ($\mathsf{n} \sim 50 - 1000$) populate the radio spectrum in large numbers. Low-frequency ($<$ 1 GHz) observations of radio recombination lines (RRLs) primarily from carbon and hydrogen offer a new, if not unique, way to probe cold, largely atomic gas and warm, ionised gas in other galaxies. Furthermore, these observations can be used to determine the physical state of the emitting regions, such as temperature and density. These properties make RRLs, potentially, a powerful tool of extragalactic interstellar medium (ISM) physics. At low radio frequencies, it is conceivable to detect RRLs out to cosmological distances when illuminated by a strong radio continuum. However, they are extremely faint ($\tau_{\rm peak} \sim 10^{-3} - 10^{-4}$) and have so far eluded detection outside of the local universe.

Aims. With observations of the radio quasar 3C 190 ($z = 1.1946$), we aim to demonstrate that the ISM can be explored out to great distances through low-frequency RRLs.

Methods. 3C 190 was observed with the High Band Antennas (HBA) of the LOw Frequency ARray (LOFAR) and processed using newly developed techniques for spectral analysis.

Results. We report the detection of RRLs in the frequency range 112 MHz – 163 MHz in the spectrum of 3C 190. Stacking 13 α-transitions with principal quantum numbers $\mathsf{n} = 266 - 301$, a peak 6σ feature of optical depth $\tau_{\rm peak} = (1.0 \pm 0.2) \times 10^{-3}$ and FWHM $= 31.2 \pm 8.3$ km s^{-1} was found at $z = 1.124$. This corresponds to a velocity offset of -9965 km s^{-1} with respect to the systemic redshift of 3C 190.

Conclusion. We consider three interpretations of the origin of the RRL emission: an intervening dwarf-like galaxy, an active galactic nucleus-driven (AGN) outflow, and the inter-galactic medium. We argue that the recombination lines most likely originate in a dwarf-like galaxy ($M \sim 10^9$ M$_\odot$) along the line of sight, although we cannot rule out an AGN-driven outflow. We do find the RRLs to be inconsistent with an inter-galactic medium origin. With this detection, we have opened up a new way

to study the physical properties of cool, diffuse gas out to cosmological distances.

K. L. Emig, P. Salas, F. de Gasperin, J. B. R. Oonk, M. C. Toribio,
H. J. A. Röttgering, and A. G. G. M. Tielens
Astronomy & Astrophysics 622, 7 (2019)

4.1 Introduction

Under typical conditions found in the interstellar medium (ISM), the recombination of electrons with singly ionised atoms can result in appreciable level populations at high principal quantum numbers (e.g. $\mathsf{n} \sim 300$). For these Rydberg atoms, the α-transitions (i.e. $\Delta\mathsf{n} = 1$) are so low in energy that they are observable at radio frequencies.

Radio recombination lines (RRLs) that have transitions stimulated by a radiation field have been observed at frequencies $\nu \lesssim 10$ GHz. The integrated strength of RRLs measured as a function of quantum number, and likewise frequency, is highly dependent upon the physical conditions of the gas. With almost 800 spectral lines (per each element) between 10 MHz and 10 GHz, this trait makes them powerful tools for understanding the physical properties of the medium, such as temperature, density, thermal pressure, and cloud size.

Within our own Galaxy, stimulated RRLs have been found in regions dominated by either hydrogen or carbon emission. In largely ionised gas with temperatures of $T_e \sim 8000$ K and densities of $n_e \sim 1$ cm^{-3} (e.g. Anantharamaiah 1985b), hydrogen RRLs are brightest. This gas phase peaks in intensity at frequencies between 250 MHz and 1000 MHz (Roshi & Anantharamaiah 2000; Zhao et al. 1996; Shaver et al. 1978; Pedlar et al. 1978). Additionally, cold ($T_e \sim 100$ K) yet diffuse ($n_e \sim 0.05$ cm^{-3}) gas can result in carbon being highly stimulated (Shaver 1975a). This gas phase occurs in the presence of a radiation field capable of ionising carbon (ionisation potential of 11.3 eV) yet not hydrogen (13.6 eV). With level populations greatly enhanced by dielectronic capture (Watson et al. 1980), carbon RRLs are the most prominent emitters at $\nu \lesssim 250$ MHz (Konovalenko & Sodin 1980; Payne et al. 1989; Oonk et al. 2017; Salas et al. 2018).

The detection of low-frequency RRLs is greatly aided towards bright radio sources as the intensity of stimulated transitions is proportional to the strength of the radio continuum, unlike the spontaneous transitions of higher frequency recombination lines associated with HII regions (e.g. Zuckerman & Ball 1974). This distinction between stimulated and spontaneous transitions is important because it indicates that low-frequency RRLs can be studied out to cosmological distances with bright radio sources. The advantages of using stimulated recombination lines to study a variety of regimes was quickly realised by Shaver (1978), including (i) the study of ionised gas in normal galaxies; (ii) the study of the physical conditions in the nuclei of galaxies, quasars and absorption-line systems; (iii) the assessment of the importance of free-free absorption in the spectra of extragalactic radio sources; and (iv) redshift determination for unidentified radio sources.

Out of the 15 extragalactic sources that have been detected via RRLs – all at $\nu > 1$ GHz and from nearby galaxies (e.g. for review see Gordon & Sorochenko 2002; Roy et al. 2008) – only the spectrum of M 82 shows clear evidence of stimulation-dominated emission (Shaver et al. 1978). Non-local sources, such as quasars, were searched at 4.8 GHz, but went undetected (Bell et al. 1984). Stifled by instrument capabilities below ∼1 GHz, few searches in extragalactic sources have been performed and even fewer outside of the local universe (Churchwell & Shaver 1979). Although RRLs can be very useful probes of the ISM, they are challenging to detect observationally. Their major obstacle stems from the very low peak optical depths of the lines (i.e. with peak fractional absorption of $10^{-3} - 10^{-4}$; Gordon & Sorochenko 2002).

An important step in extragalactic exploration at low frequencies came with the discovery of carbon RRLs at 56 MHz in the nearby starburst galaxy, M 82 (Morabito et al. 2014) using the LOw Frequency ARray (LOFAR; van Haarlem et al. 2013). The advent of sensitive low-frequency telescopes has reinvigorated the field, as the large fractional bandwidth allows for the detection of many RRL transitions simultaneously and stacking can increase the signal to noise by an order of magnitude (e.g. Oonk et al. 2015; Salas et al. 2017).

This paper presents the first result of a survey for RRLs at cosmological distances taking advantage of the capabilities afforded by LOFAR with an in-depth study of the $z = 1.1946$ radio quasar 3C190 (see Section 4.2). This has resulted in the detection of RRLs outside of the local universe for the first time. While our aim was to investigate cold clouds associated with the HI absorbing gas near to 3C 190, we uncovered something unexpected: RRLs with a large offset in redshift from the radio source.

4.2 Target

We identified 3C 190 as a candidate for RRLs as it is a bright (∼ 20 Jy at 140 MHz), steep-spectrum radio galaxy with HI detected in absorption (Ishwara-Chandra et al. 2003). It is classified as a reddened quasar (Smith & Spinrad 1980). Narrow emission lines from [Ne III], [O II], and C III] locate the source at a redshift of $z = 1.1946 \pm 0.0005$ (Stockton & Ridgway 2001). As shown in Fig. 4.1, the host galaxy of 3C 190 is the central galaxy of a dense environment that is undergoing several major and minor mergers. Optical spectra reveal the presence of an absorption system at $z = 1.19565 \pm 0.00004$ observed via Mg II λ2798 and Fe II $\lambda\lambda$2343, 2382 and $\lambda\lambda$2586, 2599 (Stockton & Ridgway 2001).

Radio observations at 1662 MHz reveal two hot spots spanning 2.6″ (Spencer et al. 1991) or 22 kpc in projection. Diffuse emission, stretching to a 4″ extent, indicates the jets may have encountered a dense medium. Using the Giant Metrewave Radio Telescope (GMRT; resolution of ∼5″ at 650 MHz), a broad and complex profile of absorbed HI was detected with five Gaussian components spanning almost 600 km s^{-1} (Ishwara-Chandra et al. 2003). Most of these components are blue-shifted with respect to 3C 190. Of particular interest for carbon RRLs is a deep ($\tau_{\rm peak} = 0.0100 \pm 0.0003$) and narrow (FWHM $= 66.8 \pm 2.2$ km s^{-1}) feature that lies at a velocity offset of -210.2 ± 1 km s^{-1}. As the region shows evidence of shocked, turbulent gas, the narrow HI likely results from the radio jet interacting with the ambient medium

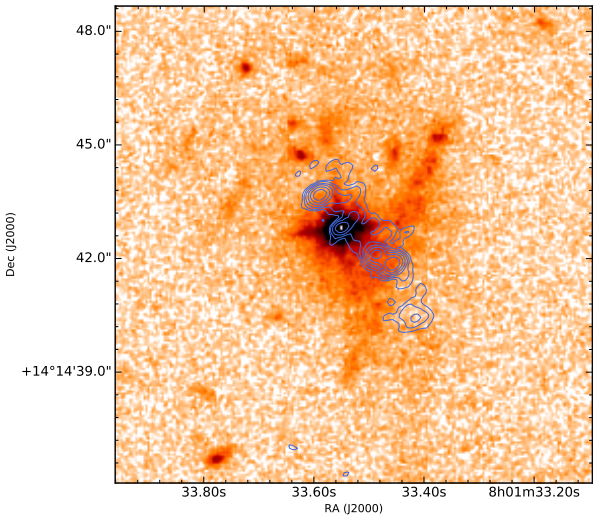


Figure 4.1: Quasar 3C 190 (centre) shown in an HST WFPC2 F702W image, where numerous satellite galaxies, a prominent linear feature, and extended diffuse emission can be seen (Stockton & Ridgway 2001) associated with the system. Superimposed are MERLIN 1658 MHz radio contours at $\sigma \cdot \log_5([1, 1.5, 2, 2.5, 3, 3.5])$ mJy/beam, showing the 22 kpc extent of the lobe hot spots. While 3C 190 is a steep spectrum source, it has a flat spectrum core, and thus we expect low-frequency emission to arise in the radio lobes.

(Ishwara-Chandra et al. 2003).

4.3 Observations and data reduction

3C 190 was observed with the LOFAR High Band Antennas (HBA) on 14 January 2017 (Project ID `LC7_027`). Four hours were spent on 3C 190, with ten minutes on the primary calibrator 3C 196 before and after. The 34 stations of the Dutch array were used in `HBA_DUAL_INNER` mode. Applying the HBA low analogue filter, we observed between 109.77 MHz and 189.84 MHz. The observing band is split into sub-bands (SBs) of 195.3125 kHz via a poly-phase filter. After which, each SB is divided into 64 channels and recorded at a frequency resolution of 3.0517 kHz. While data were taken at 1 s time intervals, RFI removal and averaging to 2 s were performed before storing the data.

Processing of the LOFAR data was performed with the SURFSara Grid processing facilities[1] (e.g. see Mechev et al. 2017, 2018). While a more detailed description of the data processing can be found in Emig et al. in prep, we summarise the steps below. Starting with the calibrator data, we flagged the first and last four edge channels of each SB, flagged for RFI using `AOflagger` (Offringa et al. 2012), selected only the core stations (max baseline $\sim$ 4 km), and averaged the data to resolutions of 6 s and 32 channels per SB (or 6.1034 kHz channels). Using `DPPP` (van Diepen & Dijkema

[1] https://www.surfsara.nl

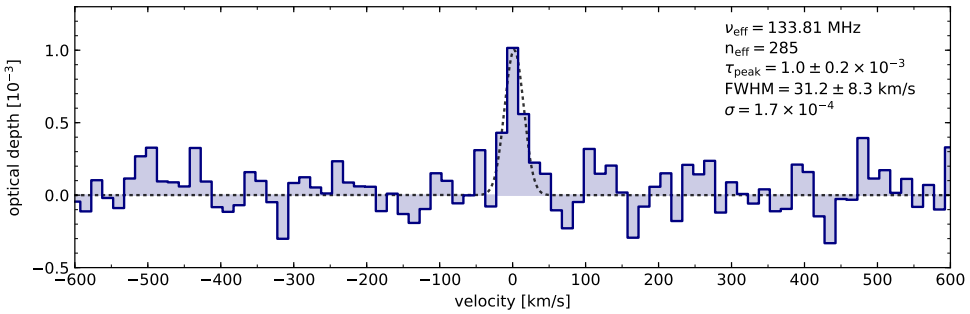


Figure 4.2: Detection of a radio recombination line in emission at quantum level $n_{\rm eff} = 285$, with a velocity centred on $z = 1.12405$ ($z = 1.12355$), originating from hydrogen (carbon). This is the average profile of a single line, effectively at 133 MHz, as a result of stacking 13 recombination lines in the spectrum of 3C 190.

2018), we solved for diagonal gain. With `LoSoTo` (de Gasperin et al. submitted) we found the median amplitude solution in time for each channel, creating a per channel bandpass solution. Then, a sixth order polynomial across 10 SBs (2 MHz) was fit to take into account the 1 MHz standing wave, the global slope of the bandpass, and smooth over its scatter (e.g. due to the poly phase filter). These effects are expected to be time independent. Next for the target data, we implemented the same flagging steps as for the calibrator and then applied the bandpass solutions. Flagging once more, for each SB we solved for phase only with `DPPP` on a 6 s time interval, using a LOFAR Global Sky Model (van Haarlem et al. 2013) generated sky model of the field. After averaging the data to a 30 s time resolution, we solved for the amplitude at full-frequency resolution. Importing the solution tables of all SBs into `LoSoTo`, we flagged for outliers (namely to catch RFI due to digital audio broadcasting with broadband effects) and smoothed the solutions in frequency space, with a running Gaussian of 4 SB FWHM, to ensure that any spectral features are not calibrated out. Once we applied the smoothed amplitude solutions, we imaged each channel with `WSCLEAN` (Offringa et al. 2014), in which multi-frequency synthesis was used per SB to make a continuum image and extract the clean components for the channel images (Offringa & Smirnov 2017).

4.4 Spectral processing

In this section we describe the processing of the data post calibration and imaging. In summary, once the channel images were created, we convolved them to the same resolution, extracted flux from a fixed aperture, removed the continuum flux, flagged SB spectra, and stacked the spectra in velocity space.

We first convolved every channel image to the same resolution of 236″, a few percent larger than the lowest resolution image, using `CASA` (McMullin et al. 2007). The flux density was then extracted from a fixed circular aperture of diameter 236″. A spectrum was created for each SB. A fit to the continuum was made to each SB

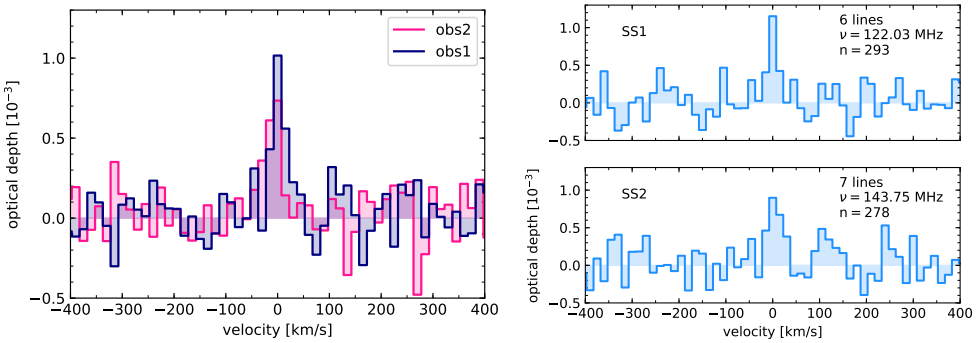


Figure 4.3: *Left:* Observation 1 (blue) as in Fig. 4.2. Observation, 2 shown in pink, is the supporting follow-up observation taken 6 months later. These spectra have been Doppler corrected to the LSR. Observation 2, with a factor of 1.6 less time spent on source, is the stack average of 10 lines at quantum number $\mathsf{n}_{\rm eff} = 287$ and $\nu_{\rm eff} = 130.9$ MHz. *Right:* We split the spectral lines from Observation 1 into 2 groups, sub-stack 1 (SS1) consisting of the 6 spectral lines with $\mathsf{n} > 285$ and sub-stack 2 (SS2) containing the 7 lines below.

spectrum with a low (1st or 2nd) order polynomial chosen to minimise the chi-square of the fit. For a given redshift, we blanked the channel of the expected frequency of the line when fitting for the continuum. At low frequencies, if stimulated transitions dominate, we have that $I_{\rm line} \approx I_{\rm cont}\tau_{\rm line}$, where the intensity we extract from the observations is $\approx I_{\rm line} + I_{\rm cont}$. Therefore, we subtracted the continuum fit and divided by it, which resulted in a measure of the optical depth.

We inspected spectra on a per SB channel basis. In each spectrum we interpolated over channels which had amplitudes higher than five times the spectrum rms. We interpolated over channels for which >60 % of the visibility data were flagged as well.

Examination of spectra on a per-SB basis was done first by eye to catch clear bandpass-related outliers, for example owing to broadband RFI. We also discarded SBs for which their rms or chi-squared of the continuum fit was a 5σ outlier (considering a rolling window of 20 SBs centred on the SB in question). Lastly, if >40 % (11) of the channels had been clipped or flagged, we did not include the SB in the stack.

In terms of the lines going into the stack, if one of the clipped channels fell within the blanking region of the expected spectral line, the line was not included in the final stack. Furthermore, we required at least three channels on either side of the blanked region to estimate the continuum, otherwise the line was discarded.

Taking the central frequency of each SB, we determined the spectral line closest in frequency and used it to convert the channels to velocity units using the radio definition of velocity, $v^{radio} = \frac{\nu_0 - \nu}{\nu_0}\, c$, where ν_0 is rest frequency. At this point, we corrected for the velocity offset necessary for Doppler tracking relative to the local standard of rest (LSR) frame. We then interpolated the velocities to a fixed grid with a channel spacing of 15 km s^{-1}.

With the spectra aligned in velocity space, the weighted mean optical depth in

each channel was found via

$$< \tau_{\rm chan} >= \frac{\Sigma_{i=0}^{N}(w_i \tau_i)}{\Sigma_{i=0}^{N} w_i}, \tag{4.1}$$

where i represents each SB going into the stack, and the weighting factor was determined by one over the noise variance of each SB spectrum, $w_i = \sigma_i^{-2}$. The effective frequency and quantum number of the stacked spectrum was determined by the weighted mean values of each SB containing a spectral line.

4.4.1 Statistical identification

We stacked RRLs across redshift space to search for features within the source and along the line of sight. This was sampled at an interval equivalent to the channel width $v = 15$ km s^{-1} or $\Delta z = 10^{-5}$ and ranged over $z = [0, 1.31]$. We note that with the flagging procedure above, the number of spectral lines in each stack at a given redshift did not remain constant. While the following methods are described extensively in Emig et al. in prep., we briefly summarise them here.

For each redshift tested, we performed a cross-correlation between a template spectrum and the pre-stacked spectra, both in units of optical depth. The template spectrum was populated with Gaussian line profiles at the location of the spectral lines that contributed to the final stack. The line profiles had a peak of one and their full width at half maximum (FWHM) was set by an assumed blanking region. We then took the cross-correlation and normalised it proportionally with the number of lines that went into the stack, i.e. the total area under the template spectrum. This was the same procedure implemented in Morabito et al. (2014), except we included a normalisation since the number of lines included at each redshift did not remain the same.

As a second test, we took a template spectrum, stacked the spectral lines at an assumed redshift, and integrated the signal within an assumed FWHM. Furthermore, we stacked and integrated the template spectrum at a range of redshifts, from $z-0.01$ to $z+0.01$, at redshift intervals of 10^{-5}. We then cross-correlated (a) with (b): (a) the integrated optical depth of the template stack as a function of redshift, and (b) the observed integrated optical depth of the stacks at each redshift. With this cross-correlation, we corroborate "mirrors" of the signal that can be found at a $\Delta z = \overline{\Delta\nu_{\mathsf{n},\rm eff}}/\nu_{\mathsf{n},\rm eff}$, or multiples thereof, where $\nu_{\mathsf{n},\rm eff}$ is the frequency of the effective n-level of the stack, and $\overline{\Delta\nu_{\mathsf{n},\rm eff}}$ is the average of the change in frequency between $\mathsf{n}_{\rm eff}$ and all other n levels included. We digress to explain the aforementioned *mirrors*. The distance between each recombination line in frequency space is unique, thus allowing us to accurately determine redshift. However, the difference in spacing between α-transitions n and $\mathsf{n}+1$ is small ($\sim$1 %); in other words, the frequencies at which recombination lines fall are close, but not quite, periodic. Therefore *mirrors* of the feature, which are broadened and reduced in peak intensity compared to the original, occur at offsets in redshift that match the frequency spacing between adjacent lines, or more precisely, $\Delta z = \overline{\Delta\nu_{\mathsf{n},\rm eff}}/\nu_{\mathsf{n},\rm eff}$. It was necessary to include the second cross-correlation on account of the low signal-to-noise regime of the lines coupled with poor estimation of the continuum over narrow SBs.

Table 4.1: Spectral properties of stacked RRLs.

$\mathsf{n}_{\rm eff}$	285	
$\nu_{\rm eff}$	133.81 MHz	
$\tau_{\rm peak}$	$(1.0 \pm 0.2) \times 10^{-3}$	19 ± 4 mJy
FWHM	13.9 ± 3.7 kHz	31.2 ± 8.3 km s^{-1}
rms	1.7×10^{-4}	3.3 mJy
$\int \tau \, {\rm d}\nu$	-14.8 ± 7.4 Hz	640 ± 320 mJy km s^{-1}

The values in the right column were determined by considering a measured flux density of 19.3 ± 3.9 Jy at 133.81 MHz.

We required that both cross-correlation methods result in a $> 5\sigma$ value at a redshift to report a detection.

4.5 Results

A significant feature ($> 5\sigma$) was found in the spectrum of 3C 190 when considering a line blanking of 15 km s^{-1}, arising from the α-transitions of hydrogen (carbon) at a redshift of $z = 1.12405$ (1.12355) $\pm$ 0.00005. A redundancy exists between carbon and hydrogen as their α-transitions are regularly offset by 149.4 km s^{-1}. The stacked feature includes 13 recombination lines of principal quantum numbers $\mathsf{n} = 266 - 301$, and it has an effective frequency of $\nu_{\rm eff} = 133.81$ MHz and quantum level of $\mathsf{n}_{\rm eff} = 285$. Increasing the blanking region further to 50 km s^{-1}, we find the averaged spectral feature (Fig. 4.2) to have a 31.2 ± 8.3 km s^{-1} width, thus an undersampled Gaussian with our velocity resolution, and an average integrated strength per line of $\int \tau {\rm d}\nu = -14.8 \pm 7.4$ Hz at 6.3σ. Properties of the line and spectrum are listed in Tab. 4.1.

The per-channel coverage is about two times lower in the blanked region versus the non-blanked. With the weighting described by Eq. 4.1, the noise outside of the blanked region is effectively $\sqrt{2}$ times lower than the rms reported in Tab. 4.1. At $z = 1.12405$, there are a total of 36 RRLs between 112 MHz – 165 MHz. However, 13 were included in the final stack. There were 22 lines that fell within three channels of the SB edge. An additional spectral line was discarded owing to a poor SB bandpass, identified as an outlier in both its rms and continuum fit.

4.5.1 Further validation

As further validation, we confirm the RRL detection (1) with a second observation, (2) by performing jack-knife tests (Miller 1974), (3) stacking in two independent subgroups, and (4) stacking other sources in the field at the same redshift. These tests all give further confidence to the detection and are described below. Additionally, stacking 3C 190 at other redshifts is consistent with noise.

(1) A second observation of 3C 190 was obtained on 03 May 2017 with the same

observational set-up, but with poor ionospheric conditions and thus only 2.5 hours of usable data. Having been taken several months apart, the Doppler correction differs by 32 km s^{-1}. As shown on the left side of Fig. 4.3 in pink, a 4.6σ feature, including ten spectral lines at $\mathsf{n}_{\rm eff} = 287$ and $\nu_{\rm eff} = 130.93$ MHz, is again seen at $z = 1.12405$ ($z = 1.12355$).

(2) Furthermore, jack-knife tests, as in Oonk et al. (2014), were performed on the stacked spectrum, in which we iteratively stack the spectrum, each time discarding one line from the stack. The line properties of the stacks do not differ substantially, indicating that a single SB is not responsible for the signal.

(3) We also split the lines into two groups (see also Oonk et al. 2014) resulting in two independent stacks, shown in the right-hand side of Fig. 4.3. Sub-stack 1 consists of six spectral lines with $\nu_{\rm eff} = 122.03$ MHz and $\mathsf{n}_{\rm eff} = 293$; SS2 consists of seven lines with $\nu_{\rm eff} = 143.75$ MHz and $\mathsf{n}_{\rm eff} = 278$. The line profiles are consistent among the two stacks. The line is more narrow but higher in its peak optical depth in SS1 as compared with SS2, which is consistent with Doppler broadening effects.

(4) Lastly, we stack the next two brightest sources in the field, 3C 191 and 4C +15.22, at $z = 1.12405$ as well. While their noise is significantly higher and thus so are their limits on the optical depth, we confirm no prominent feature is seen in their spectra. This gives further confidence that the detection is not an instrumental systematic.

4.6 Modelling and interpretation of RRL

The RRLs allow us to make constraints on the emitting gas properties arising from the observables of central velocity, line width, and line integrated strength. We discuss the constraints they place on the origin of the detected line. The difference in redshift between 3C 190 and the RRL-emitter implies a velocity difference of -9965 km s^{-1}, corresponding to a (luminosity) distance from 3C 190 of 81 Mpc.

To characterise the emitting gas, we use models of radio recombination line emission. Ideally, with multiple RRL detections distributed in frequency, models can be fit to derive gas properties. However, with only one data point, we instead explore a range of input physical conditions (density and temperature) and obtain a range in physical properties (emission measure, column density, path length, and pressure). We use these physical properties to constrain different, possible interpretations on the origin of the emitting gas.

The integrated optical depth of stimulated recombination lines at low frequencies is described by

$$\int \tau \, {\rm d}\nu = 2.046 \times 10^6 \, {\rm Hz} \cdot \exp\left(\chi_{\mathsf{n}}\right) \left(\frac{T_{\rm e}}{\rm K}\right)^{-5/2} \frac{EM}{\rm cm^{-6}\,pc} b_{\mathsf{n}} \beta_{\mathsf{n}} \tag{4.2}$$

for α transitions (e.g. Shaver 1975a; Salgado et al. 2017a). Here $\chi_{\mathsf{n}} = 1.58 \times 10^5 \, {\rm K}/(\mathsf{n}^2 T_{\rm e})$, n is the quantum level, and EM is the emission measure expressed as $EM = n_e n_{\rm ion} L_{\rm ion}$ for electron density n_e, ion density $n_{\rm ion}$, and path length $L_{\rm ion}$. The coefficients that describe the gas departure from local thermodynamic equilib-

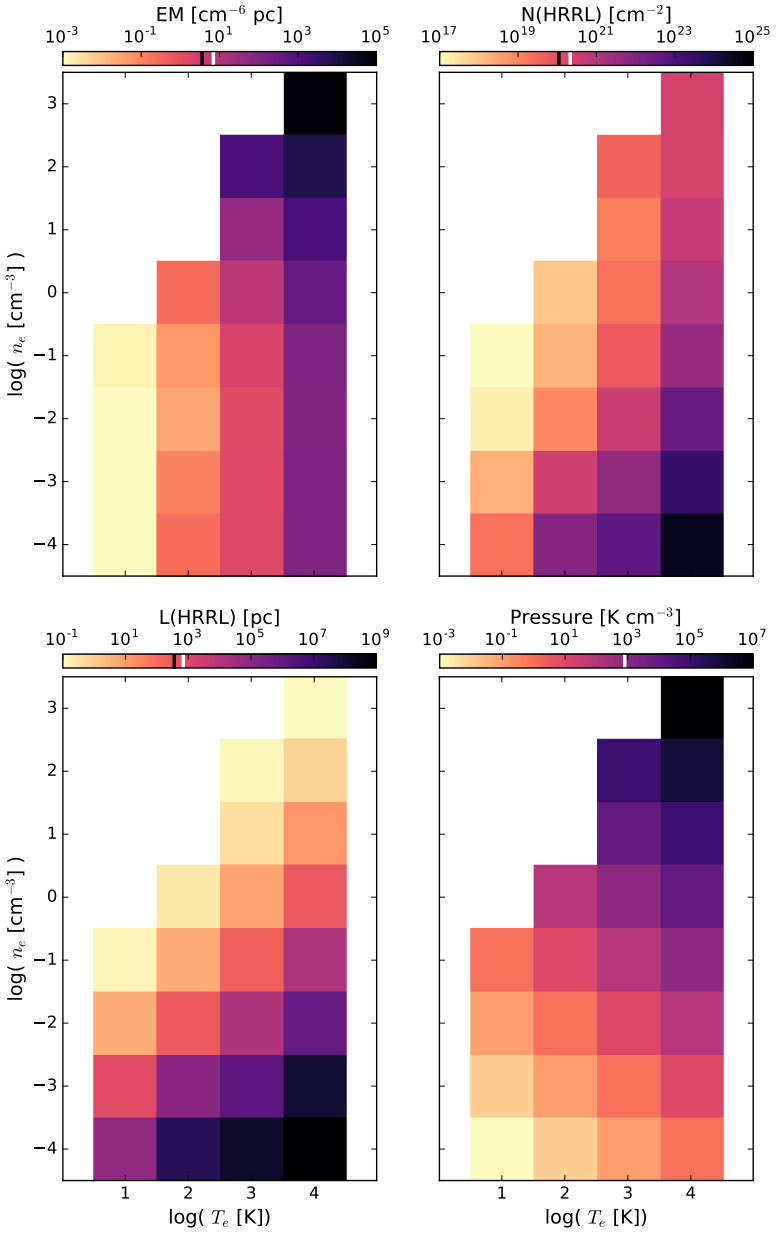


Figure 4.4: Derived physical properties from modelling of *hydrogen* RRL departure coefficients. For input physical conditions, electron density, and temperature, shown on the axes, we computed a departure coefficient ($b_n\beta_n$) via RRL modelling. Plugging in the $b_n\beta_n$, the integrated optical depth, quantum number, n_e and T_e into Eq. 4.2, we derive the emission measure (top left) and corresponding column density (top right) and path length (bottom left) of the RRL emission. Additionally we show the electron pressure (bottom right). Values representing the warm, ionised phase of a face-on (black) and an edge-on (white) Milky Way are shown with dashes in the colour bars. Hydrogen RRLs are expected to be most prominent from warm, $T \sim 1000 - 10\,000$ K, gas.

rium, $b_{\mathsf{n}}\beta_{\mathsf{n}}$, require detailed modelling of the atomic physics, for which we refer to the comprehensive low-frequency models of Salgado et al. (2017a,b).

We explored models with electron densities ranging from 10^{-4} cm^{-3} – 10^{3} cm^{-3}, electron temperatures of 10 K – 10^4 K, and with four different radiation fields. We wanted to cover a wide range of parameters that include the typical temperatures of carbon (10 K – 100 K) (Oonk et al. 2017; Salas et al. 2017) and of hydrogen (10^3 K – 10^4 K) (Anantharamaiah 1985b; Heiles et al. 1996b) RRL emitting gas. We note, the assumptions made for dielectronic capture in Salgado et al. (2017a) is relevant only for $T < 15\,000$ K (Watson et al. 1980). The densities were chosen to cover typical parameters found within our Galaxy in diffuse clouds, HII region outskirts, and extended low-density phases of the ISM (e.g. Tielens 2005; Ferriere 2001). Moreover, we incorporated more extreme conditions which resemble those of the (less dense) intergalactic medium (IGM) and (possibly, more dense) AGN environments. The radiation fields we consider have (1) only cosmic microwave background (CMB) radiation (for $z = 0$), (2) CMB radiation and a synchrotron field with $\beta = -2.6$ (e.g. de Oliveira-Costa et al. 2008; Klein et al. 2018b) and scaled to $T_r = 800$ K at 100 MHz, (3) to $T_r = 2000$ K, and (4) to $T_r = 10^5$ K. For the 19.3 Jy flux density of 3C 190 at 133.81 MHz, the synchrotron brightness temperature is given by $T_b = (1+z)\,4.7 \times 10^4$ K sr^{-1}. For an absorber 81 Mpc away with an estimated $\Omega \sim 2 \times 10^{-7}$ sr, the effective radiation temperature from 3C 190 is negligible ($T_r = 0.02$ K). This implies that if the RRLs originate in an intervening absorber, the continuum from 3C 190 does not contribute significantly to T_r. On the other hand, material that is ∼15 kpc away (or at least within a Mpc) would see a significantly higher radiation temperature, $T_r \sim 10^5$ K.

Starting from Eq. 4.2, we can solve for EM by inputting observed quantities (i.e. $\int \tau \, \mathrm{d}\nu = -14.8 \pm 7.4$ Hz at $\mathsf{n}_{\mathrm{eff}} = 285$) and the computed departure coefficient and assumed temperature. The physical properties obtained are shown in Fig. 4.4 and Fig. 4.5 for hydrogen and carbon, respectively. We assume all ionised electrons originate from their RRL element, and ionised fractions of $X_e = 1$ and $X_e = A_{\mathrm{C}} = 1.4 \times 10^{-4}$ (Sofia et al. 2004) in the respective cases. The figures show values obtained for $T_r = 2000$ K; these are representative of the radiation fields considered. The properties derived for the three coolest radiation fields do not differ by more than ∼20%. For the radiation field of $T_r = 10^5$ K, the results are qualitatively similar; they can be found in the Appendix, Fig. 4.6. We indicate values that the Milky Way would be observed to have if placed at $z = 1.124$; we assume cold gas extends to $R = 25$ kpc (Dickey et al. 2009) with parameters of the warm, ionised medium, and cold, neutral medium found in Tielens (2005). We also assume that the background continuum is emitted from a region $4''$ in extent, equivalent to 33.3 kpc at the redshift of 3C 190.

We consider five scenarios to explain the RRLs: (i) hydrogen or (ii) carbon RRLs from an intervening galaxy, (iii) hydrogen or (iv) carbon RRLs from the AGN outflow, and (v) hydrogen RRLs from the IGM. It is not immediately clear whether the RRL originates from hydrogen or carbon because of two main issues. Firstly, a regular separation of 149.4 km s^{-1} exists for the Bohr-like α-transitions between the two elements. Secondly, at the frequency at which we detected the feature, rest-frame 284 MHz, cold (carbon origin) and warm (hydrogen origin) gas phases have both

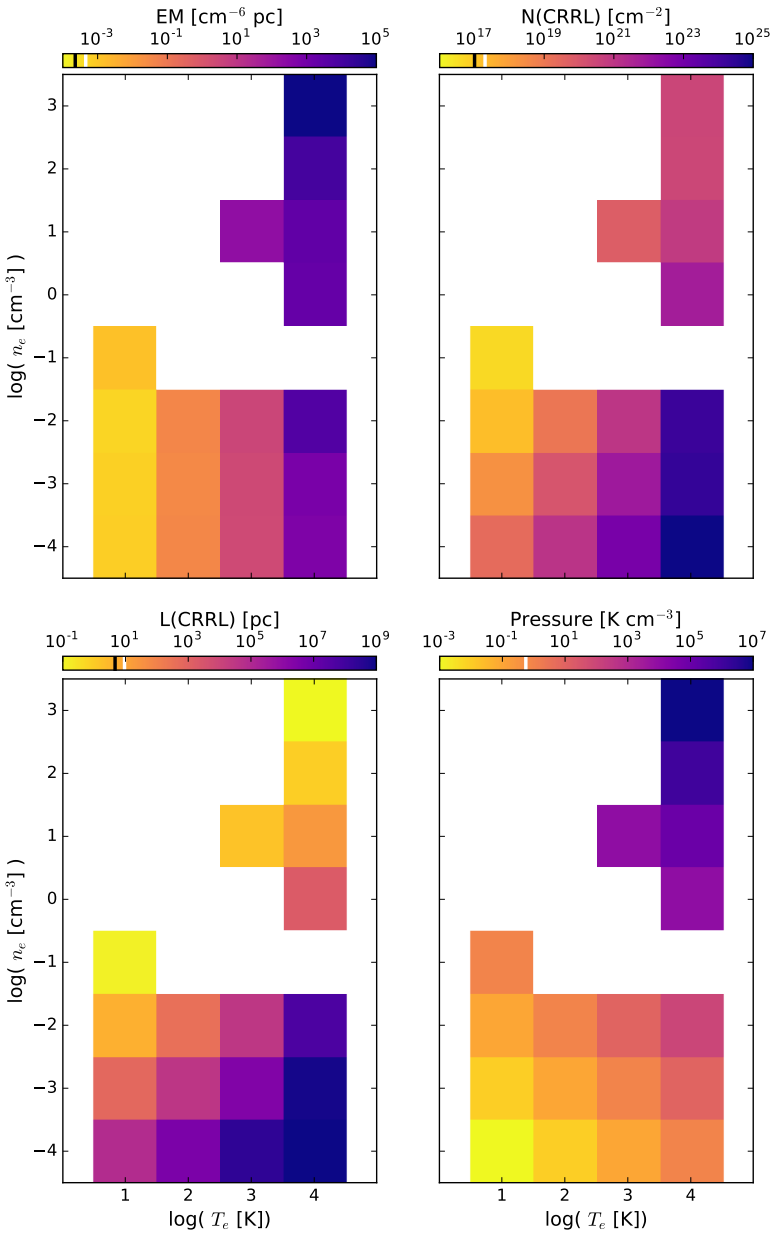


Figure 4.5: Same physical properties as in Fig. 4.4, except for *carbon* RRL modelling of the departure coefficients and the Milky Way values are representative of its cold, neutral medium. Carbon RRLs are expected to be most prominent from cold, $T \sim 10 - 100$ K, gas.

been observed in RRLs within our own Galaxy (e.g. Anantharamaiah 1985b). In the following sections, we give a brief description of each scenario and its RRL modelling constraints.

4.6.1 Intervening, dwarf-like galaxy

The RRLs would originate from the disc or in extended material of a dwarf-like galaxy along the line of sight. The galaxy should be small in mass and size such that it is not visible in Hubble Space Telescope (HST) imaging shown in Fig. 4.1 ($M_U > 27.2$). The narrow line-width of the RRL indicates a face-on orientation consistent with cool phases of the ISM for small galaxies, whether they be from the disc or from a halo component (Tumlinson et al. 2013). In this object, star formation would be possible and thus reasonable for a stimulated radiation field. Since we do not find a counterpart for the RRL in existing spectroscopic observations of the quasar (Stockton & Ridgway 2001), the RRL source should overlap with only the radio emission and not the optical. Furthermore, the high redshift and small mass suggests that the galaxy would be low in metallicity.

Hydrogen RRLs in an intervening galaxy

In scenario (i), the RRL would originate from hydrogen and thus largely ionised gas, either in the disc or in extended material of a dwarf-type galaxy.

We find sensible results (see Fig. 4.4) for models with temperatures $10^2 < T_e/\mathrm{K} < 10^5$ and densities $10^{-2} < n_e/\mathrm{cm}^{-3} < 10$, in agreement with RRLs observed in the disc of the Milky Way (Anantharamaiah 1985b; Heiles et al. 1996b). The model results for plausible interpretations of the RRL-emitting gas are listed in Table 4.2. We list the model input and derived physical properties for models with $T_r = 800$ K. Also listed is an estimate of the mass of warm ionised material; we assume a galaxy of $R = 5.5$ kpc and thus a surface filling factor of $\phi = 0.1$. We calculate the mass for a face-on disc as $M = m_H n_H \pi R^2 h$, where h is the scale height of the phase, $h = L/\phi$, and L is the path length derived from the model.

We also make note of warm, ionised gas that could originate in the halo or circumgalactic medium of a small galaxy. The densities would be slightly lower, closer to $n_e \leq 10^{-2}$ cm^{-3} (Tumlinson et al. 2017), and the path lengths longer, as the gas is distributed outside of the disc. Possibilities for this are also listed in Table 4.2, except we consider this material to be spherically distributed for the mass estimate. For spherical geometry, we let $R^3 = 3/4L \cdot (16.6\,\mathrm{kpc})^2$ to calculate the implied true radius of the clump, and extract the beam filing factor.

It is a relevant exercise to place an upper limit on the temperature of the gas assuming (non-)thermal motions Doppler-broaden the line width. As defined in Brocklehurst & Seaton (1972), a Doppler-broadened line with FWHM (Δv) is given by $\Delta v = 30.25\,\mathrm{km\ s^{-1}} \left(\frac{m_p}{m} \frac{T}{2\times10^4\,\mathrm{K}}\right)^{1/2}$, where m_p is the proton mass, m is the nuclear mass, and in this work T we take as the electron temperature. We find an upper limit for hydrogen gas of $T_e = 21\,300$ K.

As we find a broad range of physical conditions ($10^2 < T_e/\mathrm{K} < 10^{4.3}$ and $10^{-2} < n_e/\mathrm{cm}^{-3} < 10$) that could be interpreted as hydrogen RRLs in an intervening, dwarf

Table 4.2: Model results for which only *plausible* interpretations are listed.

Scenario	n_e [cm^{-3}]	T_e [K]	P/k [K cm^{-3}]	$b_n\beta_n$	EM [cm^{-6} pc]	N(RRL) [cm^{-2}]	L [pc]	ϕ	M [M$_\odot$]	Q [phot s^{-1}]	SFR [M$_\odot$ yr^{-1}]
i) galaxy, hydrogen											
	1	10^3	10^3	-28.9	7.8	2×10^{19}	7.8	0.10	2×10^8	3.6×10^{53}	3.9
	1	10^4	10^4	-205	350	1×10^{21}	350	0.10	8×10^9	2.5×10^{54}	27
	0.1	10^3	10^2	-103	2.2	7×10^{19}	220	0.10	5×10^8	1.0×10^{53}	1.1
	0.1	10^4	10^3	-587	120	4×10^{21}	12 000	0.66	$^*3 \times 10^{10}$	1.9×10^{54}	21
	0.01	10^3	10	-159	1.4	4×10^{20}	14 000	0.73	$^*3 \times 10^9$	1.4×10^{53}	1.5
ii) galaxy, carbon											
	0.1	10	$10^{3.9}$	-1.31	0.001	4×10^{16}	0.14	0.10	2×10^9		
	0.01	10	$10^{2.9}$	-3.37	0.0006	2×10^{17}	5.5	0.10	9×10^9		
iii) outflow, hydrogen											
	10	10^3	10^4	-5.23	44	1×10^{19}	0.43	0.0007	$^*9 \times 10^7$		
	10	10^4	10^5	-46.5	1600	5×10^{20}	16	0.008	$^*3 \times 10^9$		
iv) outflow, carbon											
	0.1	10	$10^{3.9}$	-0.76	0.002	8×10^{16}	0.25	0.0005	$^*4 \times 10^9$		

n_e, T_e: input electron density and electron temperature. For scenario (i) and (ii), a radiation temperature of $T_r = 800$ K was considered, and for scenario (iii) $T_r = 10^5$ K was considered. P: the thermal pressure of the input temperature and density, where $P/k \sim (n_e/X_e)T_e$. $b_n\beta_n$: the departure coefficient derived from the models. EM: the RRL emission measure derived from the model output and Eq. 4.2. N, L: the RRL column density and path length corresponding to the EM. ϕ: the surface filling factor of the object in the beam. M: the total mass of a disk-like, face-on galaxy of $R = 5.5$ kpc. Q: ionization rate needed to maintain ionized gas of the input n_e, T_e. SFR: star-formation rate determined from Q.

* assume a spherical cloud in calculating the mass.

galaxy, we estimate the number of ionising photons needed to maintain this gas phase. As described in Rubin (1968), the ionisation rate, Q, needed to maintain the implied physical conditions is given by Q [photons s^{-1}] $= 4.1 \times 10^{-10} n_e n_{\rm ion} V T_e^{-0.8}$ in a total volume V. The computed value of each plausible interpretation is shown in Table 4.2. Moreover, we calculate star formation rates (SFRs) from these ionisation rates, via SFR [M$_\odot$ yr^{-1}] $= Q \cdot 1.08 \times 10^{-53}$ (Kennicutt 1998).

Carbon RRLs in an intervening galaxy

In scenario (ii), RRLs from carbon would arise in cold, diffuse gas clouds within the mid-plane of a galaxy. With conservative constraints on thermal pressure, $P < 10^5$ K cm^{-3} (Jenkins & Tripp 2001; Herrera-Camus et al. 2017), and on the column density, N(CRRL)$< 10^{18}$ cm^{-2} such that N(HI)$< 10^{22}$ cm^{-2} for a carbon abundance of $A_C = 1.4 \times 10^{-4}$, we rule out many of the possibilities considered in Fig. 4.5. However, we find two plausible sets of physical conditions for this scenario, placing limits on electron density of $10^{-3} < n_e/{\rm cm}^{-3} < 1$ and electron temperature of $T_e < 100$ K. We list the results in Table 4.2, along with an estimate of the mass of cold neutral material as described above for the face-on disc, letting $n_H = n_e/A_C$.

Since this is a viable scenario, we derive integrated SFRs from the mass estimate of the cold, atomic gas. Referencing the relation Lopez-Sanchez et al. (2018) have found for local volume dwarf and spiral galaxies that have gas fractions dominated by HI, we find the estimated masses of 2×10^9 M$_\odot$ and 9×10^9 M$_\odot$ to have SFRs of ~ 0.1 M$_\odot$/yr and 1 M$_\odot$/yr, respectively.

4.6.2 AGN-driven outflow

In this scenario, RRLs would originate in gas outflowing as a result of the quasar or of the jet impacting the medium, for instance the optical linear feature apparent in Fig. 4.1.

There are notable reasons why this interpretation is less likely. In this system, nothing in the present literature indicates gas is moving at velocities higher than $\sim$600 km s^{-1} (Stockton & Ridgway 2001), one to two orders of magnitude below that of the RRL velocity. Furthermore, the narrow line width of the RRL is hard to maintain with a very high velocity.

Despite these indications, we investigated the scenario based on the following reasons. Jet interactions causing cold and ionised gas to outflow up to $\sim$1000 km s^{-1} have been observed in a number of AGN (e.g. Morganti et al. 2005). Via Sloan Digital Sky Survey (SDSS) optical spectra of $\sim$17 000 quasars, the majority of associated absorbers are found out to 2000 km s^{-1} – 4000 km s^{-1}, with tails to 10 000 km s^{-1}, albeit for warmer gas (Chen & Pan 2017). Additionally, warm gas with narrow line widths and velocities of -14 050 km s^{-1} has been observed in radiation-driven outflows (Hamann et al. 2011), and with velocities up to -3000 km s^{-1} in low-ionisation species, 900 pc away from the quasar (Xu et al. 2018).

In the case of 3C 190, its jet speed is estimated to be 0.22c, even out to its current $\sim$15 kpc scale. This is based on orientation (Best et al. 1995) and electron ageing (Katz-Stone & Rudnick 1997). The jet indeed appears to be interacting with the

optical linear feature seen through the diffuse radio emission surrounding the hot spots (Katz-Stone & Rudnick 1997) as well as displaced [O II] and a resulting cavity-like feature (Stockton & Ridgway 2001). Again, we point out that the displaced [O II] has a FWHM of 85 km s^{-1} and reaches velocities of -600 km s^{-1} relative to 3C 190.

Two possible outflow scenarios could be that the RRL originates in ablated [OII] material and is carried to high velocities, approaching that of the jet or wind speeds, and that the RRL cools out of the shock heated material, and as it approaches $T \sim 10^4$ K, it condenses.

For a 500 pc cloud with a velocity dispersion of 31 km s^{-1}, the dissipation time of the cloud is 10^7 yrs. With a bulk motion of 10 000 km s^{-1}, this gas could reach ∼100 kpc before dissipating. This implies that observing this type of gas cloud is conceivable.

Hydrogen RRLs in an AGN-driven outflow

In this scenario (iii), hydrogen RRLs would originate in (partially) ionised gas. With the higher densities we explore, collisional broadening may cause significant line broadening. We adopt the following expression for the Lorentzian FWHM due to collisions with electrons for α-transitions (Salgado et al. 2017b),

$$\Delta\nu_{\rm col} \approx \frac{n_e}{\rm cm^{-3}} \left(\frac{10^a \mathsf{n}^{\gamma_{\rm col}}}{\pi} \right) \rm Hz, \tag{4.3}$$

where a and $\gamma_{\rm col}$ depend on the gas temperature (values for which be found in Salgado et al. (2017b)) and n is the quantum level. We use this prescription to place an upper limit on the density of $n_e < 15$ cm^{-3} for our range of temperatures. For a density of $n_e = 10$ cm^{-3} and temperatures of $T_e = 10^3$ K – 10^4 K, the thermal pressure would be elevated compared to typical ISM values (see Fig. 4.4). High pressures have been found in various gas phases of outflowing material (e.g. Santoro et al. 2018; Oosterloo et al. 2017; Holt et al. 2011). We list physical conditions that would indeed be possible within this scenario in Table 4.2.

Carbon RRLs in an AGN-driven outflow

In scenario (iv), RRLs would originate in an outflow, but from material colder and more dense than the previous scenario. As we stated in the intervening galaxy example, we expect $N(\rm CRRL) < 10^{18}$ cm^{-2}, since the corresponding neutral hydrogen column density limit is 10^{22} cm^{-2}. We note that for gas with $T_e < 100$ K and density of $0.01 < n_e/\rm cm^{-3} < 1$, the path length derived from the model is $L = 0.25$ pc. Assuming that the gas is a sphere, this path length would imply a radius of 370 pc and hence a surface filling factor of 5×10^{-4}. These and other physical properties listed in Table 4.2 do not seem unreasonable. We conclude that this scenario could explain the RRLs we observe.

4.6.3 Hydrogen RRLs from the intervening IGM

We consider gas with typical properties of an isolated IGM cloud at $z \sim 1$, with temperature $T_e \sim 10^{4.5}$ K and densities $n_e < 10^{-4}$ cm^{-3} (McQuinn 2016). To test this scenario (v), we extended the RRL models down to densities of 10^{-4} cm^{-3}, but notice that at low densities and high temperatures, we find unreasonably large path lengths, approaching 10^5 to 10^7 kpc, to match the observed feature (see Fig. 4.4). Ultimately, we determine it is unfeasible to reproduce the observed integrated optical depth of the RRL with an IGM cloud.

4.7 Conclusions

Using LOFAR, we have identified RRLs centred at 133.8 MHz in the spectrum of 3C 190, as a result of stacking 13 α-transitions at $z = 1.124$ (Figure 4.2). This is the first detection of RRLs outside of the local universe.

At low frequencies, recombination lines can occur in diffuse gas from stimulated transitions in hydrogen and carbon. Since their transitions are regularly spaced ∼150 km s^{-1} apart, an ambiguity exists in determining the species of origin. However, carbon and hydrogen RRLs originate in distinctly different types of gas. Carbon arises in cold, neutral gas clouds, and hydrogen RRLs arise in warm, largely ionised material.

In this paper, we demonstrate how RRLs can be used to study the physical properties in these types of gas clouds. We model the non-LTE effects responsible for strong stimulation; inputting physical conditions (temperature and density), we constrain the origin of the gas (Table. 4.2). We find the RRL could be explained by hydrogen ($10^2 < T_e/\mathrm{K} < 10^{4.3}$ and $10^{-2} < n_e/\mathrm{cm}^{-3} < 10$) or carbon ($T_e < 10^2$ K and $10^{-3} < n_e/\mathrm{cm}^{-3} < 1$) emission in an intervening, dwarf galaxy ($M \sim 10^9$ M$_\odot$), roughly 80 Mpc from 3C 190. Although we consider it to be less likely, we cannot rule out the possibility that the RRL emitter is outflowing from the radio-loud quasar at ∼10 000 km s^{-1} and from hydrogen or carbon. Lastly, we rule out an IGM origin.

Since the RRLs in the spectrum of 3C 190 have no counterpart in existing observations, follow-up investigations are crucial. Observations targeting RRLs at a higher and a lower frequency give the best indication of success; they would distinguish between the carbon and hydrogen RRL origin and further constrain the physical conditions of the gas. An intervening galaxy seen only against the radio emission may reveal itself, slightly offset from the quasar, as an absorber in IFU observations. taken with high-spatial resolution. If the RRL originates from cold, neutral gas, HI 21cm absorption may be found at the redshifted frequency. HI has been searched for over only a small velocity interval centred on the systemic velocity and gas at ∼10 000 km s^{-1} would have been missed.

The detection of RRLs in the spectrum of 3C 190 has opened up a new way to study the physical properties of the ISM out to cosmological distances. The LOFAR Two Metre Sky Survey (Shimwell et al. 2017) of the Northern Hemisphere is uniquely suited to characterise RRL emitters in a large population of sources.

Acknowledgements

The authors would like to thank the referee for constructive and positive feedback. The authors would like to thank Alan Stockton, Anita Richards, and Ralph Spencer for their help in acquiring supplementary data. We also thank Aayush Saxena, Madusha Gunawardhana, Rogier Windhorst, Turgay Culgar, and Nastasha Wijers for useful discussions. KLE, PS, JBRO, HJAR, and AGGMT acknowledge financial support from the Netherlands Organization for Scientific Research (NWO) through TOP grant 614.001.351. AGGMT acknowledges support through the Spinoza premier of the NWO. MCT acknowledges financial support from the NWO through funding of Allegro. FdG is supported by the VENI research programme with project number 639.041.542, which is financed by the NWO. Part of this work was carried out on the Dutch national e-infrastructure with the support of the SURF Cooperative through grant e-infra 160022 & 160152.

This paper is based (in part) on results obtained with International LOFAR Telescope (ILT) equipment under project code `LC7_027`. LOFAR (van Haarlem et al. 2013) is the LOw Frequency ARray designed and constructed by ASTRON. It has observing, data processing, and data storage facilities in several countries, which are owned by various parties (each with their own funding sources) and are collectively operated by the ILT foundation under a joint scientific policy. The ILT resources have benefitted from the following recent major funding sources: CNRS-INSU, Observatoire de Paris and Universite d'Orleans, France; BMBF, MIWF-NRW, MPG, Germany; Science Foundation Ireland (SFI), Department of Business, Enterprise and Innovation (DBEI), Ireland; NWO, The Netherlands; The Science and Technology Facilities Council, UK; and the Ministry of Science and Higher Education, Poland.

Software: APLpy (Robitaille & Bressert 2012), astropy (The Astropy Collaboration 2018), CASA (McMullin et al. 2007), CRRLpy (Salas et al. 2016), DPPP (van Diepen & Dijkema 2018), iPython (Perez & Ganger 2007), LoSoTo (de Gasperin et al. submitted), matplotlib (Hunter 2007), PYBDSF (Mohan & Rafferty 2015), WSCLEAN (Offringa et al. 2014)

4.A RRL Modelling for high radiation temperatures

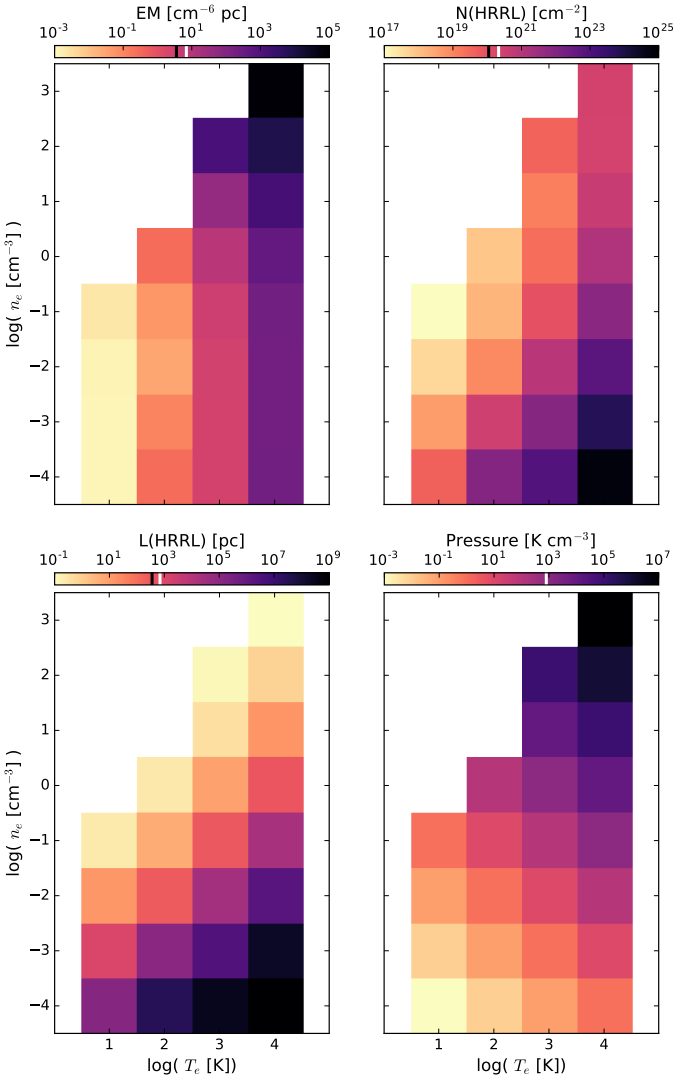


Figure 4.6: (cont.)

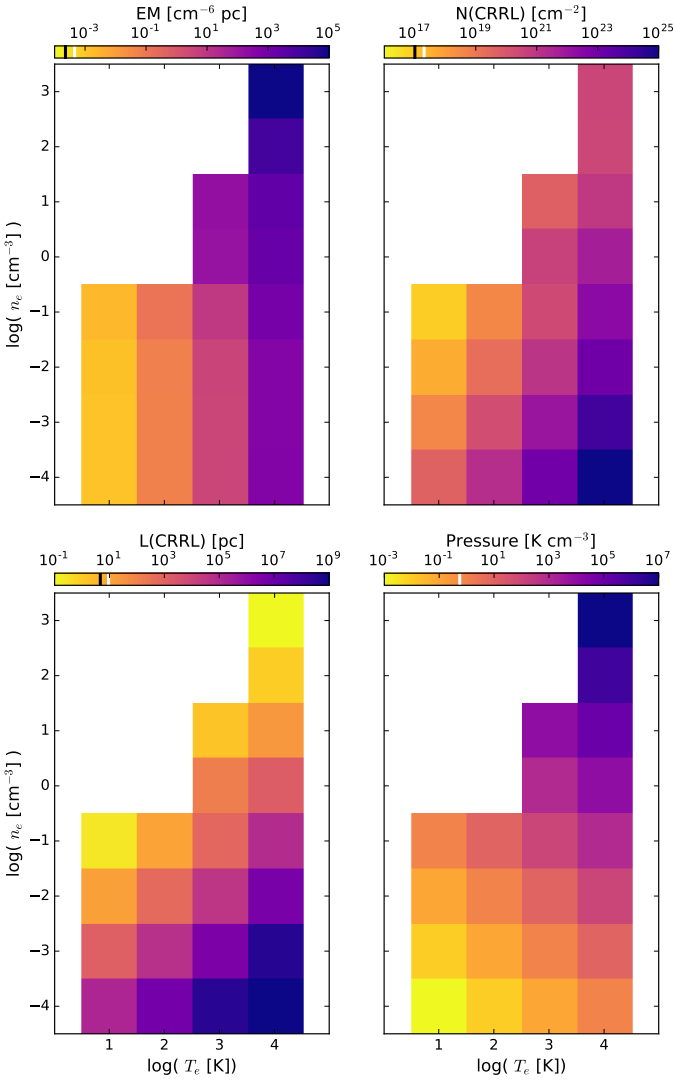


Figure 4.6: Same physical properties as in Fig. 4.4 and 4.5, except these results for hydrogen (left) and carbon (right) have been derived with $T_r = 10^5$ K.

5 | Searching for the largest bound atoms in space

Abstract

Context. Radio recombination lines (RRLs) at frequencies $\nu < 250$ MHz trace the cold, diffuse phase of the interstellar medium, and yet, RRLs have been largely unexplored outside of our Galaxy. Next-generation low-frequency interferometers such as LOFAR, MWA, and the future SKA will, with unprecedented sensitivity, resolution, and large fractional bandwidths, enable the exploration of the extragalactic RRL universe.

Aims. We describe methods used to (1) process LOFAR high band antenna (HBA) observations for RRL analysis, and (2) search spectra for RRLs blindly in redshift space.

Methods. We observed the radio quasar 3C 190 ($z \approx 1.2$) with the LOFAR HBA. In reducing these data for spectroscopic analysis, we placed special emphasis on bandpass calibration. We devised cross-correlation techniques that utilize the unique frequency spacing between RRLs to significantly identify RRLs in a low-frequency spectrum. We demonstrate the utility of this method by applying it to existing low-frequency spectra of Cassiopeia A and M 82, and to the new observations of 3C 190.

Results. Radio recombination lines have been detected in the foreground of 3C 190 at $z = 1.12355$ (assuming a carbon origin) owing to the first detection of RRLs outside of the local universe (first reported in Emig et al. 2019). Toward the Galactic supernova remnant Cassiopeia A, we uncover three new detections: (1) stimulated Cϵ transitions ($\Delta \mathsf{n} = 5$) for the first time at low radio frequencies, (2) Hα transitions at 64 MHz with a full width at half-maximum of 3.1 km s^{-1} the most narrow and one of the lowest frequency detections of hydrogen to date, and (3) Cα at $v_{\rm LSR} \approx 0$ km s^{-1} in the frequency range 55–78 MHz for the first time. Additionally, we recover Cα, Cβ, Cγ, and Cδ from the -47 km s^{-1} and -38 km s^{-1} components. In the nearby starburst galaxy M 82, we do not find a significant feature. With previously used techniques, we reproduce the previously reported line properties.

Conclusion. RRLs have been blindly searched and successfully identified in Galactic (to high-order transitions) and extragalactic (to high redshift) observations with our spectral searching method. Our current searches for RRLs in LOFAR observations

are limited to narrow (< 100 km s^{-1}) features, owing to the relatively small number of channels available for continuum estimation. Future strategies making use of a wider band (covering multiple LOFAR subbands) or designs with larger contiguous frequency chunks would aid calibration to deeper sensitivities and broader features.

K. L. Emig, P. Salas, F. de Gasperin, J. B. R. Oonk, M. C. Toribio, A. P. Mechev, H. J. A. Röttgering, and A. G. G. M. Tielens
Astronomy & Astrophysics (2020)

5.1 Introduction

Recombination lines that are observable at low radio frequencies ($\lesssim 1$ GHz) involve transitions with principal quantum numbers $\mathsf{n} \gtrsim 200$. They trace diffuse gas ($n_e \approx 0.01$ – 1 cm^{-3}) that can be considered cool in temperature ($T_e \approx 10 - 10^4$ K).

Observations in our Galaxy suggest that the most prominent radio recombination lines (RRLs) at $\nu \lesssim 250$ MHz arise from cold ($T_e \approx 10 - 100$ K), diffuse ($n_e \approx 0.01$ – 0.1 cm^{-3}) gas within diffuse HI clouds and in clouds surrounding CO-traced molecular gas (Roshi & Kantharia 2011; Salas et al. 2018). This reservoir of cold gas, referred to as "CO-dark" or "dark-neutral" gas, is missed by CO and HI emission observations. Yet it is estimated to have a comparable mass to the former two tracers (Grenier et al. 2005) and is the very site where the formation/destruction of molecular hydrogen transpires.

In addition, RRLs are compelling tools to study the physics of the interstellar medium (ISM) because they can be used to determine physical properties of the gas, specifically the temperature, density, path-length and radiation field (Shaver 1975b; Salgado et al. 2017b). Pinning down these properties is key to describing the physical state of a galaxy and understanding the processes of stellar feedback. RRL modeling depends, not on chemical-dependent or star-formation modeling, but on (redshift-independent) atomic physics. Since low-frequency RRLs are stimulated transitions, they can be observed to high redshift against bright continuum sources (Shaver 1978). With evidence of stimulated emission being dominant in local extragalactic sources at $\sim$1 GHz (Shaver et al. 1978), it is plausible that RRLs can be observed out to $z \sim 4$. Low-frequency RRLs, therefore, have a unique potential to probe the ISM in extragalactic sources out to high redshift.

The physical properties of gas can be determined when RRLs are observed over a range of principal quantum numbers. However, they are extremely faint, with fractional absorption of $\sim 10^{-3}$ or less. At frequencies of $\sim$150 MHz, RRLs have a $\sim$1 MHz spacing in frequency. By $\sim$50 MHz, their spacing is $\sim$0.3 MHz. Large fractional band-widths enable many lines to be observed simultaneously. On one hand this helps to constrain gas properties (e.g. Oonk et al. 2017, hereafter O17). On the other, it enables deeper searches through line-stacking (e.g. Balser 2006).

The technical requirements needed for stimulated RRL observations can be summarized as follows: (1) large fractional bandwidths that span frequency ranges 10 – 500 MHz for cold, carbon gas and 100 – 800 MHz for (partially) ionized, hydrogen gas; (2) spectral resolutions of $\sim$0.1 kHz for Galactic observations and $\sim$1 kHz

for extragalactic; (3) high sensitivity per channel; and (4) spatial resolutions which ideally resolve the $\lesssim$1–100 pc emitting regions. These requirements have inhibited wide-spread, in-depth studies of low-frequency RRLs in the past, largely due to the low spatial resolutions and the narrow bandwidths of traditional low-frequency instruments – owing to the difficulty of calibrating low frequency observations affected by the ionosphere.

However, with next generation low frequency interferometers, such as the Low Frequency Array (LOFAR; van Haarlem et al. 2013), the Murchison Widefield Array (MWA; Tingay et al. 2013), and the future Square Kilometer Array (SKA), new possibilities are abound for the exploration of the ISM through RRLs. LOFAR has currently been leading the way, thanks to its raw sensitivity and the flexibility of offering high spectral and spatial resolutions.

LOFAR operates between 10 MHz – 90 MHz via low band antennas (LBA) and 110 MHz – 250 MHz via high band antennas (HBA). The array consists of simple, inexpensive dipole antennas grouped into stations. LOFAR is an extremely flexible telescope, offering multiple observing modes (beam-formed, interferometric) and vast ranges of spectral, timing and spatial resolutions. It is the first telescope of its kind in the Northern Hemisphere and will uniquely remain so for the foreseeable future.

The first Galactic RRL analyses with LOFAR have been directed towards Cassiopeia A (Cas A), a bright supernova remnant whose line of sight intersects gas within the Perseus Arm of the Galaxy. These studies have highlighted the capability of RRL observations, and through updated modeling of atomic physics (Salgado et al. 2017a,b), have laid important ground work for the field in a prototypical source. It was shown that with recombination lines spanning principal quantum numbers of $\mathsf{n} = 257 - 584$ the electron temperature, density, and path-length of cold, diffuse gas can be determined to within 15 percent (O17). Wide bandwidth observations, especially at the lowest observable frequencies (11 MHz), can be used to constrain gas physical properties together with the α, β and γ transitions of carbon in a single observation (Salas et al. 2017, hereafter S17). Through pc-resolution and comparisons with other cold gas tracers, it was shown that low-frequency RRLs indeed trace CO-dark molecular gas on the surfaces of molecular clouds (Salas et al. 2018, hereafter S18). Finally, observations towards Cygnus A demonstrated that bright extragalactic sources can also be used to conduct Galactic pinhole studies (Oonk et al. 2014).

The first extragalactic observations with LOFAR are showing that low frequency RRLs provide means to trace cold, diffuse gas in other galaxies and out to high redshift. These studies focused on M 82 (Morabito et al. 2014, hereafter M14), a nearby prototypical starburst galaxy, and the powerful radio quasar 3C 190 at z$\sim$1.2 (Emig et al. 2019). While these searches are important first steps that show RRL detections are possible, they also indicate that detailed analyses of stacking are necessary (Emig et al. 2019).

In this article, we cover that much needed detailed look. We describe the methods behind the detection of RRLs in 3C 190 (Emig et al. 2019). We explain processing of LOFAR 110–165 MHz observations for spectroscopic analysis (Section 5.2). We then focus on the methods used to search across redshift space for the presence of RRLs in a low-frequency spectrum (Section 5.3). We apply this technique to existing 55–78 MHz LOFAR observations of Cas A (Section 5.4) and demonstrate that it can

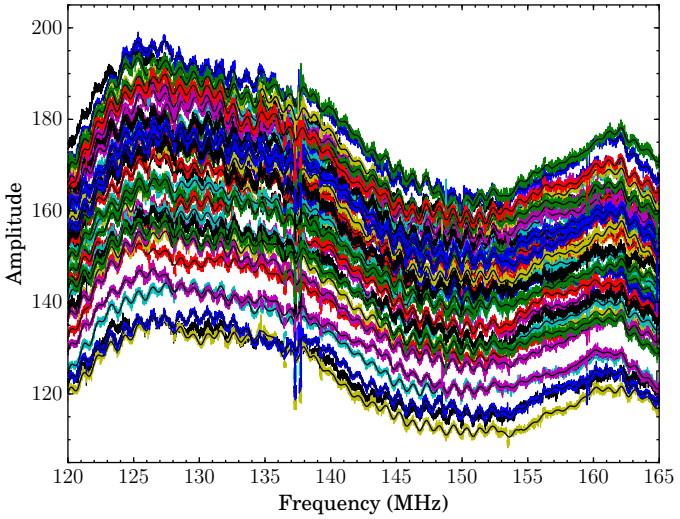


Figure 5.1: Bandpass solutions of the XX polarization towards the primary calibrator 3C196, in which stations are represented with different colors. These solutions demonstrate the global shape of the bandpass, as well a ∼1 MHz ripple which results from a standing wave. Unflagged RFI is still present between 137–138 MHz. All core-stations have the same cable length and thus standing wave of the same periodicity. The fit to the solutions of each station, which is transferred to the target, is shown in black.

be used to recover known RRLs in the spectrum, in addition to previously unknown features. We then focus on the LOFAR observations of M 82 in Section 5.5. Section 5.6 covers the application of our spectral search to 3C 190. We discuss the utility of the method and implications for future observations in Section 5.7. Conclusions are given in Section 5.8.

5.2 Spectroscopic Data Reduction

In this section, we cover the implementation of direction-independent spectroscopic calibration for LOFAR HBA(-low; 110–190 MHz) interferometric observations. Special attention is placed on the bandpass as it is one of the most crucial steps and underlies the main motivations for our strategy.

5.2.1 HBA Bandpass

The observed bandpass of LOFAR's HBA-low (110–190 MHz filter) is a result of the system response across frequency. It is dependent upon both hardware and software related effects. The physical structure and orientation of the dual-polarization dipole antennas influence the global shape of the bandpass. Once the sky signal enters through the antennas, an analog beamformer forms a single station beam. This beam (and the model of it) is frequency dependent (except at the phase center). The signal is transferred from the station through coaxial cables to a processing cabinet. As

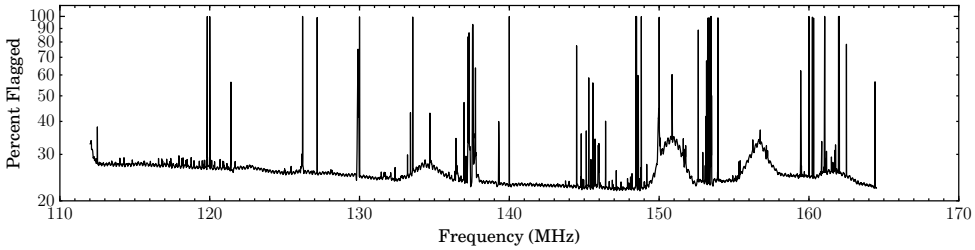


Figure 5.2: Percentage of flagged visibilities per channel in the calibrated target data. The total percentage here includes 10 remote stations and 4 core stations that have been flagged. Broad bumps centered at frequencies of about 135, 151, 157, and 162 MHz show broad-band RFI that results from intermodulation products of DAB amplifiers.

a result of an impedance mismatch between the cables and receiver units, standing waves are imprinted in the signal, in proportion to the length of the cable. Standing waves causing the ∼1 MHz ripple are apparent in Figure 5.1. An analog filter is then applied; this is responsible for the roll-off at either end of the bandpass. Next, after analog-to-digital conversion, a polyphase filter (PPF) is applied to split the data into subbands 195.3 kHz wide, which each have fixed central frequencies (for a given filter and digital-converter sampling frequency). The data are transported to the off-site correlator via optical fibers. A second PPF is applied, now to each subband, to re-sample the data into channels. This PPF imprints sinusoidal ripples within a subband. This can and is corrected for by the observatory. However, after the switch to the COBALT correlator in 2014, residuals of the PPF are present at the ∼1% level. This PPF is also responsible for in-subband bandpass roll-off which renders edge channels unusable. In-subband roll-off together with fixed subband central frequencies results in spectral observations that are processed with non-contiguous frequency coverage.

Radio frequency interference (RFI) is another major contributor to frequency dependent sensitivity. Of particular hindrance that has increased over the past years (for comparison, see Offringa et al. 2013) is digital audio broadcasting (DAB). DABs are broadcast in 1.792 MHz wide channels in the frequency ranges 174– 195 MHz. However, with the output power reaching non-linear proportions ($\propto P^3$), intermodulation products of the DABs appear at frequencies of 135, 151, 157, and 162 MHz, as can be seen in Figure 5.2. The sensitivity reached and the stability of the bandpass is affected in these regions, rendering data unusable for RRL studies at the central most frequencies.

5.2.2 Procedure

In this section we describe the reduction of 3C 190 observations. An overview of the steps we take in our data reduction strategy include: flagging and RFI removal of calibrator data, gain and bandpass solutions towards the primary calibrator, flagging and RFI removal of target data, transfer of bandpass solutions to the target, self-calibrated phase and amplitude corrections towards the target, and imaging. These steps are described in detail below.

Processing was performed with the SURFSara Grid processing facilities[1] making use of LOFAR Grid Reduction Tools (Mechev et al. 2017, 2018). It relies on the LOFAR software `DPPP` (van Diepen & Dijkema 2018), `LoSoTo` (de Gasperin et al. 2019), `wsclean` (Offringa et al. 2014), and `AOFlagger` (Offringa et al. 2012) to implement the necessary functions.

The observations of 3C 190 were taken with the LOFAR HBA-low on Jan 14, 2017 (Project ID: `LC7_027`). The set up was as follows: four hours were spent on 3C 190, with ten minutes on the primary calibrator 3C 196 before and after. The 34 stations of the Dutch array took part in the observations. Frequencies between 109.77 MHz and 189.84 MHz were recorded and divided into subbands of 195.3125 kHz. Each subband was further split into 64 channels and recorded at a frequency resolution of 3.0517 kHz. While taken at 1 s time intervals, RFI removal and averaging to 2 s were performed by the observatory before storing the data.

Pre-processing and flagging

Before calibration, we first implemented a number of flagging steps. Using `DPPP`, we flagged the calibrator measurement sets for the remote station baselines, keeping only the 24 core stations (CS). Since similar ionospheric conditions are found above stations this close in proximity (Intema et al. 2009; de Gasperin et al. 2018), this heavily reduces direction dependent effects, and this also avoids the added complication of solving for the sub-microsecond drifting time-stamp of the remote stations (e.g. van Weeren et al. 2016).

As the CS of the HBA are split into two ears, we filtered out the ear-to-ear cross-correlations. Four channels (at 64 channel resolution) at both edges of each subband are flagged to remove bandpass roll-off. With `AOFlagger`, we used an HBA-specific flagging strategy to further remove RFI. We flagged all data in the frequency ranges 170 MHz – 190 MHz due to the DABs (see Section 5.2.1). Additionally we flagged stations CS006HBA0, CS006HBA1, CS401HBA0, CS501HBA1 due to bandpass discontinuity. The data were then averaged in time and in frequency to 6 s and 32 channels per subband (or 6.1034 kHz).

Calibration solutions towards the primary calibrator

With the scan of 3C 196 (ObsID: L565337) taken at the start of the observation, we first smoothed the visibilities and weights with a Gaussian weighting scheme that is proportional with one over the square of the baseline length (e.g. see de Gasperin et al. 2019). With `DPPP`, we obtained diagonal (XX and YY) gain solutions towards 3C 196, at full time resolution and with a frequency resolution of 1 subband, while filtering out baselines shorter than 500 m to avoid large-scale sky emission. An 8 component sky-model of 3C 196 was used (courtesy of A. Offringa).

Next, we collected the solution tables from each subband and imported them into `LoSoTo`. For each channel, we found its median solution across time, the results of which are shown in Figure 5.1. After 5σ clipping, we fit the amplitude vs. frequency solutions with a rolling window (10 subbands wide) polynomial (6th order). With

[1] https://www.surf.nl

a window of 10 subbands, we attempted to fit over subband normalization issues, interpolate over channels which were flagged or contained unflagged RFI — e.g. RFI-contaminated channels between 137–138 MHz in Figure 5.1 — and avoid transferring per channel scatter to the target. The fit to these solutions is also shown in Figure 5.1.

Calibration and imaging of the target Field

Flagging and averaging of the target data were performed as described in Section 5.2.2. We then applied the bandpass solutions found with the primary calibrator 3C 196. We next smoothed the visibilities with the baseline-dependent smoother. Considering that ionospheric effects were minimal, the CS are all time-stamped by the same clock, and our target 3C 190 is a bright and dominant source, we solved explicitly for the diagonal phases with `DPPP` with a frequency resolution of one subband and at full time resolution, while filtering out baselines shorter than 500 m. We performed this self-calibration using a Global Sky Model (van Haarlem et al. 2013), which included 128 sources down to 0.1 Jy within a 5 degree radius. We then applied these solutions to full resolution data (32 channels, 6 s).

To correct for beam errors and amplitude scintillation, we performed a "slow" amplitude correction. We first averaged the data down to a 30 s time-resolution, then smoothed the visibilities with the baseline-based weighting scheme. While again filtering out baselines shorter than 500 m, we used `DPPP` to solve the amplitude only, every 30 sec and twice per subband. Before applying this amplitude correction, the solution tables from each subband are imported into `LoSoTo`. Using `LoSoTo`, we clipped outliers and smoothed the solutions in frequency-space with a Gaussian of full width half maximum (FWHM) covering four subbands. These solutions are then applied to full resolution data (32 channels per subband, 6 s).

Our last step was to create an image for each channel. With `wsclean`, a multi-frequency synthesis image was created for each subband, from which the clean components are extracted and used to create channel images of greater depth. Channel images were created with Briggs 0.0 weighting out to 11x11 square degrees field of view. We convolved every channel image to the same resolution of $236''$, a few percent larger than the lowest-resolution image, using `CASA` (McMullin et al. 2007). The flux density was then extracted from a fixed circular aperture of diameter $236''$, and a spectrum was created for each subband.

5.3 Searching RRLs in redshift space

The second main focus of this paper covers our method to search for RRLs blindly across redshift space. RRLs may not be detected individually, but wide-bandwidth observations enable detections as a result of stacking. Since the frequency spacing between each recombination line is unique ($\nu \propto \mathsf{n}^{-2}$), a unique redshift can be blindly determined with the detection of two or more lines. In stacking across redshift space, there are two main issues that require caution. The first is the low N statistics involved in the number of lines (10 – 30 spectral lines in HBA, 20–50 in LBA) used to determine the stack averaged profile. The second is the relatively small number of channels available to estimate the continuum in standard (64 channels or less per

subband) LOFAR observations. The method we employ does not depend on the unique set up of LOFAR and can be applied to observations with other telescopes.

The main steps of the method include:

1. assume a redshift and stack the spectra at the location of availables RRLs; repeat for a range of redshifts (see Section 5.3.1)

2. cross-correlate the observed spectrum in optical depth units with a template spectrum populated with Gaussian profiles at the location of the spectral lines for a given redshift; repeat over a range of redshifts (see Section 5.3.2)

3. cross-correlate the integrated optical depth of the stacked spectrum across redshift with the integrated optical depth of a template spectrum across redshift in order to corroborate mirror signals (see Section 5.3.3)

We compare the values of the normalized cross-correlations, and identify outliers assuming a normal distribution. Here we note that a single cross-correlation value is not necessarily meaningful in itself, but it is the relative comparison of the cross-correlation values across redshifts which identifies outliers. We require both cross-correlations result in a 5σ outlier at each redshift, as deemed necessary from simulated spectra (Section 5.3.4). Once a significant feature is identified by these means, we subtract the best fit of the RRL stack. We then repeat the procedure to search for additional transitions or components. In the sections below, we describe each step in further detail.

5.3.1 Stacking RRLs

We began processing the spectra by flagging. We manually flagged subbands with clearly poor bandpasses. We Doppler corrected the observed frequencies as Doppler tracking is not supported by LOFAR. We flagged additional edge channels that are affected by bandpass roll-off. Before removing the continuum, we flagged and interpolated over channels for which >50% of the visibility data were flagged as well as channels with a flux density greater than five times the standard deviation.

For a given redshift, we blanked the channels (assuming a certain line-width) at the expected frequency of the line when estimating the continuum. For stimulated transitions at low frequencies, we have that $I_{\rm line} \approx I_{\rm cont}\tau_{\rm line}$, where the intensity extracted from the observations is $I_{\rm obs} \approx I_{\rm line} + I_{\rm cont}$. Therefore, subtracting a continuum fit and dividing by it resulted in the optical depth, $(I_{\rm obs} - I_{\rm cont})/I_{\rm cont} = \tau_{\rm line}$. The continuum was fit with a 1st or 2nd order polynomial, chosen based on the χ^2 of the fit. Considering χ^2 of the fit and the rms of the subband, we flagged subbands which have outlying values.

Taking the central frequency of each subband, we found the RRL closest in frequency and used it to convert frequency units into velocity units. We then linearly interpolated the velocities to a fixed velocity grid, which has a frequency resolution equal to or greater than the coarsest resolution of all subbands. We weight subbands by their rms ($w = \sigma^{-2}$). We then stack-averaged all of the N subbands available,

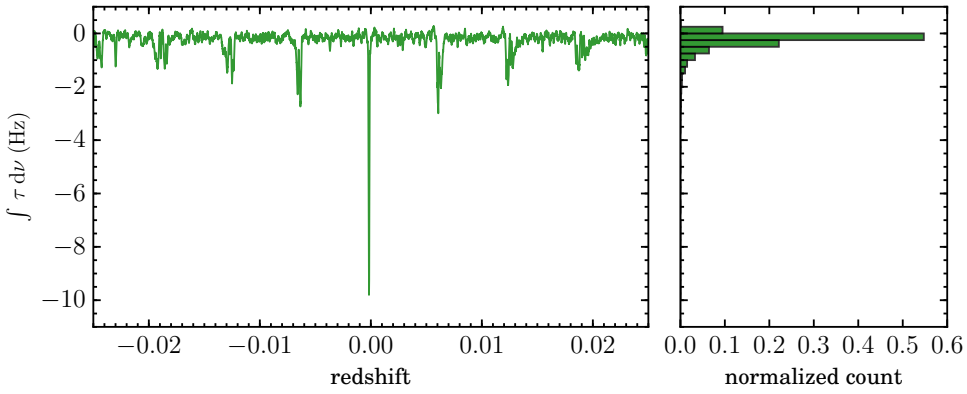


Figure 5.3: Cas A spectra has been stacked for Cα RRLs across a range of redshifts. The plot on the *left* shows the optical depth integrated at the central velocities (-7.7 km s^{-1} to +19.8 km s^{-1}, which encompasses the -47 and -38 km s^{-1} components) at each redshift. Here we clearly see an outlying peak at the redshift ($z = -0.000158$) of the -47 km s^{-1} component. The histogram on the *right* shows the binned distribution of integrated optical depth. Cas A α-transition stacking demonstrates our method in a high signal-to-noise regime.

where the optical depth of each channel is given by

$$< \tau_{\rm chan} >= \frac{\Sigma_{i=0}^{N}(w_i \tau_i)}{\Sigma_{i=0}^{N} w_i} \tag{5.1}$$

and i represents each subband going into the stack. An effective frequency, $\nu_{\rm eff}$, of the stacked spectral line was determined from the weighted mean frequencies among the lines stacked. We determined an effective principal quantum number, $\mathsf{n}_{\rm eff}$, by taking the integer quantum number of the line that was closest in frequency to $\nu_{\rm eff}$.

The error of each channel of the stacked spectra reflects the standard deviation of a weighted mean, which is $\sigma_{<\rm chan>} = (\Sigma_{i=0}^{N} w_i)^{-1/2}$. When comparing the integrated optical depth at each redshift, we integrated within a region one half of the blanked region and centered at $v_{\rm stack} = 0$ km s^{-1}. The spectral noise per channel of the stack at each redshift is determined by taking the weighted standard deviation of all channels outside the line-blanking region.

We show an example of stacking across redshift in Figure 5.3.

5.3.2 Spectral cross-correlation

For each redshift, we performed a cross-correlation between a template spectrum and the observed spectrum, both in units of optical depth and frequency. The template spectrum was populated with Gaussian line profiles at the location of the spectral lines that contributed to the final stack. The line profiles were normalized to a peak optical depth of unity and their FWHM was set by half the width of the line-blanked region. We then took the cross-correlation and normalized it proportionally with the number of lines that went into the stack, i.e. the total area under the template spectrum. This was the procedure that Morabito et al. (2014) implemented for a

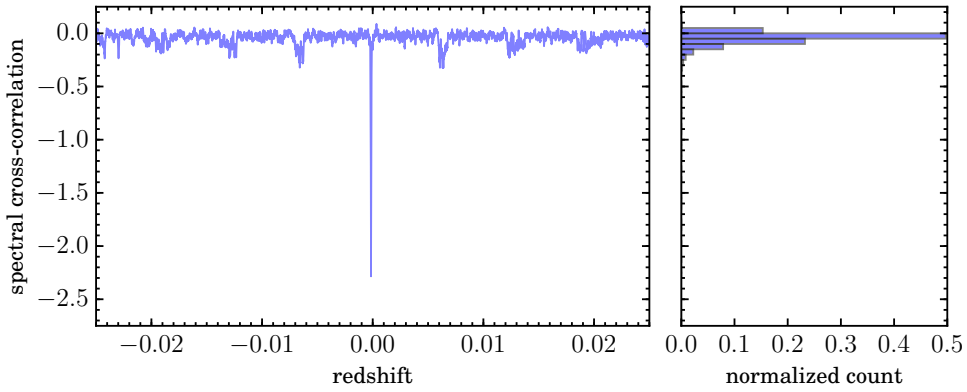


Figure 5.4: Cas A spectra have been cross-correlated for Cα RRLs with a template spectrum across a range of redshifts. The plot on the *left* shows the spectral cross-correlation value at each redshift. Again we clearly see an outlying peak at the redshift ($z = -0.000158$) of the -47 km s^{-1} component. The histogram on the *right* shows the binned distribution of cross-correlation values. Cas A α-transition stacking demonstrates our method in a high signal-to-noise regime.

single redshift, except we included a normalization since the number of lines stacked at each redshift did not remain constant. We show an example of cross-correlating the spectrum across a range of redshifts in Figure 5.4.

5.3.3 Stack cross-correlation

With a second cross-correlation, we corroborate "mirrors" of the signal that can be found at $(\Delta z)_{\rm mirror} = \overline{\Delta\nu_{\mathsf{n},\mathsf{n}+\mathsf{N}}}/\nu_{\mathsf{n}}$, or multiples thereof, where ν_{n} is the frequency of the effective n-level of the stack, and $\overline{\Delta\nu_{\mathsf{n},\mathsf{n}+\mathsf{N}}}$ is the average of the change in frequency between the effective n and all of the N other n-levels. Since the difference in frequency spacing between α-transitions n and $\mathsf{n}+1$ is small ($\sim$1 %), mirrors of the feature, which are more broadly distributed (in velocity space) and reduced in peak intensity, occur at offsets in redshift that match the frequency spacing between adjacent lines.

We identified redshifts with an outlying ($> 4\sigma$) value in the normalized spectral cross-correlation. For those redshifts, we took its template spectrum (see Section 5.3.2), stacked the template lines at an assumed redshift, $z_{\rm cen}$, and integrated the optical depth. We then stacked and integrated the template spectrum at a range of redshifts covering $\approx z_{\rm cen} \pm (\Delta z)_{\rm mirror}$. We then cross-correlated (a) the integrated optical depth of the template stack as a function of redshift, with (b) the observed integrated optical depth of the data at each redshift (see Figures 5.5 & 5.6). We found it best to restrict $(\Delta z)_{\rm mirror}$ such that only one mirror of the feature was present.

5.3.4 Validation with Synthetic Spectra

We developed this procedure first on synthetic spectra. The synthetic spectra were constructed by injecting Gaussian noise, distributed about an optical depth of 0. Radio recombination line profiles were populated at frequencies covering 110 – 160

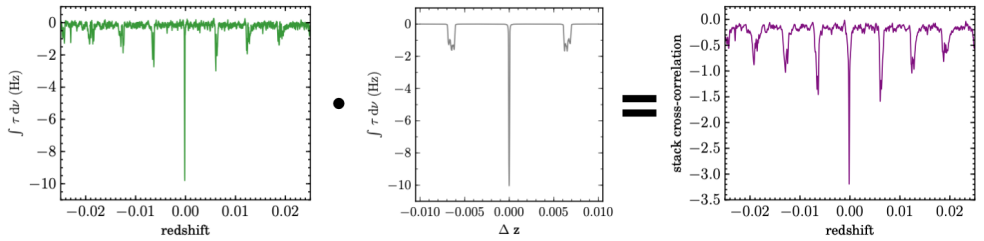


Figure 5.5: Here we demonstrate the stack cross-correlation. Cas A spectra have been stacked for Cα RRLs across a range of redshifts. The plot on the *left* shows the optical depth integrated at the central velocities at each redshift. The *middle* plot shows the integrated optical depth resulting from stacking the template spectrum. It has been done for a redshift of $z = 0.000158$. We cross-correlate the left plot with the middle plot to obtain the stack cross-correlation on the *right*.

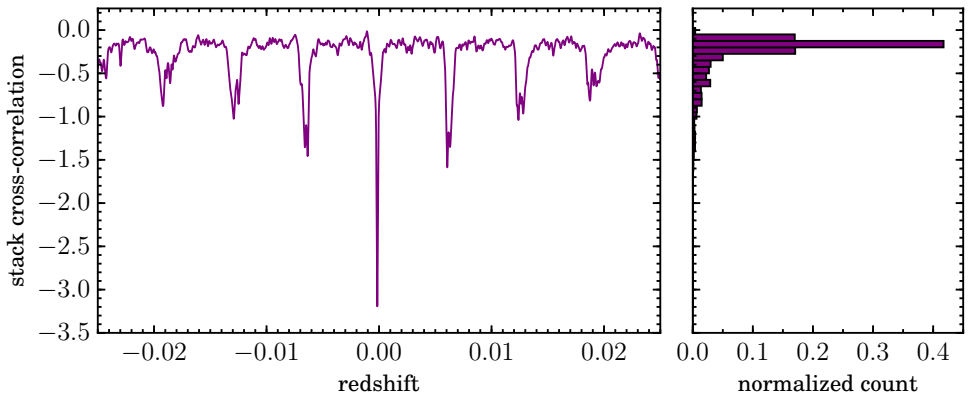


Figure 5.6: The integrated optical depth of the observed Cas A spectra is cross-correlated with the expected template spectrum. Here we clearly see an outlying peak at the redshift ($z = -0.000158$) of the -47 km s^{-1} component. The histogram on the *right* shows the binned distribution of integrated optical depth. Cas A α-transition stacking demonstrates our method in a high signal-to-noise regime.

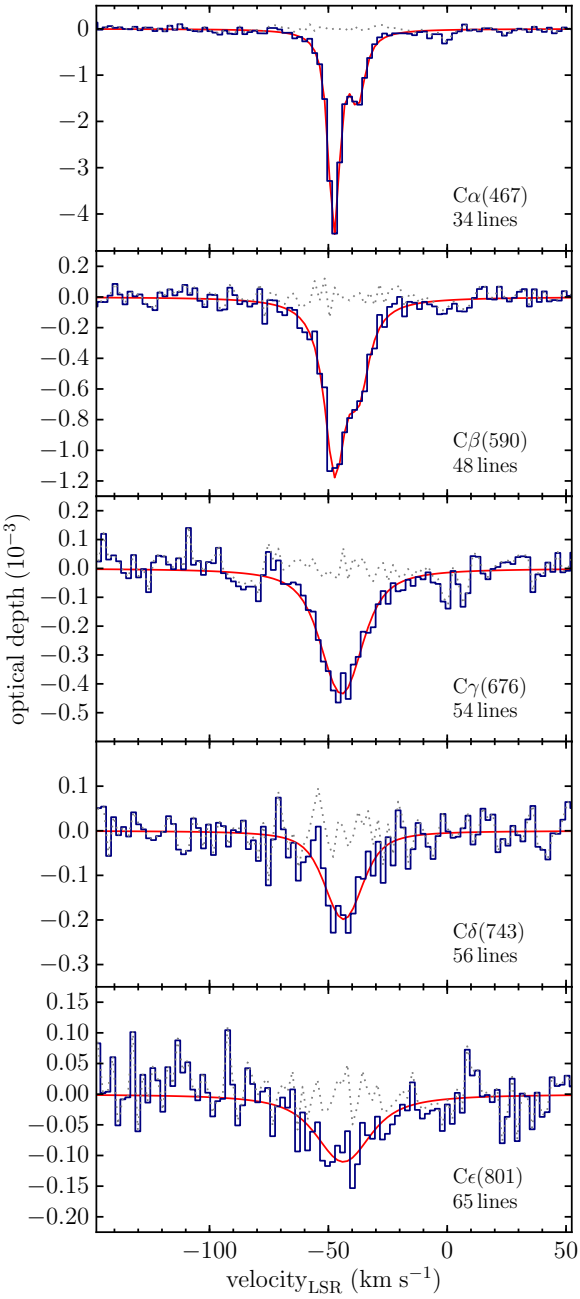


Figure 5.7: The stack averaged profiles for the carbon transitions associated with the $v_{\rm LSR} = -47$ km s^{-1} and $v_{\rm LSR} = -38$ km s^{-1} components in 55 MHz – 78 MHz observations of Cas A. These have all been significantly identified by our method.

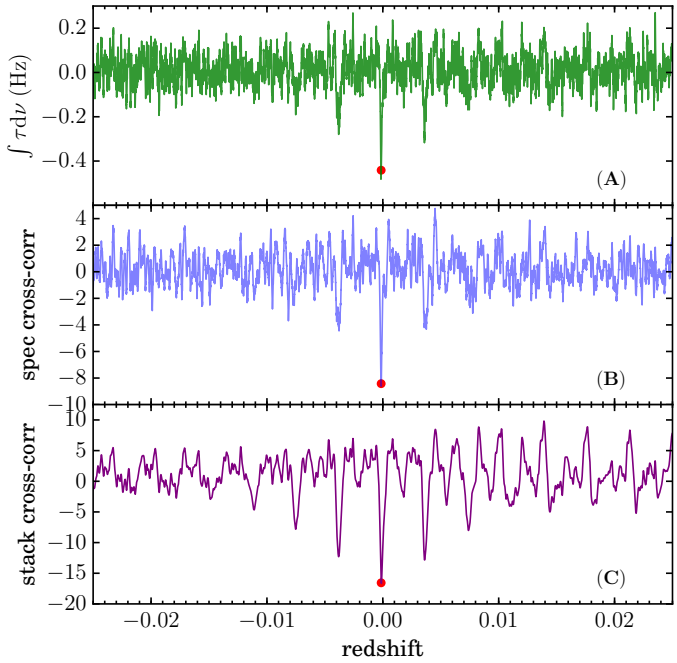


Figure 5.8: The three steps of our method applied to ϵ-transition stacking in the spectrum of Cas A: **(A)** the integrated optical depth at each redshift (see Section 5.3.1) where "mirrors" are seen at e.g. $z_m = \pm 0.004$, **(B)** the spectral cross-correlation (see Section 5.3.2), and **(C)** the stack cross-correlation (see Section 5.3.3). This example demonstrates the behavior of the method in the presence of broad RRLs.

MHz. Fifteen trials were made; the trials covered a variety of physical conditions (line properties), noise (signal-to-noise) regimes, and redshifts out to $z = 2$. Note, we did not insert continuum flux and started the procedure from the continuum subtraction stage. The tests were focused on a low signal-to-noise regime where continuum estimation and removal do not substantially affect the noise properties of the stacked line results.

Results with these synthetic spectra showed that radio recombination lines which resulted in a stacked line profile whose Gaussian fit had a peak signal-to-noise of 2.7σ could be successfully recovered in the cross-correlations at the 5σ level. The results of these tests also showed that it was essential for both the spectral and the stack cross-correlations to achieve a 5σ outlier at the input redshift. When these criteria were satisfied no false positives were recovered.

5.4 Cassiopeia A

The line-of-sight towards the supernova remnant Cas A has been the focus of detailed investigations of low-frequency RRLs in our Galaxy. We make use of LOFAR spectra between 55 MHz – 80 MHz first presented in O17. The flux densities ($\sim 2 \times 10^4$ Jy at 55 MHz) were extracted from a 14 x 14 arcmin^2 aperture, covering the entire source. The spectra have a channel width of 0.381 kHz (512 channels per subband), which corresponds to velocity resolutions between 1.4 km s^{-1} – 2.1 km s^{-1} over this frequency range. Starting from spectra in units of flux density that had been Doppler corrected to the Local Standard of Rest (LSR) reference frame, we implemented flagging as described in Section 5.3.1. Since the velocities of the components are known before hand, we directly specified the line-blanking regions then estimated the continuum using the line-free regions. With spectra in optical depth units, we iterated through the steps of our method. This differs from the implementation of M82 and 3C 190 in that it does not include continuum subtraction for each redshift. During stacking, we interpolate to a spectrum with a velocity resolution of 2.1 km s^{-1}, and we cover the redshift ranges $-0.025 \leq z \leq 0.025$, Nyquist sampling redshift intervals, $\delta z_{\rm sample} = 3 \times 10^{-6}$, which corresponds to about 0.9 km s^{-1}.

We initiated the procedure by stacking for the most prominent spectral lines in the spectrum, the Cα transitions of the -47 km s^{-1} and -38 km s^{-1}velocity components associated with gas in the Perseus Arm of our Galaxy. Two components are clearly distinguishable, but they overlap with one another. Therefore we defined a line-blanking region (-7.7 km s^{-1}, 19.8 km s^{-1}) that was centered on the most prominent component, -47 km s^{-1}, but encapsulated both components. The integrated optical depth which resulted from stacking for Cα RRLs at each redshift is shown in Figure 5.3. The redshift ($z_{-47{\rm km/s}} = -0.000151$) corresponding to $v_{\rm LSR}$ = -47 km s^{-1} is most prominent. Mirrors of the signal are clearly visible at multiples of $z_m \approx \pm 0.007$; they degrade in peak intensity the larger the multiple of z_m. Figure 5.4 shows the spectral cross-correlation. The template spectrum has been populated with two Voigt fits (see Figure 5.7, Table 5.1) to the -47 km s^{-1} and -38 km s^{-1} components rather than Gaussians with a FWHM in proportion to the line-blanked region, as line-broadening due to radiation and pressure broadening (for more details see Sec. 5.4.1) are present

Table 5.1: Properties of the carbon line profiles that have been detected in the spectrum of Cas A.

Line	n range	Frequency (MHz)	Line center (km s^{-1})	Lorentz FWHM (km s^{-1})	Total FWHM (km s^{-1})	$\int \tau \, d\nu$ (Hz)
Cα(467)	436 – 489	64.38	-47.6 ± 0.9	5.0 ± 0.3	6.1 ± 0.7	6.7 ± 0.8
			-37.7 ± 1.0	6.7 ± 1.1	7.8 ± 1.7	3.7 ± 0.9
Cα(469)	436 – 489	63.41	-0.5 ± 1.2	-	3.6 ± 0.5	0.23 ± 0.05
Cβ(590)	549 – 616	63.79	-47.3 ± 1.1	10.0 ± 0.5	10.7 ± 0.8	5.8 ± 0.5
			-37.2 ± 1.3	9.9 ± 1.1	10.7 ± 1.5	2.7 ± 0.5
Cγ(676)	628 – 705	63.53	-44.4 ± 1.2	14.2 ± 2.3	22.8 ± 3.1	7.8 ± 1.1
Cδ(743)	691 – 775	63.77	-43.6 ± 1.5	11.4 ± 1.7	21.2 ± 3.2	2.8 ± 0.5
Cϵ(801)	744 – 835	63.53	-43.7 ± 2.0	22.2 ± 3.2	29.4 ± 3.0	3.7 ± 0.5

The uncertainties quoted are 1σ.

in these lines (O17).

We stack the template spectrum over a redshift range of $z_{-47kms} \pm 0.01$, as shown in Figure 5.5. The result of cross-correlating this template function with the integrated signal vs redshift of the true spectrum is shown in Figure 5.6. In this high signal-to-noise test case, all three steps of the method identify the -47 km s^{-1} component with large significance. To subtract this component (and make more sensitive searches for additional components), we grouped the spectral lines into six different stacks and subtracted the best fit at the location of the spectral line. Grouping lines into six stacks minimized residuals due to slow changes in line properties at different frequencies. We then repeated the procedure, stacking for Cβ, Cγ, Cδ, Cϵ, Cζ, and again Cα. After each transition was tested for and significantly identified, we subtracted the best fit line profile, and then continued with testing the next transition.

5.4.1 Carbon RRL Results

We significantly identify the presence of six unique transitions associated with the -47 (and -38) km s^{-1} component(s) of the Perseus Arm of the Galaxy and one transition associated with a 0 km s^{-1}velocity component in the Orion Arm of the Galaxy. In addition to Cα associated with the -47 and -38 km s^{-1} components, we show the detections of Cβ, Cγ, and Cδ at this frequency for the first time. As a first for low-frequency RRLs, we have also significantly identified (7.8σ) the Cϵ transition. The stacked spectra and line profiles are shown in Figure 5.7. We show the steps of our method that recover the ϵ-transitions in Figure 5.8. The properties of all Cas A line profiles can be found in Table 5.1. The errors of each quantity were determined from the variance of each variable as determined by the fit. When fitting the β lines, the Gaussian width was fixed to the values derived from the α lines: 1.2 km s^{-1} and 1.1 km s^{-1} for the -47 and -38 km s^{-1} components, respectively. Additionally, the best fit of the γ line was used to set the Gaussian width of the δ and ϵ lines at 6.0 km s^{-1}since the two velocity components are blended for these transitions.

The CRRLs towards Cas A have been exhaustively analyzed in O17, S17, and S18. From the data available at $\mathsf{n} < 580$ for the -47 km s^{-1} gas component, spanning roughly 30 MHz – 1 GHz, densities and temperatures have been determined to be $n_e(-47) = 0.04 \pm 0.005$ cm^{-3}, $T_e(-47) = 85 \pm 5$ K with a path-length of $L_{C+}(-47) = 35.3 \pm 1.2$ pc (O17). Similarly the -38 km s^{-1} component was found to have the properties $n_e(-38) = 0.04 \pm 0.005$ cm^{-3}, $T_e(-38) = 85 \pm 10$ K, and $L_{C+}(-38) = 18.6 \pm 1.6$ pc. For principal quantum numbers $\mathsf{n} > 550$ line blending has not allowed for the two velocity components to be analyzed independently. S17 analyzed CRRLs present at 10–33 MHz from the blend of these velocity components. Slightly less dense gas with a somewhat stronger radiation field dominated the absorption lines at very low frequencies: $T_e = 60 - 98$ K, $T_{r,100} = 1500 - 1650$ K and $n_e = 0.02 - 0.035$ cm^{-3}. S17 note the difference between the RRLs observed at low and high quantum number could arise from (1) a variation in the spectral index of the background continuum source across the 5′ face, weighting the gas differently at higher and lower frequencies, (2) different cloud depths and thus the single slab with which the gas was model would not hold, and (3) the physical conditions of the -38 and -47 km s^{-1} components may not be the same.

Here, we focus on the new data uncovered by our stacking procedure. In Figure 5.9, we compare the observed line-width of the CRRLs as a function of principal quantum number with models for line broadening due to the Doppler effect, pressure broadening, as well as radiation broadening (S17). The line-widths of α and β lines are measured from only the -47 km s^{-1} component, where as the line-widths of the γ, δ, and ϵ lines are a blend of the -47 and -38 km s^{-1} components. Our results show that the α, β, and γ line-widths are slightly over-estimated compared with prior measurements, although well within error. The broad line widths we measure are likely contaminated by higher order transitions which have not been subtracted from the data and by the blending of the two velocity components. In the more detailed analyses of S17 (also applied to O17) and Stepkin et al. (2007), the spectral lines were subtracted and an additional baseline correction was performed before re-stacking for the final spectra. S17 goes on to validate the line-width measurement error and the integrated optical depth error by applying the approach to synthetic spectra. They confirmed that for principal quantum numbers $\mathsf{n} < 800$ the line properties were reproduced accurately to within 16 percent; this also applies to the O17 results. Since a validation of this kind goes beyond the scope of this paper, we underline the greater certainty in the detailed analysis of the S17 measurements.

Figure 5.10 shows the integrated optical depth as a function of principal quantum number together with model constraints, adapted from the literature compilation of S17. The integrated optical depth reflects the sum of -47 and -38 km s^{-1}components. We converted the integrated optical depth of higher order transitions into an equivalent α optical depths through $I_{\Delta\mathsf{n}=1} = I_{\Delta\mathsf{n}} \frac{\Delta\mathsf{n} \cdot M_{\Delta\mathsf{n}}}{M_{\Delta\mathsf{n}=1}}$, where $\Delta\mathsf{n} = \mathsf{n}\prime - \mathsf{n}$ and $M_{\Delta\mathsf{n}}$ is the approximate oscillator strength (Menzel 1968). We note that at high n involved here, difference in non-LTE effects are considered negligible ($\sim 1\%$). The integrated optical depths that we measure of the higher order ($\Delta\mathsf{n} > 1$) transitions generally lie above existing values. Again, since the maximum error induced by the processing procedure has been quantified in S17, we defer to those measurements. Our values are most likely over-estimated primarily due to residuals of e.g. α (see Figure 5.11) and β spectral lines which were not perfectly subtracted. The scatter in the literature for $\mathsf{n} \gtrsim 500$ reflects the difficulty in determining the baseline of the continuum in broad, overlapping spectral lines.

Lastly, we turn our attention to the Cα component at $v_{\rm LSR} = 0$ km s^{-1} associated with the Orion Arm. While it was not reported by O17 at these frequencies, we significantly identify it (5.2σ) here with a rather faint optical depth of $\int \tau {\rm d}\nu = (0.23 \pm 0.05)$ Hz (see Figure 5.11). Additional line properties can be found in Table 5.1. This result demonstrates the utility of our stacking methods for a narrow line in a low signal-to-noise regime. In Figure 5.11 **(C)**, we see that low-level residuals from the imperfect subtraction of the bright -47 km s^{-1} component is present in the stack cross-correlation.

There are now five detections in total of CRRLs associated with the 0 km s^{-1} component. A component at this velocity was clearly visible in WSRT P-band data at $\mathsf{n} = 267$ and in LOFAR 33 – 45 MHz data ($\mathsf{n} \sim 588$) (O17). Stepkin et al. (2007) also find an α and β line centered at $v_{\rm LSR} = -1.6$ km s^{-1}, blended with the -47 km s^{-1}component in their low 26 MHz observations using the UTR-2 telescope. Visual comparisons of this component with other cold diffuse gas tracers, such as ^{13}CO(1-

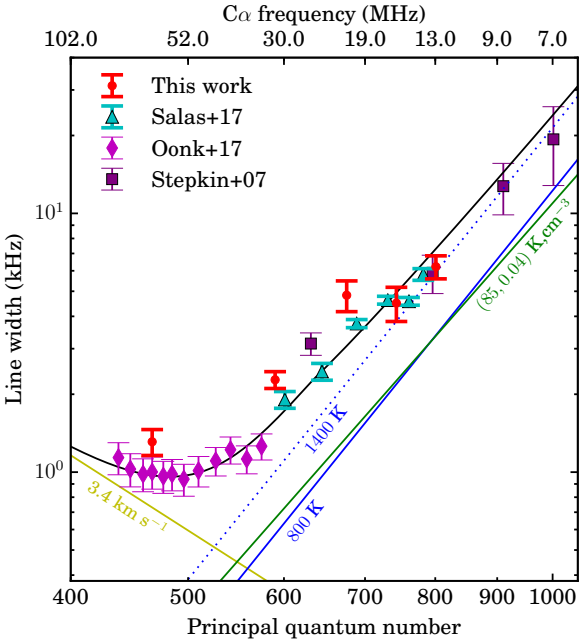


Figure 5.9: Line-width for the -47 km s^{-1} velocity component (or the sum of blended -47 and -38 km s^{-1} components) as a function of principal quantum number. The red points show the measured line-widths from this work (Table 5.1); the Cα(467) and Cβ(590) are derived only from the -47 km s^{-1} velocity component, while the higher order transitions represent a single fit to the blended -47 and -38 km s^{-1} components. The cyan triangles show the line-widths of the Cα and Cβ transitions of the -47 km s^{-1} velocity component from S17. The purple diamonds show the line-widths of the -47 km s^{-1} component from O17. The purple squares show the Cα, Cβ, Cγ and Cδ data, for which the -47 and -38 km s^{-1} components are blended, from Stepkin et al. (2007). The colored lines show the contribution from Doppler broadening (yellow line), pressure broadening (green line) and radiation broadening (blue lines). The solid black line shows the model which best fit the line-widths from S17: a Doppler line-width of 3.4 km s^{-1}, $T_e = 85$ K, $n_e = 0.04$ cm^{-3} and a radiation field which is a combination of a power law with $T_{r,100} = 800$ K and $\alpha = -2.6$ plus a contribution to the radiation field from Cas A.

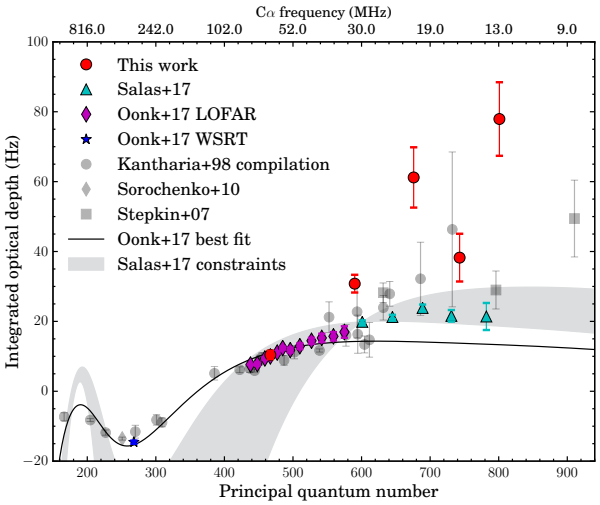


Figure 5.10: Integrated optical depth as a function of principal quantum number for the sum of the Perseus arm components at -47 and -38 km s^{-1}adapted from S17. The cyan triangles show the 10–33 MHz LOFAR observations of S17. The purple diamonds represent the 33–78 MHz LOFAR detections and the blue star shows the WSRT detection (O17). Pre-LOFAR literature data are shown in gray data points (Kantharia et al. 1998; Stepkin et al. 2007; Sorochenko & Smirnov 2010). The black solid line is the best-fitting model from O17. The gray shaded region covers the models which correspond to the physical constraints from S17.

0) and [CI] (S18) and HI (Payne et al. 1989), provide additional evidence for a cold neutral medium (CNM) component. We compared detections of the -47 km s^{-1} CRRL component with these in the Orion Arm by looking at the integrated optical depth as a function of principal quantum number. After scaling the Orion detections by a factor of 20 (to match values at $\mathsf{n} \sim 500$), there are indications that the transition from absorption to emission occurs at higher n in this component, indicating that the gas may be less dense and/or have a warmer radiation field (e.g. see constraints of Figure 5.10). However, the re-scaled values and errors of these low signal-to-noise detections are within 3σ of the -47 km s^{-1} component and therefore consistent with its derived properties. Deep observations, particularly at higher frequencies where the line is in emission, would provide useful constraints for future modeling.

5.4.2 Hydrogen RRL Results

We also report the detection of an Hα emission line as a result of stacking 43 lines at an average frequency of 64.08 MHz, one of the lowest frequency detections to date (see also Oonk et al. 2019, towards the Galactic Center at 63 MHz) and a valuable probe of the cold partially ionized phase of the ISM. This feature (see Table 5.2 for line properties) has been significantly identified in each step of our method (see Figure 5.11), most prominently in the stack cross-correlation with a 5.4σ confidence. Figure 5.11 shows the feature identified at a redshift $z = 0.000337$ or $v_{\rm LSR} = 101$ km s^{-1} as these stacks are relative to Cα RRL, for which the Hα transitions are

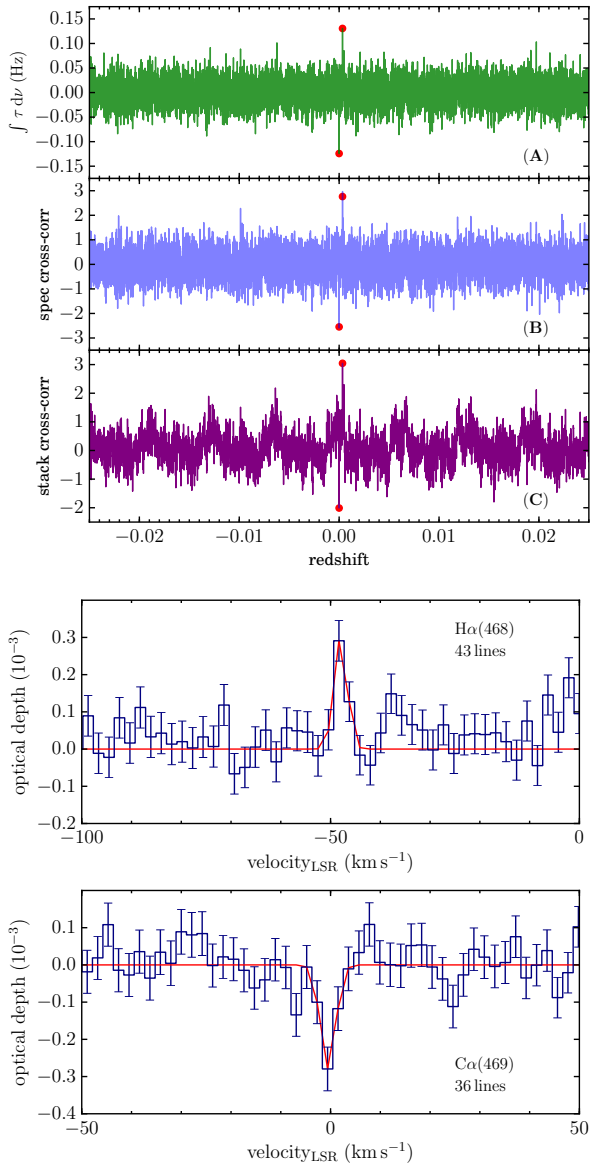


Figure 5.11: *Left:* Same as in Figure 5.8, except applied to Cα-transitions in Cas A spectra. Fits to the -47 km s^{-1} and -38 km s^{-1}components (and all higher order transitions) have been subtracted in the spectrum of Cas A. Residuals remaining from imperfect subtraction of the brightest α transitions from -47 km s^{-1} and -38 km s^{-1}components can be see in **(C)**. These results demonstrate the behavior of the method in the presence of narrow, low signal-to-noise spectral lines. The redshifts ($z_H = 0.000337$ and $z_C = -0.000002$, with respect to stacking for Cα) of the two detections are shown in red. *Right:* The stacked spectrum of Hα (top) associated with the -47 km s^{-1} component, one of the lowest frequency detections of Hydrogen RRLs. Cα (bottom) associated with the Orion Arm, detected at this frequency for the first time.

Table 5.2: Hydrogen RRL detections towards Cas A.

n	Frequency (MHz)	Line center (km s^{-1})	FWHM (km s^{-1})	$\int \tau \, d\nu$ (Hz)
250	418.35	-47.9 ± 0.2	4.6 ± 0.3	-1.98 ± 0.24
267	343.7	-47.4 ± 0.14	3.81 ± 0.34	-1.96 ± 0.15
309	222	-	-	-1.15 ± 0.58
468	64.08	-47.9 ± 1.1	3.1 ± 0.5	-0.21 ± 0.04

In this work we report the Hα(468) detection. We also provide the parameters of Hα(250) (Sorochenko & Smirnov 2010), Hα(267) (O17), and Hα(309) (Oonk et al. 2015).

regularly separated by 149.4 km s^{-1}. With respect to Hα RRLs, the central velocity is $v_{\rm LSR} = -47.9 \pm 1.1$ km s^{-1}. We find a feature of $\int \tau d\nu = -0.21 \pm 0.04$ Hz, consistent with a 3σ upper limit of -0.42 Hz in this frequency range (O17).

The line-width is narrow (FWHM $= 3.1 \pm 0.5$ km s^{-1}) in comparison to our channel resolution (2.1 km s^{-1}); considering that we have interpolated the spectra to a fixed velocity grid, it is plausible that the peak of the line is underestimated and the width is overestimated, while preserving the integrated optical depth. However, we note that our line-width is consistent within error of the FWHM of lines at higher frequencies (Table 5.2).

The measured line-widths indicate that pressure (and radiation) broadening are not dominant effects to the line-width at this frequency, since they would cause an increase in line-width towards lower frequencies. Instead the roughly constant line-width indicates a Doppler broadened profile. Assuming a purely Doppler broadened line-profile, an upper limit on the gas temperature is given by $T_e < 2 \times 10^4 {\rm K} \left(\frac{30.25 \, {\rm km \, s^{-1}}}{\Delta v} \right)^2$ (Equation 4.7, Brocklehurst & Seaton 1972) assuming hydrogen gas, where Δv is the FWHM in units of km s^{-1}. We derive an upper limit of $T_e < 210 \pm 0.5$ K, directly attributing this feature to the cold neutral medium.

We can also place a strict upper limit on the electron density if we assume the line-width is set by collisional broadening. Under this assumption, we can solve for the electron density in units of cm^{-3} as $n_e = \Delta\nu_{\rm col} \, (10^a {\rm n}^\gamma / \pi)^{-1}$, where $\Delta\nu_{\rm col}$ is the FWHM in units of Hz, n is the principal quantum number, and a and γ depend on the gas temperature (Salgado et al. 2017b). We refer to Salgado et al. (2017b) where the collisional coefficients were tabulated as $a = -9.620$ and $\gamma = 5.228$ for a temperature of $T_e = 200$ K, a temperature which is within 5% of our upper limit from the Doppler profile. We find an upper limit to the electron density of $n_e < 0.10 \pm 0.02$ cm^{-3}.

HRRLs at this frequency were searched for in O17 but went undetected. Nonetheless, with their higher frequency detection and two literature values, the HRRL gas properties were modeled. O17 found that the same physical conditions which best described the cold, diffuse CRRL gas ($T_e = 85$ K and $n_e = 0.04$ cm^{-3}) did not best fit the HRRL emitting gas. Alternatively, the models suggested that the hydrogen RRLs are arising from a colder and denser gas ($T_e = 30 - 50$ K and $n_e = 0.065 - 0.11$

cm^{-3}).

We re-model the physical conditions (with the procedure described in O17) to derive updated constraints on the physical properties of the HRRL gas (see Figure 5.12). The results provide further evidence for a component with distinctly different conditions than the CRRL gas: an electron temperature of $T_e = 30-50$ K, an electron density of $n_e = 0.045-0.75$ cm^{-3}, an emission measure of $EM_{H+} = 0.00064-0.0018$ pc cm^{-6}, and a radiation temperature of $T_{r,100} = 800-2000$ K. In Figure 5.12, we plot the best fit — $T_e = 40$ K, $n_e = 0.06$ cm^{-3}, and $EM_{H+} = 0.0012$ pc cm^{-6} — as a function of integrated line strength and principal quantum number. Figure 5.13 plots all line-width and modelling constraints of the HRRL gas together with the CRRL properties.

Under the assumptions that (1) HRRLs and CRRLs originate from the same gas phase and (2) HRRLs are only ionized by cosmic rays (CR), constraints on the CR ionization rate can be determined from the ratio of the Hydrogen and Carbon RRL integrated optical depths (e.g. Neufeld & Wolfire 2017; Sorochenko & Smirnov 2010). The results from our models challenge assumption (1). While the line-widths and central velocities of the two tracers do not require different phases, it is possible that the gas motion is perpendicular to the line-of-sight. A high spatial resolution ($2'$) analysis of the CRRLs towards Cas A (S18) highlighted additional relevant effects: there are variations in peak line-strength by as much as a factor of seven within this region, and through comparisons with other cold gas tracers, a spatially resolved transition of the gas from a diffuse atomic state to a dense molecular cloud was identified. Spatially resolving HRRLs across the region would allow us to structurally locate the emission and how its intensity is distributed. In conclusion, a higher resolution analysis of HRRLs is needed to determine the validity of assumptions (1) and (2) in constraining the CR ionization rate.

5.5 M 82

M 82 is a prototypical starburst galaxy located ∼3.5 Mpc away (Jacobs et al. 2009), with a systemic velocity of $v_{\mathrm{sys}} = 209 \pm 4$ km s^{-1} (Kerr & Lynden-Bell 1986). It was observed with the LOFAR LBA and reported to have CRRLs in absorption (M14) centered at $v_{\mathrm{LSR}} = 211.3^{+0.7}_{-0.5}$ km s^{-1} or $z_{\mathrm{cen}} = 0.00070$. The feature has a peak optical depth of $2.8^{+0.12}_{-0.10} \times 10^{-3}$ and a FWHM of $30.6^{+2.3}_{-1.0}$ km s^{-1}providing an integrated optical depth of $\int \tau \mathrm{d}\nu = 21.3$ Hz; errors were derived from the Gaussian fit. We were motivated to test our methods on the M 82 data as it makes for the most direct comparison to the 3C 190 spectra, where the spectral feature is narrow, has low signal-to-noise and the fraction of line-blanked channels to line-free channels ($\gtrsim 10\%$) starts to induce non-negligible effects.

We used the spectra extracted from the 50 MHz – 64 MHz observations of an unresolved M 82 (Morabito, priv. comm., M14). The data have a channel width of 6.10 kHz, which corresponds to 28.6 km s^{-1} – 36.7 km s^{-1} over this frequency range. We Doppler corrected the spectra to the Barycentric frame. During stacking, we interpolated each spectrum to a fixed velocity grid with a channel resolution of 36.7 km s^{-1}. Before implementing the flagging and stacking routine described in

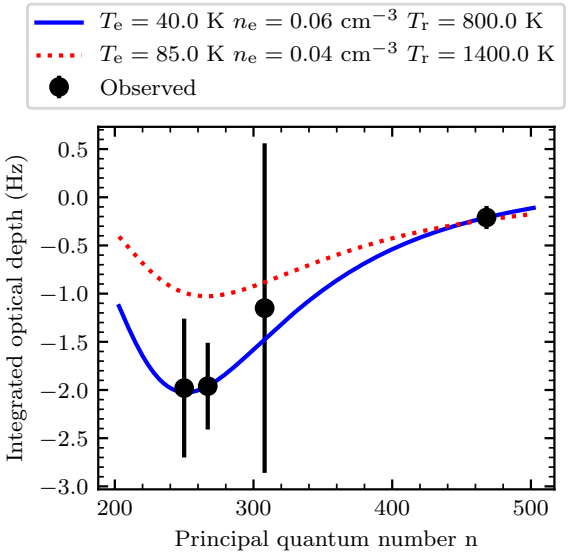


Figure 5.12: The integrated optical depth as a function of principal quantum number, n, for the Hydrogen RRLs detections associated with the -47 km s^{-1} velocity component towards Cas A. The figure shows the values from Table 5.2 using 3σ error bars. The blue, solid line shows the best fit model. The red, dotted line shows the best fit model (O17) to Carbon RRL detections spanning quantum numbers 225 – 550 (e.g. see Figure 5.10).

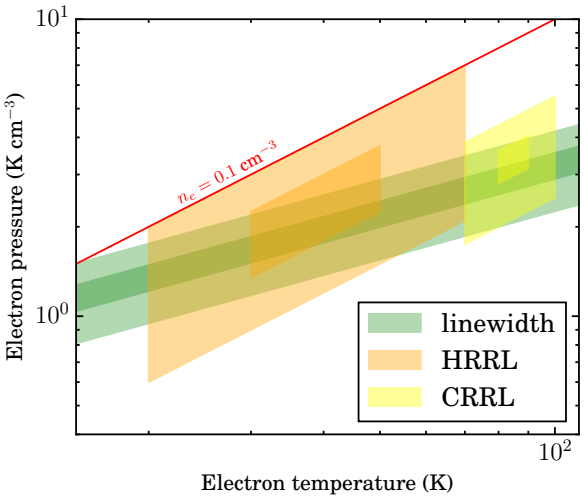


Figure 5.13: The electron pressure as a function of electron temperature of the -47 km s^{-1} velocity component towards Cas A. The red solid line shows the upper limit of a Pressure broadened line-width. Shaded regions represent 1σ and 3σ uncertainty. The green shaded regions show the line-width constraints from the combined Doppler, pressure and radiation broadening terms for $T_{r,100} = 1400$ K (S17). The orange shaded regions show the physical properties of the HRRLs constrained from modeling the integrated optical depth. The yellow shaded regions show the physical properties of the CRRLs constrained from modeling the integrated optical depth (O17). This plot shows the distinction between the HRRLs and CRRLs, where HRRLs arise from slightly colder, denser regions in the cloud.

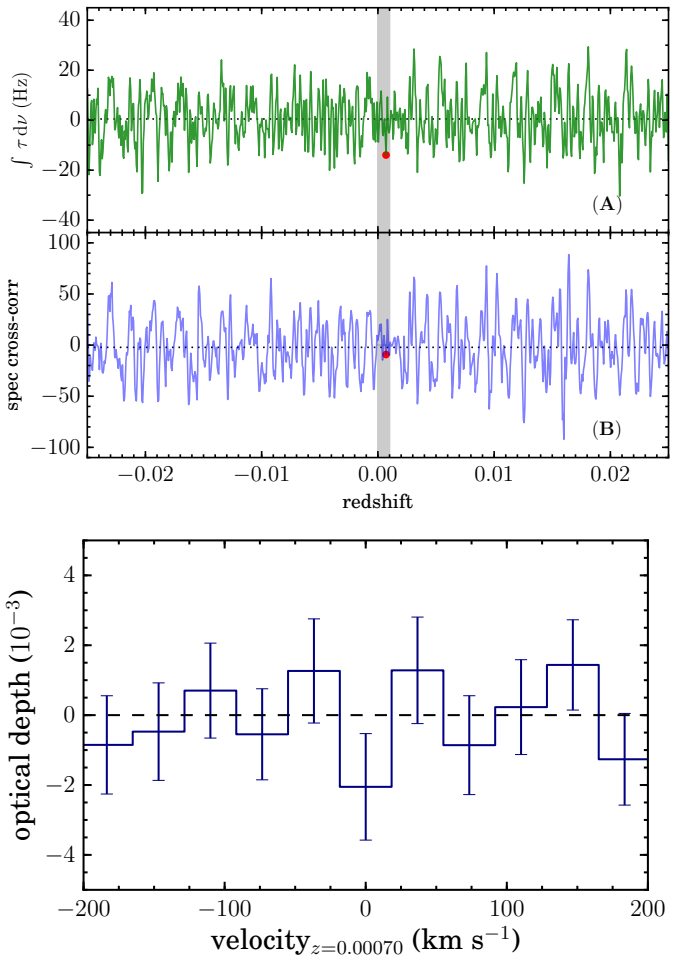


Figure 5.14: Our results for M 82. *Left:* Same as in Figure 5.8, except applied to the Cα-transitions of M 82. We did not proceed with the stack cross-correlation because no significant outlier was identified in (B). The shaded region in gray shows the redshift range probed by M14. The red data point shows marks $z_{\rm M82} = 0.00070$, the redshift corresponding to the peak of RRL emission, as reported by M14. *Right:* The stacked spectrum of M 82, in velocity units, with respect to $z_{\rm M82} = 0.00070$. The error bars reflect the standard deviation of a weighted mean (see Sec. 5.3.1)

Section 5.3.1, we flagged two channels at the starting edge (lower in frequency) of the subband and one channel at the ending edge (higher in frequency). We cover the redshift ranges $-0.025 \leq z \leq 0.025$ in stacking, sampled at redshift intervals of $\delta z_{\rm sample} = 6 \times 10^{-5}$. We considered line-blanking regions of ± 25 km s^{-1}, ± 50 km s^{-1}, and ± 100 km s^{-1}.

We did not find a significant outlier in the integrated optical depth across redshift or in the spectral cross-correlation (Sec. 5.3.2) using any of the line-blanking widths. An example, with a line-blanking region of ± 50 km s^{-1}, is shown in Figure 5.14. The redshift at which RRLs were reported in M14 is shown in the plots with a red, circular data point. Since we could not identify an outlier in the spectral cross-correlation, we did not proceed with the stack cross-correlation (Sec. 5.3.3).

In Figure 5.14 we also show the CRRL stacked spectrum centered at $z = 0.00070$. The optical depth in the central channel is $(2.1 \pm 1.5) \times 10^{-3}$. The standard deviation in the optical depth of channels within ± 200 km s^{-1} when excluding the central channel is $\sigma = 9.7 \times 10^{-4}$. This noise (which is an rms per channel) is about 1.5 times lower than the error attributed to the central channel. This mismatch arises from (1) interpolation and (2) non-uniform coverage across the channels. We compared the noise in the subband spectra before and after interpolation, and found that interpolation reduced the noise by 20 – 30 percent; an overall 6 percent was due to an effective averaging of higher resolution channels to the coarse grid. This affects all channels of the spectrum. Furthermore, the number of data points averaged in channels which are line free is 20 – 30 percent higher than in the line channels; this results in a lower rms in the line-free channels only. Accounting for these two effects results in an rms per channel of $\sigma = 1.4 \times 10^{-3}$, a peak optical depth of $(2.5 \pm 1.4) \times 10^{-3}$, and for an effective frequency of 55.5 MHz, an integrated optical depth of $\int \tau {\rm d}\nu = 18 \pm 10$ Hz.

The properties of the spectral feature that we find centered at $z = 0.00070$ are consistent with the values reported by M14. However, the significance relative to the noise is 1.8σ. When we compare the value of the integrated optical depth at $z = 0.00070$ with values obtained for all of the other redshifts tested (see Figure 5.14), its value is 1.5 times the standard deviation.

Moreover, the value of the spectral cross-correlation at $z = 0.00070$ does not appear significant in comparison to the values across the full range of redshifts (0.35 times the standard deviation of the cross-correlation values), nor in the redshift range probed by M14 (z = 0.0 – 0.001; see gray, shaded region in Figure 5.14). The latter may be unexpected, given that M14 found a peak in the cross-correlation when the stack was centered on a redshift of $z = 0.00073$. We note that the spectral cross-correlation we have implemented differs from the M14 approach in three ways, which we explain in the following paragraphs.

Firstly, a separate fit and subtraction of the continuum is done at each redshift. It is essential to include this step because, by definition, residuals of the fit have been minimized in the line-free channels; thus any spectral search in the "off" redshifts, the line-free channels, will be consistent with noise. Furthermore, the error within the line-blanked channels will be amplified since the continuum has not been directly estimated from those channels. This is especially true when the number of channels available to estimate the continuum (about 25 channels) is small (Sault 1994).

Secondly, we generate a new template spectrum for each redshift. The number

of spectral lines available for stacking varies across redshift. For example, there are 8 fewer spectral lines available to stack at $z = 0$ compared with $z = 0.00070$. The cross-correlation would naively appear lower at $z = 0$ only because fewer lines fall within a searchable region.

This leads to the third difference, which is that we have normalized the cross-correlation value by the number of spectral lines stacked (or equivalently, by the area of the cross-correlation function). Without a normalization, the cross-correlation values are not directly comparable for functions of different areas.

We went about reproducing the M14 results following their methods, in order to understand the limitations of our method and verify the results. The major change in the procedures implemented to obtain the line profile involves smoothing the combined subband spectra with a filter rather than interpolating the subband spectra to a fixed velocity grid and averaging. The minor changes involved in flagging were also included. We reproduced the spectrum of M14 and investigated the noise properties, the line properties, and the distribution in those values when stacked across redshift. We describe the procedure and discuss the results in the Appendix 5.B. The spectrum we obtained has line properties consistent with those of M14. However, with an updated noise estimate (Figure 5.19), the feature has a 2.2σ strength.

We conclude that deeper LOFAR observations will be needed to investigate CRRLs in the spectrum of M 82. Since the line-width of the reported feature is less than a channel width, the peak strength would be underestimated and width overestimated; processing the data at higher spectral resolution will also be necessary. A 3σ upper limit on the integrated optical depth is 26.5 Hz for a line-width of 36 km s^{-1}.

5.6 3C 190

We identified 3C 190 as a candidate to search for RRLs at high redshift as it is a bright radio source at LOFAR frequencies, compact (4 arcsec in size), and has been detected via HI absorption (Ishwara-Chandra et al. 2003) and Mg II absorption (Stockton & Ridgway 2001), which are both indicative of cold gas. The data were Doppler corrected to the Barycentric rest frame, and the stacked spectra were interpolated to a velocity resolution of 15 km s^{-1}. We searched the redshift ranges $-0.01 < z < 1.22$ and considering line-blanking regions of $\pm$7.5 km s^{-1}, $\pm$30 km s^{-1}, $\pm$45 km s^{-1}, and $\pm$60 km s^{-1}.

We find one significant outlier with the line-blanking region of $\pm$7.5 km s^{-1}. However, no outliers were found with other line-blanking regions. The results of the method when line-blanking $\pm$7.5 km s^{-1} are shown in Figure 5.15. We ultimately find α-transition RRLs at $z = 1.12355$ assuming carbon, as reported in Emig et al. (2019). In Figure 5.15, we also show the spectrum when extending the line-blanking region to $\pm$25 km s^{-1}as in Emig et al. (2019). We do not find a significant signal at the systemic redshift of 3C 190 or at the redshifts of the reported absorption features and place a 3σ upper limit of 4.6 Hz for a line width of $\pm$7.5 km s^{-1}. If the noise scaled in proportion to the line-blanking regions, we would expect an upper limit of 18.6 Hz for a line width of $\pm$45 km s^{-1}, however given the distribution of integrated optical depth, we find a 3σ upper limit of 22.7 Hz is more representative. This indicates

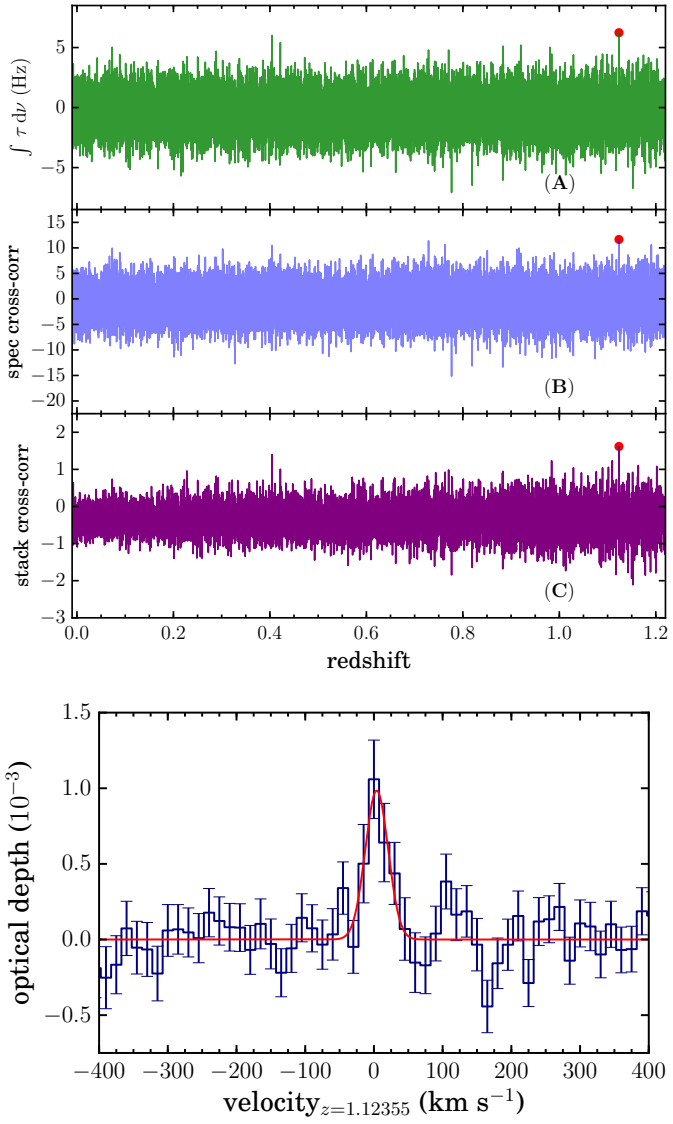


Figure 5.15: Our results for 3C 190. *Left:* Same as in Figure 5.8 except applied to the Cα-transitions of 3C 190. The red data point marks the significantly identified redshift of $z = 1.12355$. *Right:* The spectrum resulting from stacking Cα-transitions at $z = 1.12355$ in 3C 190, using a line-blanking region of ± 25 km s^{-1}.

that systematics such as the poly-phase filter residuals are present in the data, and noise is being amplified within the line-blanked region due to improper continuum subtraction. We find the noise to be amplified by a factor of 1.22 when roughly 4 of 26 channels are blanked.

5.7 Discussion of Methods

The data reduction strategy we have laid out covers basic calibration with an improvement on the bandpass calibration, which is of particular importance for RRL observations. Future strategies that include corrections for ionospheric effects, a phase offset present in some stations, as well as including stations with longer baselines to increase spatial resolution could be beneficial to the strategy here (van Weeren et al. 2016; Williams et al. 2016). Although with the set up and configuration processed here, these are not dominant effects. Prefactor[2] 3.0 would enable this and has been designed with spectroscopic studies in mind (de Gasperin et al. 2019).

We have presented a three step method to search for radio recombination lines most pertinently in a low signal-to-noise regime. This method can be applied to observations of any telescope with sufficiently large fractional bandwidth that allow for spectral stacking. This includes recombination lines at higher frequencies — for example, in the L band, where 17 α-transitions lie between rest frequencies of 1–2 GHz. We have demonstrated our method in a variety of regimes.

As a proof of concept, we showed the behavior of stacked recombination lines with high signal-to-noise (see Figures 5.3 – 5.6). When comparing the optical depth integrated within a fixed region for a range of input redshifts/velocities, we see the actual redshift is clearly the most prominent in comparison. However, since the lines are also detected with significance individually, redshifts corresponding to mirrors of the signal are of course, detected with significance as well. In this regime the spectral cross-correlation (Section 5.3.2) identifies the actual redshift most uniquely, whereas the stack cross-correlation (Section 5.3.3) is more sensitive to the stack mirrors and therefore the relative significance of the mirrors is enhanced.

We also show and examine low signal-to-noise features, and we do so in two regimes: broad features and narrow features. Broad features can be defined such that their line width is greater than the difference in frequency spacing between adjacent lines of principal quantum number n, $\Delta\nu_{\rm FWHM} \gtrsim (\Delta\nu_{\mathsf{n}-1,\mathsf{n}} - \Delta\nu_{\mathsf{n},\mathsf{n}+1})$. In this case, at redshifts corresponding to the mirrors of the stack, the slight misalignment in frequency/velocity space fall within the FWHM of the line, causing the ratio between the integrated signal and the mirrors to be closer to unity. An example of this regime is the ϵ-transition search, were the line width is $\Delta\nu_{\rm FWHM} = 6.4$ kHz and $(\Delta\nu_{\mathsf{n}-1,\mathsf{n}} - \Delta\nu_{\mathsf{n},\mathsf{n}+1}) = 1$ kHz for $\mathsf{n} = 801$. Conversely, narrow features we define as $\Delta\nu_{\rm FWHM} \lesssim (\Delta\nu_{\mathsf{n}-1,\mathsf{n}} - \Delta\nu_{\mathsf{n},\mathsf{n}+1})$, where the narrowness of the line causes minimal resonance in the mirrors, and the ratio of the integrated optical depth between z_{cen} and z_{mirror} approaches $1/N$ with N being the number of lines stacked. An example of this is Hα, where $\Delta\nu_{\rm FWHM} = 0.7$ kHz and $(\Delta\nu_{\mathsf{n}-1,\mathsf{n}} - \Delta\nu_{\mathsf{n},\mathsf{n}+1}) = 4.0$ kHz for $\mathsf{n} = 468$.

[2]https://github.com/lofar-astron/prefactor.git

When stacking broad spectral lines, the integrated optical depth at the redshift of the stack mirrors is maximally enhanced. Therefore, the stack mirrors may be clearly distinguished (see Figure 5.8) and the ratio between the integrated signal at the actual redshift and at the mirror redshifts is less extreme. In the presence of well behaved bandpasses and high N statistics, the integrated optical depth and spectral cross-correlation may be sufficient to identify the feature. However, the stack cross-correlation results in a distribution of redshifts, with uncertainties of $\delta z = \Delta\nu_{\mathsf{n},\mathsf{n}+1}/\nu_{\mathsf{n}}$. When, for example, bandpass estimation is distorted at "off" redshifts (by the very presence of the spectral lines) or in the case of low N statistics, additional scatter in the integrated optical depth could result and the use of the stack cross-correlation becomes essential.

On the other hand, narrow features produce only a low level mirror of the stack at flanking redshifts, and in all three steps of the method, only the actual redshift of the source (no mirror redshifts) appear as outliers (see Figure 5.11). In this regime, the stack cross-correlation becomes the most sensitive probe.

Useful tips that we have learned using our method include (1) the significance of the feature is maximized in each step of the procedure when the optical depth is integrated within a region roughly half the size of the line-blanking region; (2) it is essential to minimize residuals of high signal-to-noise features in order to reliably probe additional low signal-to-noise features; and, (3) the stack cross-correlation is most effective when only one mirror is included in the cross-correlation function.

Focusing briefly on the implications of integrated optical depth distributions, we find that the apparent amplification of noise shown in the distribution of integrated optical depth in M 82 and 3C 190 indicates that the noise in the stacks is not purely Gaussian; several effects may be at play. Systematics may be present in the spectra. The continuum removal may amplify noise within the line-blanked region, with too few channels to estimate the continuum. Lastly, fluctuations due to low N statistics and non-uniform coverage across channels may be contributing to additional scatter. For example, in Cas A the number of line-free channels is large (~ 200) and the fraction of line to continuum channels is typically ≤ 0.01, resulting in a negligible error of estimating the continuum within the line-region. The fraction of velocity channels in which the coverage is uniform is also larger. Indeed, across redshift, we find that the scatter in the peak optical depth of the stack matches the median channel noise.

Nonetheless, the detection of an RRL in the first AGN we searched with LOFAR holds promise for future exploration. For typical low-frequency RRL optical depths of $10^{-3} - 10^{-4}$, it is reasonable and feasible to search in the 3CR sources (Edge et al. 1959; Bennett 1962) of the northern hemisphere, which have flux densities > 9 Jy at 178 MHz. With 328 sources in the northern hemisphere and an average source density of 1.6×10^{-2} deg^{-2}, about 0.8 3CR sources will fall in a $\sim$50 deg^2 pointing of the LOFAR Two Meter Sky Survey (Shimwell et al. 2019). In light of our results above, it will be necessary to perform a continuum estimation and subtraction across multiple subbands, for the survey frequency resolution of 16 channels per subband, to reach adequate sensitivities.

5.8 Conclusion

RRLs are largely unexplored at low frequencies although they uniquely can provide long sought after physical conditions of the cold-neutral (and warm-ionized) phase(s) of the ISM. We have described methods to calibrate and extract RRLs in low-frequency (< 170 MHz) spectra. Starting with LOFAR observations that are optimized for extragalactic sources (where line widths may be 10 – 100 km s^{-1} and can plausibly be probed to $z \sim 4$), we discussed spectroscopic data reduction. We then showed a procedure in which spectra are stacked and cross-correlated to identify low signal-to-noise features. One cross-correlation is taken between a template spectrum and the observed spectrum, both in optical depth units, where the location of each line is utilized. A second cross-correlation incorporates the average spacing between lines (and their line width); the integrated optical depth over a range of redshifts is cross-correlated with the distribution of the template spectrum over a range of redshifts, corroborating what we refer to as "mirrors" of the stack at flanking redshifts.

Our method was developed to search blindly in redshift for RRLs in the LOFAR HBA spectrum of 3C 190, in which we have identified an RRL in emission at $z = 1.12355 \pm 0.00005$ (assuming a carbon origin). This was the first detection of RRLs outside of the local universe (Emig et al. 2019). To demonstrate and test the limitations of the method, we also apply it to existing LOFAR observations of the sources Cas A and M 82.

We have re-analyzed the 55 – 78 MHz LOFAR spectra of Cas A. Using our methods, we discover three new detections in the data, plus the original detections of Oonk et al. (2017). We significantly detect Cα(n=467), Cβ(590), Cγ(676), Cδ(743), and Cϵ(801) transitions associated with the line-of-sight -47 km s^{-1} and/or -38 km s^{-1} components. This is the first detection of an ϵ-transition (Δn = 5) at low radio frequencies. We also find Hα(468) in emission at 64.08 MHz with $\int \tau d\nu = (-0.21 \pm 0.04)$ Hz and a FWHM of 3.1 km s^{-1} resulting in one of the lowest frequency and most narrow detections of hydrogen. The line-width directly associates this hydrogen with the cold, ionized component of the ISM; this is further supported by our updated modeling of the gas physical properties with best fit conditions of $T_e = 40$ K, $n_e = 0.06$ cm^{-3}, and $EM_{H+} = 0.0012$ pc cm^{-6}. Additionally, we detect Cα associated with the Orion Arm at 0 km s^{-1}at these frequencies for the first time.

For the 55 – 64 MHz spectra of the nearby starburst galaxy M 82, we recover the line properties reported by Morabito et al. (2014) and find the integrated optical depth to be $\sim 2\sigma$ relative to the noise. A 3σ upper limit on the integrated optical depth is 26.5 Hz. Follow-up LOFAR observations reaching deeper sensitivities and higher spectral resolution will be worthwhile.

We find that LOFAR observations using 32 channels per subband is not optimal for RRL studies. Because the number of channels available to estimate the continuum is low, the noise in the line blanking region is amplified. Currently continuum subtraction can only be estimated within a single subband. As the non over-lapping nature of the narrow (195.3 kHz) subbands makes a smooth bandpass calibration difficult, future observing bands with larger contiguous frequency coverage would enable deeper searches of RRLs in extragalactic sources.

Acknowledgements

The authors would to thank Leah Morabito for providing the LOFAR LBA spectra of M 82 and for the discussions, and Reinout van Weeren for guidance and careful review of the manuscript.

KLE, PS, JBRO, HJAR and AGGMT acknowledge financial support from the Dutch Science Organization (NWO) through TOP grant 614.001.351. AGGMT acknowledges support through the Spinoza premier of the NWO. MCT acknowledges financial support from the NWO through funding of Allegro. FdG is supported by the VENI research programme with project number 639.041.542, which is financed by the NWO. Part of this work was carried out on the Dutch national e-infrastructure with the support of the SURF Cooperative through grant e-infra 160022 & 160152.

This paper is based (in part) on results obtained with International LOFAR Telescope (ILT) equipment under project codes `LC7_027, DDT002`. LOFAR (van Haarlem et al. 2013) is the Low Frequency Array designed and constructed by ASTRON. It has observing, data processing, and data storage facilities in several countries, that are owned by various parties (each with their own funding sources), and that are collectively operated by the ILT foundation under a joint scientific policy. The ILT resources have benefited from the following recent major funding sources: CNRS-INSU, Observatoire de Paris and Universite d'Orleans, France; BMBF, MIWF-NRW, MPG, Germany; Science Foundation Ireland (SFI), Department of Business, Enterprise and Innovation (DBEI), Ireland; NWO, The Netherlands; The Science and Technology Facilities Council, UK; Ministry of Science and Higher Education, Poland.

5.A Subband spectra of Cas A

In Figure 5.16, we show Cas A spectra for which the continuum has been removed, in units of optical depth as a function of frequency. These spectra are shown prior to stacking in order to illustrate typical properties of the observations. For example, each subplot shows the spectrum of a single subband and the channels remaining unflagged out of the original 512 channels per subband. How often spectral lines of the various transitions – α, β, γ, δ, ϵ – fall and how often they overlap can be grasped. In addition, the channels used to estimate the continuum are those outside of the shaded regions.

5.B Spectral properties of M 82 applying M14 criteria

In this section we first discuss the criteria implemented in M14 and compare it with the procedure generally used in this paper (Section 5.5). We reproduce the spectrum of M14 and analyze the noise properties for the same assumed redshift. Finally, using the M14 flagging and processing, we implement our three step method to search for a significant outlier.

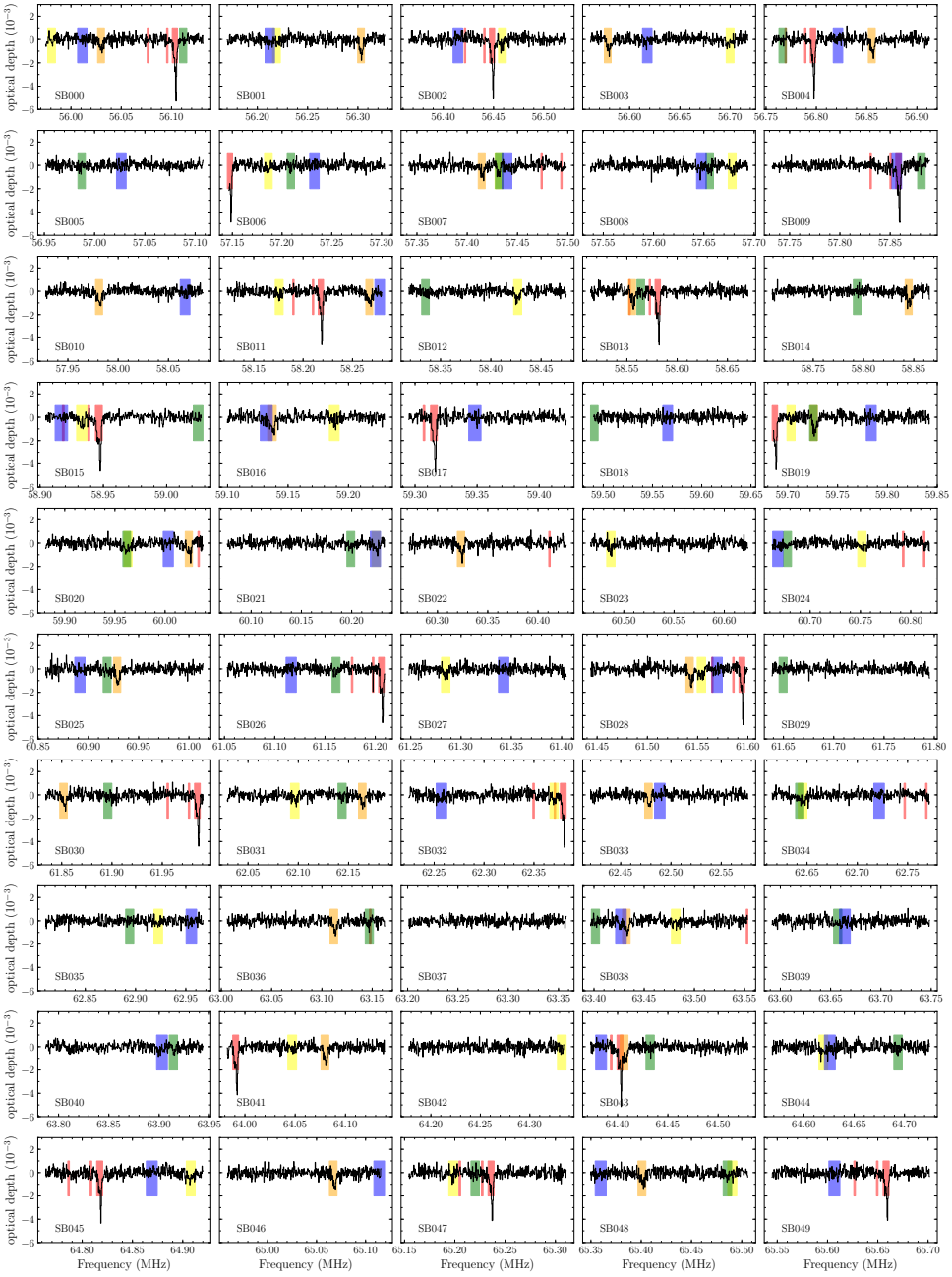


Figure 5.16: *(cont.)*

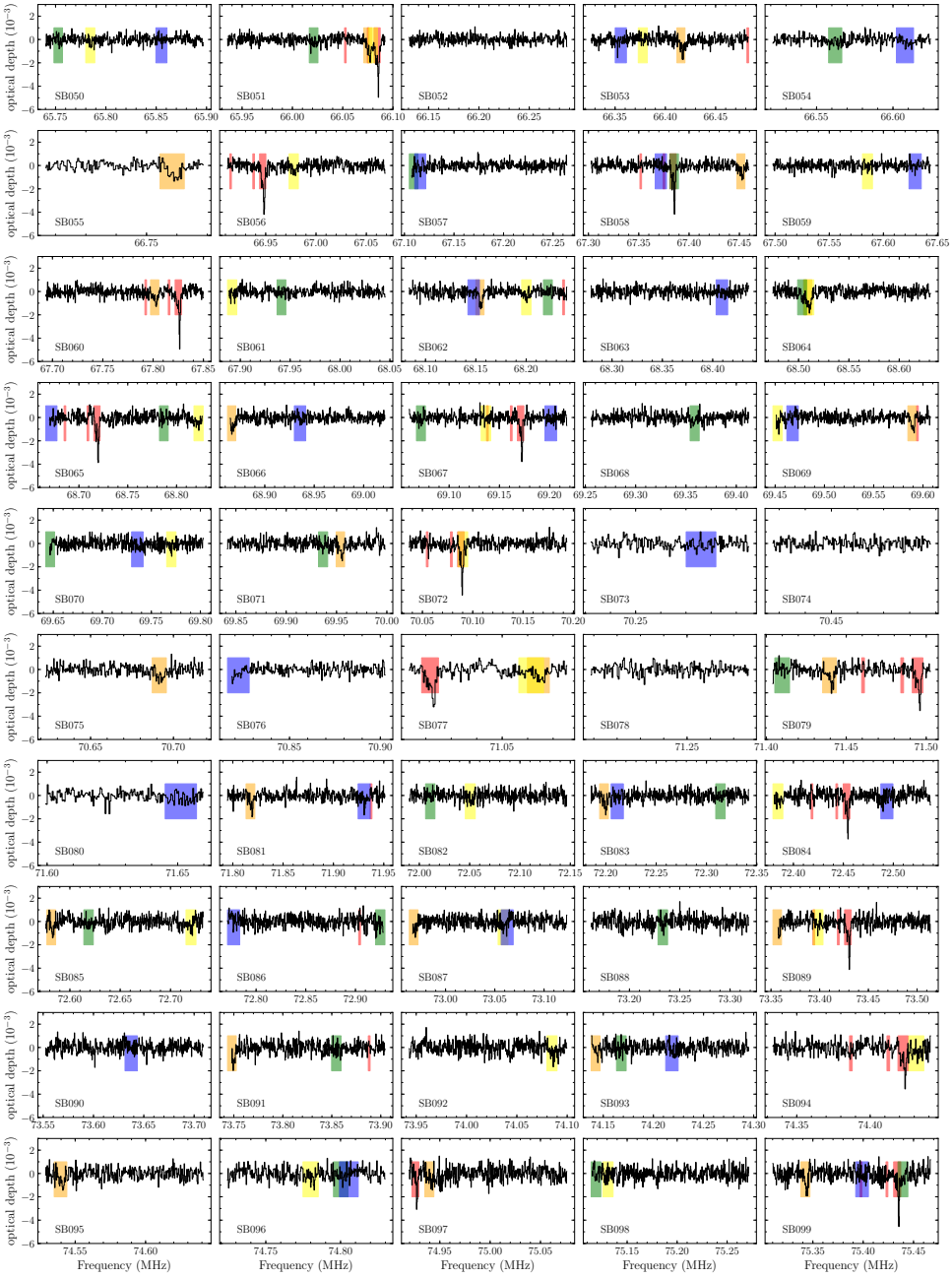


Figure 5.16: *(cont.)*

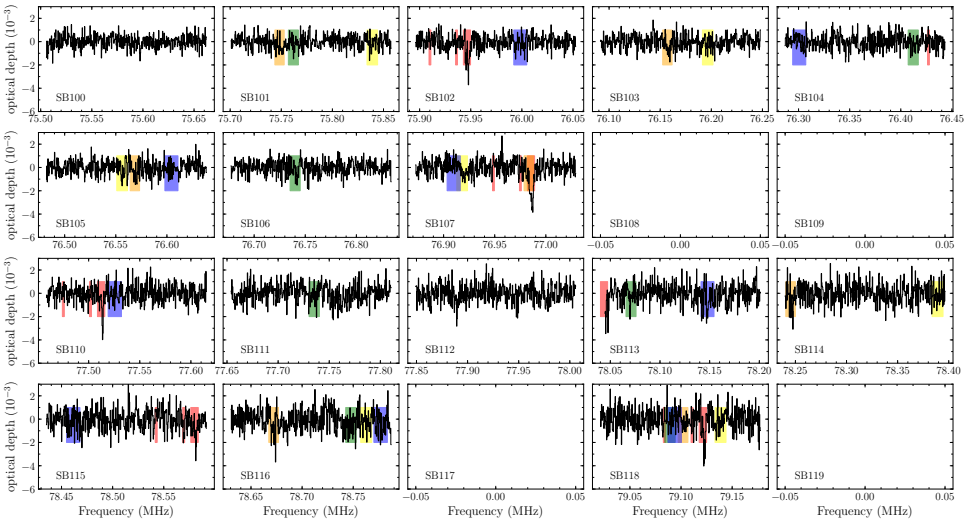


Figure 5.16: The subband spectra of Cas A, shown after continuum removal, as the optical depth as a function of frequency. The red shaded regions cover the location of α-transitions, orange regions cover β-transitions, yellow regions cover γ-transitions, green regions cover δ-transitions, and blue regions cover ϵ-transitions. We note that the widest shaded regions in red cover the spectral lines of the -47 km/s and -38 km/s Cα components; two recurring thin shaded regions in red represent the weak, narrow spectral lines of the 0 km s^{-1} Cα component and Hα at -47 km s^{-1}.

The stacking procedure used by M14 differs from our own. Instead of interpolating the spectra to a fixed velocity grid and averaging, they combine the unique velocity samples of the data and smooth the values with a Savitzky-Golav filter (Savitzky & Golay 1964). This filtering has the superior advantages that when tuned properly, the peak and width line properties are well preserved (Savitzky & Golay 1964; Bromba & Ziegler 1981; Press & Teukolsky 1990). On the other hand, it requires that data be uniformly sampled, which is not the case for our stacked RRL spectra, and since the filter-width requires precise tuning (Bromba & Ziegler 1981), it is more cumbersome for our aim of performing large-scale systematic searches.

Next, we replicated the M14 results, with the procedures outline in that paper. We computed a Doppler correction to the barycentric rest-frame and subtracted $v_{\rm Doppler} = 10.24$ km s^{-1}. We flagged five subbands with large rms. A redshift of $z = 0.00073$ was chosen by M14 because a peak in their spectral cross-correlation was found at this redshift and it maximized the signal-to-noise of the $z = 0.00070$ feature. Thus, we assumed $z = 0.00073$ and compared the expected frequency of the CRRL α-transitions with the frequencies of the subband channels. We discard all subbands that do not contain a spectral line and additionally remove subbands where the spectral line falls closer than six channels from the subband edge. Then the first three channels and the last three channels were flagged. At this point, 23 subbands remained for processing.

We then converted the spectra into velocity units. Comparing the line-blanking region of ± 50 km s^{-1} with the channel velocities, we found that all 23 subbands had at

least one unflagged channel on either side of the line-blanking region. M14 note that one subband did not have at least one channel remaining between the line-blanking region and the unflagged channels on the subband edge, and thus that subband was discarded. To identify the additional subband flagged, we identified three subbands which had a line blanking region separated from the edge by one channel: SB107, SB120, SB168. We created stacked spectra (see below for stacking procedure) in which just one of the subbands was flagged. We fit a Gaussian to the final spectrum and compared the noise in the line free region. When SB168 was excluded, the peak of the Gaussian, the center velocity of the fit, the FWHM, and the noise matched closest with the values reported in M14. Thus we additionally flagged SB168.

We blanked channels within ± 50 km s^{-1} of the line center and used the line-free channels to make a fit to the continuum. Each subband contained 31 channels to start with, after 6 total were flagged, and typically 3 were line-blanked, this resulted in typically 22 channels available to fit the continuum. The order of the polynomial fit to the continuum was not chosen by eye (among a first or second order) as done in M14; instead, a first and second order polynomial was fit, and the polynomial order which resulted in a lower χ^2 was taken. We divided all channels by the fit and subtracted one, obtaining spectral units in terms of optical depth. The optical depth values of a subband were given a weight $w = \sigma^{-1}$, where σ is the standard deviation of line-free channels.

With these 22 subbands, we analyze the noise properties. We compute the noise of each subband by finding the standard deviation of the line-free channels. We plot this as a function of subband frequency in Figure 5.17. The median value is 6.9×10^{-3}. We then collect the weighted optical depth values in all the line-free channels and plot them in a histogram (Figure 5.18). The best fit Gaussian to the distribution has a mean value of $\mu = -1.4 \times 10^{-3}$ and a width of $\sigma = 7.3 \times 10^{-3}$. This shows that the uncertainty reflected in the data, before smoothing (or averaging) is 7.3×10^{-3}.

These noise properties contrast with M14. In Figure 1 of M14, the standard deviation of the optical depth is plotted as a function of subband frequency. The median value appears to be roughly $\sigma_{\rm median} \approx 1.5 \times 10^{-2}$. However, the optical depth standard deviation of channels that have a velocity in the range of $50 {\rm km\,s^{-1}} < |v| \lesssim 150$ km s^{-1} is quoted as 5×10^{-3}, prior to any smoothing of the data. In other words, M14 reported that the channels within the narrow velocity range they selected have a standard deviation that is a factor of three times lower than the standard deviation of the full sample of data. This indicates that the noise properties within that region do not reflect the actual uncertainty of the data.

We continued with the next step of processing in which the subband channels are aligned in velocity space and smoothed with the Savitzky-Golay filter. We show the results of Savitzky-Golay smoothing with a 31 data point filter width and a first order polynomial – matching the properties which represent the main result of M14. The spectrum we reproduce is shown in Figure 5.19, overplotted on the spectrum of M14 (courtesy L. Morabito). The properties of the spectral feature we reproduced are consistent within error. The best fit Gaussian (also shown in Figure 5.19) to the feature that we reproduced has the following properties: a peak of $(-2.8 \pm 0.2) \times 10^{-3}$, a central velocity of (-10.6 ± 1.2) km s^{-1}, and a FWHM of (35.7 ± 2.8) km s^{-1}. The optical depth standard deviation in the line-free channels of our spectrum is

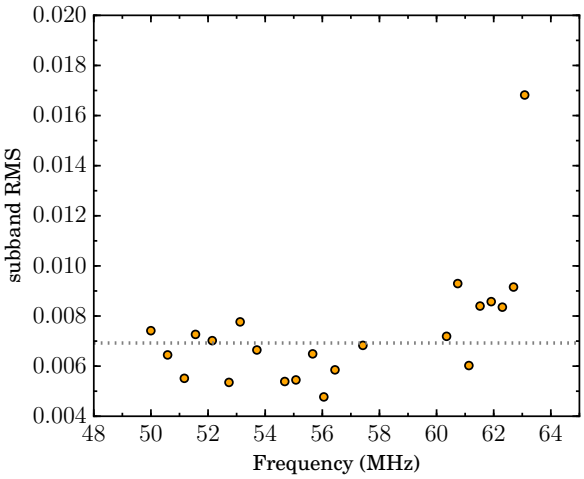


Figure 5.17: The RMS of each M 82 subband plotted against the average frequency of the subband. The gray dotted lines represents the median value of 6.9×10^{-3}. These subbands were flagged and processed following the criteria of M14. The RMS was determined from line-free channels only.

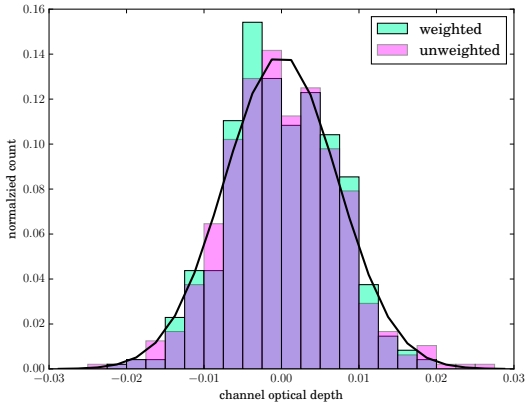


Figure 5.18: Histogram of the optical depth of line-free channels, shown before (pink) and after (green) weighting with their overlap in purple. The black solid line shows the Gaussian fit to the weighted subbands. Unweighted, the fit to the distribution has a mean of $\mu_{uw} = -1.4 \times 10^{-3}$ and a standard deviation of $\sigma_{uw} = 7.2 \times 10^{-3}$. Weighted, the best fit Gaussian has properties of $\mu_w = -1.3 \times 10^{-3}$ and $\sigma_w = 7.3 \times 10^{-3}$. These subbands of M 82 were flagged and processed following the criteria of M14.

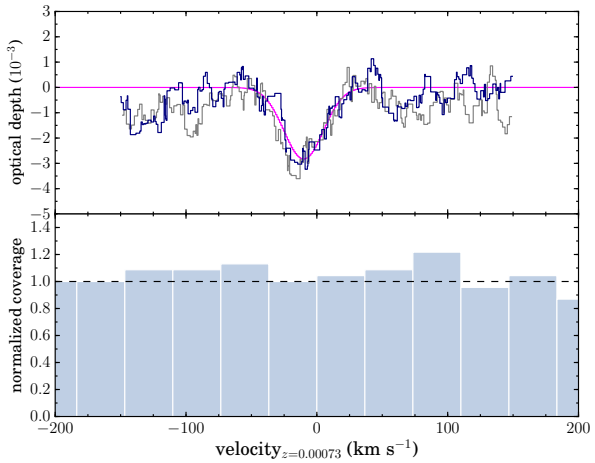


Figure 5.19: The top plot shows the optical depth as a function of velocity for the RRL spectrum of M 82 stacked at $z_{\rm M82} = 0.00073$ and smoothed with a Savitzky-Golay filter. In navy blue is the spectrum we reproduced following the criteria of M14; the magenta line shows the best fit Gaussian. The standard deviation in the line-free channels is 6.5×10^{-4}. In gray is the final M14 spectrum (courtesy L. Morabito). The bottom plot shows the coverage, or number of subband data points, within velocity bins of 36.7 km s^{-1}. The coverage count has been normalized by 23, the number of data points in the bin where the line peak falls.

$\sigma = 6.5 \times 10^{-4}$.

A property of the Savitzky-Golay filter is that the noise is inversely proportional with $\sqrt{N}$ number of points of the filter width (Savitzky & Golay 1964). Thus, we expect the noise (7.3×10^{-3}) to be smoothed to roughly 1.3×10^{-3} in our spectrum. However, this is a factor of about two greater than the standard deviation we measure (6.5×10^{-4}). Incidentally, in the M14 analysis, spectra with a noise of 1.5×10^{-2} are expected to have a noise of 2.7×10^{-3} after smoothing. This is a factor of about 10 larger than the 3.3×10^{-4} noise reported, and it is also consistent with the peak value of the RRL feature reported.

The narrow velocity range was selected by M14 because the coverage in velocity was approximately uniform (to within 20%, e.g. see Figure 5.19). Here coverage refers to the number of subband data points that fall within a fixed velocity bin. However it is possible to extend the coverage to larger velocities by including flanking subbands which do not contain a spectral line. When including these subbands, we find the coverage continues to stay within 20% (with a median value of an additional 13%) out to ± 550 km s^{-1}. We plot the spectrum which includes the additional subbands and its normalized coverage in Figure 5.20. The noise we find in line-free channels out to ± 550 km s^{-1} is 1.1×10^{-3}, which is now much closer to the expected noise of 1.3×10^{-3} for completely uniform coverage. Quantifying the effects of applying the filter on non-uniformly sampled data is beyond the scope of this paper. Given this additional uncertainty and since the coverage is 13% lower in the line region, we adopt the noise value of 1.3×10^{-3}. Thus, we find both the peak signal-to-noise and integrated optical depth equal to a 2.2σ result.

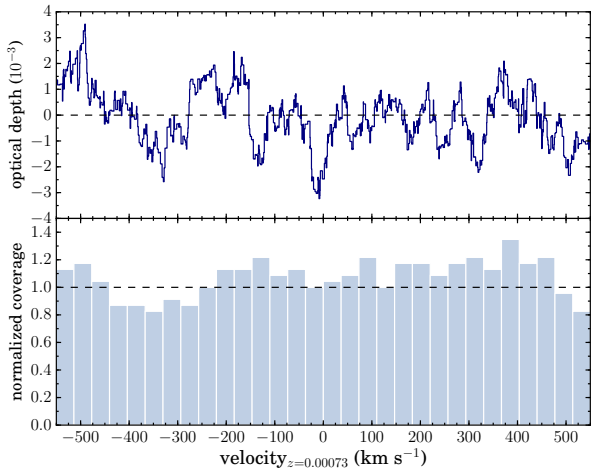


Figure 5.20: Optical depth as a function of velocity, same as in Figure 5.19, except we have extended the coverage to larger velocities by including subbands of M 82 flanking those which contain a line. The standard deviation in the line-free channels after correcting for non-uniform coverage is 1.3×10^{-3}.

As a final test, we assume the same flagging and processing criteria of M14 and apply our three step method. We make an RRL stack at each redshift (same redshift range described in Section 5.5). Instead of integrating within a region half the size of the line-blanking region ($\pm$25 km s^{-1}), we taylor this to specifically match the FWHM of the line (as it will maximize the signal-to-noise of the feature) and integrate within $\pm$18 km s^{-1}. The results across redshift are shown in Figure 5.21. The mean is -0.1 Hz and the standard deviation of the values across redshift is 8.7 Hz. This agrees with the 8.8 Hz expectation from noise of 1.3×10^{-3} integrated over 36 km s^{-1} at 56.5 MHz. The integrated optical depth at $z = 0.00073$, which is 9.0 Hz, is also consistent with noise. As the second step of our method, we apply the spectral cross correlation. We find the cross-correlation value to be 1.0σ in comparison to the distribution. We note that relative to $z = 0.00070$, we find an integrated optical depth of 11.2 Hz and a cross-correlation value that is 1.2 times the standard deviation of the distribution.

In conclusion, we find that smoothing subband spectra with a Savitzky-Golay filter and implementing the flagging procedure of M14 produces a stacked CRRL feature centered at $z = 0.00070$ with a significance of 2.2σ. Comparing the integrated optical depth at this redshift with the value obtained across a range of redshifts produces a value that is 1.0σ.

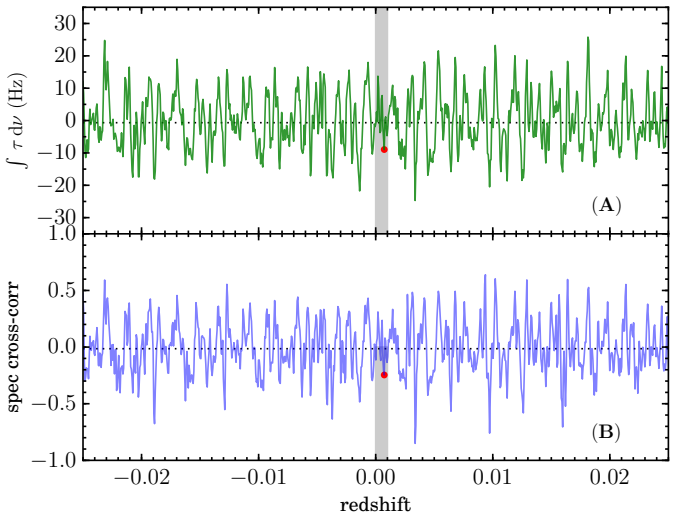


Figure 5.21: Same as in Figure 5.14 where we have implemented a stacking and spectral cross-correlation across a range of redshifts, except in this case the spectra have been combined and smoothed via a Savitzky-Golay filter. We mimicked the procedure M14 used to flag, process, and smooth the data of M 82 as best as possible.

Bibliography

Abdullah, A., & Tielens, A. G. G. M. 2020, A&A, 639, 110

Ackermann, M., Ajello, M., Allafort, A., et al. 2011, Science, 334, 1103

Adamo, A., & Bastian, N. 2016, in Orig Stellar Clust, ed. S. Stahler (Springer), 27

Adamo, A., Kruijssen, J. M. D., Bastian, N., Silva-Villa, E., & Ryon, J. 2015, MNRAS, 452, 246

Adamo, A., Östlin, G., Zackrisson, E., et al. 2011, MNRAS, 415, 2388

Albacete Colombo, J. F., Drake, J. J., Flaccomio, E., et al. 2018, ApJSS, arXiv:1806

Alves, M. I., Calabretta, M., Davies, R. D., et al. 2015, MNRAS, 450, 2025

Anantharamaiah, K. R. 1985a, JApA, 6, 177

—. 1985b, JApA, 6, 203

—. 1986, JApA, 7, 131

Anantharamaiah, K. R., & Goss, W. M. 1990, in Radio Recomb Lines 25 Years Investig IAU Colloq 125, ed. M. A. Gordon & R. L. Sorochenko (Puschino: Dordrecht: Kluwer), 267–275

Anantharamaiah, K. R., & Goss, W. M. 1996, ApJL, 466, 13

Anantharamaiah, K. R., Viallefond, F., Mohan, N. R., Goss, W. M., & Zhao, J. H. 2000, ApJ, 537, 613

Anantharamaiah, K. R., Zhao, J.-H., Goss, W. M., & Viallefond, F. 1993, ApJ, 419, 585

Anantharamaiah, K. R., Zhao, J. H., Goss, W. M., & Viallefond, F. 1995, J Astrophys Astron Suppl, 16, 273

André, P., Di Francesco, J., Ward-Thompson, D., et al. 2014, in Protostars Planets VI, ed. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning (University of Arizona Press), 27–51

Araya, E., Baan, W. A., & Hofner, P. 2004, ApJSS, 154, 541

Arias, M., Vink, J., Zhou, P., et al. 2019, AJ, 158, 253

Arzoumanian, D., André, P., Didelon, P., et al. 2011, A&A, 529, L6

Asgekar, A., Oonk, J. B. R., Yatawatta, S., et al. 2013, A&A, 551, L11

Bally, J. 2016, ARAA, 54, 491

Balser, D. 2006, AJ, 132, 2326

Balser, D. S., Wenger, T. V., Goss, W. M., Johnson, K. E., & Kepley, A. A. 2017, ApJ, 844, 73

Beerer, I. M., Koenig, X. P., Hora, J. L., et al. 2010, ApJ, 720, 679

Bell, M. B., & Seaquist, E. R. 1977, A&A, 56, 461

—. 1978, ApJ, 223, 378

—. 1980, ApJ, 238, 818

Bell, M. B., Seaquist, E. R., Mebold, U., Reif, K., & Shaver, P. A. 1984, A&A, 130, 1

Bendo, G. J., Beswick, R. J., D'Cruze, M. J., et al. 2015, MNRAS, 450, L80
Bendo, G. J., Henkel, C., D'Cruze, M. J., et al. 2016, MNRAS, 463, 252
Bendo, G. J., Miura, R. E., Espada, D., et al. 2017, MNRAS, 472, 1239
Bennett, A. S. 1962, MNRAS, 68, 163
Bennett, C. L., Fixsen, D. J., Hinshaw, G., et al. 1994, ApJ, 434, 587
Berkhuijsen, E. M. 1972, A&AS, 5, 263
Berkhuijsen, E. M., Mitra, D., & Müller, P. 2006, AN, 327, 82
Berlanas, S. R., Herrero, A., Comerón, F., et al. 2018, A&A, 612, 50
Berlanas, S. R., Wright, N. J., Herrero, A., Drew, J. E., & Lennon, D. J. 2019, MNRAS, 484, 1838
Berlanas, S. R., Herrero, A., Comerón, F., et al. 2020, A&A
Best, P. N., Bailer, D. M., Longair, M. S., & Riley, J. M. 1995, MNRAS, 275, 1171
Bochkarev, N. G., & Sitnik, T. G. 1985, Astrophys Space Sci, 108, 237
Bohnenstengel, H. D., & Wendker, H. J. 1976, A&A, 52, 23
Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013a, ARAA, 51, 207
Bolatto, A. D., Warren, S. R., Leroy, A. K., et al. 2013b, Natur, 499, 450
Bressert, E., Ginsburg, A., Bally, J., et al. 2012, ApJL, 758, L28
Brocklehurst, M., & Seaton, M. J. 1972, MNRAS, 157, 179
Bromba, M. U. A., & Ziegler, H. 1981, Anal Chem, 53, 1583
Calzetti, D., Kennicutt, R. C., Engelbracht, C. W., et al. 2007, ApJ, 666, 870
Cardelli, J. A., Meyer, D. M., Jura, M., & Savage, B. D. 1996, ApJ, 467, 334
Carilli, C. L., & Walter, F. 2013, ARAA, 51, 105
Chaisson, E. J., & Rodriguez, L. F. 1977, ApJ, 214, L111
Chandar, R., Fall, S. M., Whitmore, B. C., & Mulia, A. J. 2017, ApJ, 849, 128
Chen, Z.-F., & Pan, D.-S. 2017, ApJ, 848, 79
Chou, R. C. Y., Peck, A. B., Lim, J., et al. 2007, ApJ, 670, 116
Chowdhury, A., & Chengalur, J. N. 2019, MNRAS, 486, 42
Churchwell, E., & Shaver, P. A. 1979, A&A, 77, 316
Churchwell, E., Povich, M. S., Allen, D., et al. 2006, ApJ, 649, 759
Comerón, F., Djupvik, A. A., Schneider, N., & Pasquali, A. 2016, A&A, 586, 46
—. 2020, A&A
Comerón, F., & Pasquali, A. 2012, A&A, 543, 1
Comeron, F., & Torra, J. 1999, A&A, 349, 605
Comerón, F., & Torra, J. 2001, A&A, 375, 539
Condon, J. J. 1992, ARAA, 30, 575
Condon, J. J., & Ransom, S. M. 2016, Essential Radio Astronomy (Princeton University Press), 361
Cong, H. I. L. 1977, PhD thesis, Columbia University
Cotton, W. D. 2017, PASP, 129, 094501
Dale, J. E., Ercolano, B., Bonnell, I. A., et al. 2015, MNRAS, 451, 987
Dale, J. E., Ngoumou, J., Ercolano, B., & Bonnell, I. A. 2014, MNRAS, 442, 694
de Gasperin, F., Mevius, M., Rafferty, D. A., Intema, H. T., & Fallows, R. A. 2018, A&A, 615, 179
de Gasperin, F., Dijkema, T. J., Drabent, A., et al. 2019, A&A, 622, A5
de Oliveira-Costa, A., Tegmark, M., Gaensler, B. M., et al. 2008, MNRAS, 388, 247
de Pree, C. G., Gaume, R. A., Goss, W. M., & Claussen, M. J. 1996, ApJ, 464, 788

de Wit, W. J., Testi, L., Palla, F., & Zinnecker, H. 2005, A&A, 437, 247
Deb, S., Kothes, R., & Rosolowsky, E. 2018, MNRAS, 481, 1862
Decarli, R., Walter, F., Aravena, M., et al. 2016, ApJ, 833, 70
Dickey, J. M., & Lockman, F. J. 1990, ARAA, 28, 215
Dickey, J. M., Strasser, S., Gaensler, B. M., et al. 2009, ApJ, 693, 1250
Downes, D., & Rinehart, R. 1966, ApJ, 144, 937
Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium (Princeton University Press), 510
Eastwood, M. W., Anderson, M. M., Monroe, R. M., et al. 2018, AJ, 156, 32
Edge, D. O., Shakeshaft, J. R., McAdam, W. B., Baldwin, J. E., & Archer, S. 1959, MNRAS, 68, 37
Eisner, B. A., Ott, J., Meier, D. S., & Cannon, J. M. 2019, ApJ, 882, 95
Elmegreen, B. G., & Lada, C. J. 1977, ApJ, 214, 725
Elvis, M., Wilkes, B. J., McDowell, J. C., et al. 1994, ApJSS, 95, 1
Emig, K. L., Salas, P., de Gasperin, F., et al. 2020a, A&A, 634, 138
—. 2019, A&A, 622, A7
Emig, K. L., Bolatto, A. D., Leroy, A. K., et al. 2020b, ApJ, 903, 50
Emonts, B., Raba, R., Moellenbrock, G., et al. 2019, in ADASS XXIX, ASP Conf Ser
Erickson, W. C., McConnell, D., & Anantharamaiah, K. R. 1995, ApJ, 454, 125
Ferriere, K. M. 2001, RvMP, 73, 1031
Field, G. B., Goldsmith, D. W., & Habing, H. J. 1969, ApJ, 155, L149
Gandhi, P., Isobe, N., Birkinshaw, M., et al. 2011, Publ Astron Soc Japan, 63, 505
Gibson, S. J., Taylor, A. R., Higgs, L. A., & Dewdney, P. E. 2000, ApJ, 540, 851
Gies, D. R. 1987, ApJSS, 64, 545
Ginsburg, A., & Kruijssen, J. M. D. 2018, ApJL, 864, L17
Ginsburg, A., Goss, W. M., Goddi, C., et al. 2016, A&A, 595, 27
Ginsburg, A., Bally, J., Barnes, A., et al. 2018, ApJ, 853, 171
Ginsburg, A., Anderson, L. D., Dicker, S., et al. 2020, ApJSS, 248, 24
Goddard, Q. E., Bastian, N., & Kennicutt, R. C. 2010, MNRAS, 405, 857
Goldberg, L. 1966, ApJ, 144, 1225
Goldsmith, P. F., Yildiz, U. A., Langer, W. D., & Pineda, J. L. 2015, ApJ, 814, 133
Gordon, M. A., & Sorochenko, R. L. 2002, Radio Recombination Lines. Their Physics and Astronomical Applications (Dordrecht: Kluwer Academic Publishers), 360
Gorski, M., Ott, J., Rand, R., et al. 2017, ApJ, 842, 124
Greenhill, L. J., Moran, J. M., & Herrnstein, J. R. 1997, ApJL, 481, L23
Grenier, I. A., Casandjian, J.-M., & Terrier, R. 2005, Science, 307, 1292
Groppi, C., Walker, C., Kulesa, C., et al. 2009, in 20th Int Symp Sp Terahertz Technol, ed. E. Bryerton, A. Kerr, & A. Lichtenberger, Charlottesville, VA, USA, 90
Haffner, L. M., Dettmar, R. J., Beckman, J. E., et al. 2009, RvMP, 81, 969
Haid, S., Walch, S., Seifried, D., et al. 2018, MNRAS, 478, 4799
Hainich, R., Ramachandran, V., Shenar, T., et al. 2019, A&A, 621, 85
Hamann, F., Kanekar, N., Prochaska, J. X., et al. 2011, MNRAS, 410, 1957
Hamann, W.-R., & Gräfener, G. 2004, A&A, 427, 697
Harper-Clark, E., & Murray, N. 2009, ApJ, 693, 1696
Harris, S. 1973, MNRAS, 162, 5
Haslam, C. G. T., Salter, C. J., Stoffel, H., & Wilson, W. E. 1982, A&AS, 47, 1

Haslam, C. G. T., Wilson, W. E., Graham, D. A., & Hunt, G. C. 1974, A&AS, 13, 359
Heckman, T. M., Armus, L., & Miley, G. K. 1990, ApJSS, 74, 833
Heiles, C. 1994, ApJ, 436, 720
—. 1997, ApJ, 481, 193
Heiles, C., Koo, B.-C., Levenson, N. A., & Reach, W. T. 1996a, ApJ, 462, 326
Heiles, C., Reach, W. T., & Koo, B.-C. 1996b, ApJ, 466, 191
Heiles, C., & Troland, T. H. 2003, ApJ, 586, 1067
Henkel, C., Muehle, S., Bendo, G., et al. 2018, A&A, 615, 155
Herrera-Camus, R., Bolatto, A., Wolfire, M., et al. 2017, ApJ, 835, 201
Hollenbach, D. J., & Tielens, A. G. G. M. 1999, RvMP, 71, 173
Holt, J., Tadhunter, C. N., Morganti, R., & Emonts, B. H. C. 2011, MNRAS, 410, 1527
Hopkins, P. F., Kereš, D., Oñorbe, J., et al. 2014, MNRAS, 445, 581
Hopkins, P. F., Richards, G. T., & Hernquist, L. 2007, ApJ, 654, 731
Hopkins, P. F., Wetzel, A., Kereš, D., et al. 2018, MNRAS, 480, 800
Hoyle, F., & Ellis, G. R. A. 1963, Aust J Phys, 16, 1
Hunter, J. D. 2007, CSE, 9, 90
Intema, H. T., van der Tol, S., Cotton, W. D., et al. 2009, A&A, 501, 1185
Ishwara-Chandra, C. H., Dwarakanath, K. S., & Anantharamaiah, K. R. 2003, JApA, 24, 37
Izumi, T., Nakanishi, K., Imanishi, M., & Kohno, K. 2016, MNRAS, 459, 3629
Jacobs, B. A., Rizzi, L., Tully, R. B., et al. 2009, AJ, 138, 332
Jenkins, E. B., & Tripp, T. M. 2001, ApJSS, 137, 297
Jenkins, E. B., Tripp, T. M., Jenkins, E. B., & Tripp, T. M. 2011, ApJ, 734, 65
Johnson, L. C., Seth, A. C., Dalcanton, J. J., et al. 2016, ApJ, 827, 33
Kalberla, P. M., & Kerp, J. 2009, ARAA, 47, 27
Kanekar, N., & Briggs, F. H. 2004, New Astron Rev, 48, 1259
Kantharia, N. G., & Anantharamaiah, K. R. 2001, JApA, 22, 51
Kantharia, N. G., Anantharamaiah, K. R., & Payne, H. E. 1998, ApJ, 506, 758
Karachentsev, I. D., Tully, R. B., Dolphin, A., et al. 2007, AJ, 133, 504
Katz-Stone, D. M., & Rudnick, L. 1997, ApJ, 479, 258
Kennicutt, R. C. J. 1998, ARAA, 36, 189
Kepley, A. A., Chomiuk, L., Johnson, K. E., et al. 2011, ApJL, 739, L24
Kerr, F. J., & Lynden-Bell, D. 1986, MNRAS, 221, 1023
Kessler, M. F., Steinz, J. A., Anderegg, M. E., et al. 1996, A&A, 315, L27
Kewley, L. J., Nicholls, D. C., & Sutherland, R. S. 2019, ARAA, 57, 511
Kim, C.-G., & Ostriker, E. C. 2015, ApJ, 802, 99
Kim, C.-G., Ostriker, E. C., Kim, W.-T., et al. 2013, ApJ, 776, 1
Kim, J.-G., Kim, W.-T., & Ostriker, E. C. 2018, ApJ, 859, 68
Klein, U., Lisenfeld, U., & Verley, S. 2018a, A&A, 611, 55
—. 2018b, A&A, 611, 55
Konovalenko, A. A., & Sodin, L. G. 1980, Natur, 283, 360
Krause, M. G. H., Charbonnel, C., Bastian, N., & Diehl, R. 2016, A&A, 587, 53
Krause, M. G. H., Offner, S. S. R., Charbonnel, C., et al. 2020, SSRv, 216, 64
Krieger, N., Bolatto, A. D., Walter, F., et al. 2019, ApJ, 881, 43
Krieger, N., Bolatto, A. D., Leroy, A. K., et al. 2020, ApJ, 897, 176
Kroupa, P. 2001, MNRAS, 322, 231
Kruijssen, J. M. D. 2012, MNRAS, 426, 3008

Krumholz, M. R. 2014, Phys Rep, 539, 49
Krumholz, M. R., McKee, C. F., & Bland-Hawthorn, J. 2019, ARAA, 57, 227
Kulkarni, S. R., & Heiles, C. 1988, in Galact extragalactic radio Astron, 2nd edn., ed. G. A. Verschuur & K. I. Kellermann (Berlin and New York: Springer-Verlag), 95–153
Lada, C. J., & Lada, E. A. 2003, ARAA, 41, 57
Landecker, T. L. 1984, AJ, 89, 95
Landecker, T. L., & Wielebinski, R. 1970, AuJPA, 16, 1
Lang, C. C., Goss, W. M., & Morris, M. 2001, AJ, 121, 2681
Le Duigou, J.-M., & Knodlseder, J. 2002, A&A, 392, 869
Leitherer, C., Schaerer, D., Goldader, J. D., et al. 1999, ApJSS, 123, 3
Lenc, E., & Tingay, S. J. 2009, AJ, 137, 537
Leroy, A. K., Bolatto, A. D., Ostriker, E. C., et al. 2018, ApJ, 896, 126
Levy, R. C., Bolatto, A. D., Sánchez, S. F., et al. 2019, ApJ, 882, 84
Levy, R. C., Bolatto, A. D., Leroy, A. K., et al. 2021, ApJ arxix eprint, 2011.05334, 30
Li, H., Vogelsberger, M., Marinacci, F., & Gnedin, O. Y. 2019, MNRAS, 487, 364
Linden, S. T., Murphy, E. J., Dong, D., et al. 2020, ApJSS, 248, 25
Linden, S. T., Evans, A. S., Rich, J., et al. 2017, ApJ, 843, 91
Lockman, F. J. 1976, ApJ, 209, 429
—. 1989, ApJSS, 71, 469
Longmore, S. N., Kruijssen, J. M. D., Bastian, N., et al. 2014, in Protostars Planets VI (University of Arizona Press), 291
Lopez-Sanchez, A. R., Lagos, C. D. P., Young, T., & Jerjen, H. 2018, MNRAS
Lucas, R., & Liszt, H. S. 2000, A&A, 358, 1069
Luisi, M., Anderson, L. D., Bania, T. M., et al. 2018, PASP, 130, 084101
Madau, P., & Dickinson, M. 2014, ARAA, 52, 415
Manti, S., Gallerani, S., Ferrara, A., et al. 2016, MNRAS, 456, 98
Marchesi, S., Ajello, M., Marcotulli, L., et al. 2018, ApJ, 854, 49
Marconi, A., Oliva, E., Van Der Werf, P. P., et al. 2000, A&A, 357, 24
Marston, A. P., Reach, W. T., Noriega-Crespo, A., et al. 2004, ApJSS, 154, 333
Matzner, C. D. 2002, ApJ, 566, 302
McCray, R., & Snow, T. P., J. 1979, ARAA, 17, 213
McDonald, A. R., Muxlow, T. W. B., Wills, K. A., Pedlar, A., & Beswick, R. J. 2002, MNRAS, 334, 912
McKean, J. P., Godfrey, L. E. H., Vegetti, S., et al. 2016, MNRAS, 463, 3143
Mckee, C. F., & Ostriker, J. P. 1977, ApJ, 218, 148
McKee, C. F., & Williams, J. P. 1997, ApJ, 476, 144
McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, in Astron Data Anal Softw Syst XVI ASP Conf Ser, Vol. 376 (Astronomical Society of the Pacific), 127
McQuinn, M. 2016, ARAA, 54, 313
Mebold, U., Shaver, P. A., Bell, M. B., & Seaquist, E. R. 1980, A&A, 82, 272
Mechev, A. P., Oonk, J. B. R., Danezi, A., et al. 2017, in Proc Int Symp Grids Clouds 2017, Taipei
Mechev, A. P., Plaat, A., Oonk, J. B. R., Intema, H. T., & Röttgering, H. J. A. 2018, Astron Comput, 2, 117
Menten, K. M., Reid, M. J., Forbrich, J., & Brunthaler, A. 2007, A&A, 474, 515
Menzel, D. H. 1968, Natur, 218, 756

Mezger, P. O. 1978, A&A, 70, 565
Michiyama, T., Iono, D., Nakanishi, K., et al. 2020, ApJ, 895, 85
Miller, R. G. 1974, Biometrika, 61, 1
Mills, E. A. C., Gorski, M., Bolatto, A. D., et al. 2020, ApJ, submitted
Mingozzi, M., Cresci, G., Venturi, G., et al. 2019, A&A, 622, 146
Mohan, N., & Rafferty, D. 2015, PyBDSF: Python Blob Detection and Source Finder
Mohan, N. R., Anantharamaiah, K. R., & Goss, W. M. 2001, ApJ, 557, 659
—. 2002, ApJ, 574, 701
Mohan, N. R., Goss, W. M., & Anantharamaiah, K. R. 2005, A&A, 432, 1
Mok, A., Chandar, R., & Fall, S. M. 2020, ApJ, 893, 135
Moorwood, A. F. M., van der Werf, P. P., Kotilainen, J. K., Marconi, A., & Oliva, E. 1996, A&A, 308, L1
Morabito, L. K., Oonk, J. B. R., Salgado, F., et al. 2014, ApJL, 795, L33
Morganti, R., & Oosterloo, T. 2018, A&ARv, 26, 4
Morganti, R., Tadhunter, C. N., & Oosterloo, T. A. 2005, A&A, 444, L9
Motte, F., Bontemps, S., Schilke, P., et al. 2007, A&A, 476, 1243
Murphy, E. J., Dong, D., Momjian, E., et al. 2018, ApJSS, 234, 24
Murray, N. W., & Rahman, M. 2010, ApJ, 709, 424
Neufeld, D. A., & Wolfire, M. G. 2017, ApJ, 845, 163
Niklas, S., Klein, U., & Wielebinski, R. 1997, A&A, 322, 19
Ochsendorf, B. B., Brown, A. G. A., Bally, J., & Tielens, A. G. G. M. 2015, ApJ, 808, 111
Ochsendorf, B. B., Cox, N. L. J., Krijt, S., et al. 2014, A&A, 563, 65
Odenwald, S. F., Campbell, M. F., Shivanandan, K., et al. 1990, AJ, 99, 288
Oey, M. S., Herrera, C. N., Silich, S., et al. 2017, ApJ, 849, L1
Oey, M. S., Meurer, G. R., Yelda, S., et al. 2007, ApJ, 661, 801
Offringa, A. R., & Smirnov, O. 2017, MNRAS, 471, 301
Offringa, A. R., van de Gronde, J. J., & Roerdink, J. B. T. M. 2012, A&A, 539, 95
Offringa, A. R., de Bruyn, A. G., Zaroubi, S., et al. 2013, A&A, 549, A11
Offringa, A. R., McKinley, B., Hurley-Walker, N., et al. 2014, MNRAS, 444, 606
Olivier, G. M., Lopez, L. A., Rosen, A. L., et al. 2020, ApJ, eprint, eprint arXiv:2009.10079
Oonk, J. B. R., Alexander, E. L., Broderick, J. W., Sokolowski, M., & Wayth, R. 2019, MNRAS, 487, 4737
Oonk, J. B. R., Morabito, L. K., Salgado, F., et al. 2015, in Proc Adv Astrophys with Sq Km Array 2014, 139
Oonk, J. B. R., van Weeren, R. J., Salas, P., et al. 2017, MNRAS, 465, 1066
Oonk, J. B. R., van Weeren, R. J., Salgado, F., et al. 2014, MNRAS, 437, 3506
Oosterloo, T., Oonk, J. B. R., Morganti, R., et al. 2017, A&A, 608
Ossenkopf, V., & Henning, T. 1994, A&A, 291, 943
Oster, L. 1961, RvMP, 33, 525
Ostriker, J. 1964, ApJ, 140, 1529
Pabst, C., Higgins, R., Goicoechea, J. R., et al. 2019, Natur, 565, 618
Pabst, C. H. M., Goicoechea, J. R., Teyssier, D., et al. 2020, A&A, 639, 2
Panwar, N., Sharma, S., Ojha, D. K., et al. 2020, ApJ, eprint, arXiv:2010.11065
Patra, N., Subrahmanyan, R., Sethi, S., Shankar, N. U., & Raghunathan, A. 2015, ApJ, 801, 138
Payne, H. E., Anantharamaiah, K. R., & Erickson, W. C. 1989, ApJ, 341, 890

Pedlar, A., Davies, R. D., Hart, L., & Shaver, P. A. 1978, MNRAS, 182, 473
Pellegrini, E. W., Baldwin, J. A., & Ferland, G. J. 2011, ApJ, 738, 34
Perez, F., & Ganger, B. E. 2007, CSE, 9, 21
Peters, W. M., Lazio, T. J. W., Clarke, T. E., Erickson, W. C., & Kassim, N. E. 2011, A&A, 525, 128
Phookun, B., Anantharamaiah, K. R., & Goss, W. M. 1998, MNRAS, 295, 156
Piddington, J. H., & Minnett, H. C. 1952, Aust J Sci Res, 5, 17
Piepenbrink, A., & Wendker, H. J. 1988, A&A, 191, 313
Pineda, J. L., Horiuchi, S., Anderson, L. D., et al. 2019, ApJ, 866, 1
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2011, A&A, 536, 19
Portegies Zwart, S. F., McMillan, S. L. W., & Gieles, M. 2010, ARAA, 48, 431
Press, W. H., & Teukolsky, S. A. 1990, Comput Phys, 4, 669
Price, S. D., Egan, M. P., Carey, S. J., Mizuno, D. R., & Kuchar, T. A. 2001, AJ, 121, 2819
Puxley, P. J., Brand, P. W. J. L., Moore, T. J. T., Mountain, C. M., & Nakai, N. 1991, MNRAS, 248, 585
Puxley, P. J., Brand, P. W. J. L., Moore, T. J. T., et al. 1989, ApJ, 345, 163
Puxley, P. J., Mountain, C. M., Brand, P. W. J. L., Moore, T. J. T., & Nakai, N. 1997, ApJ, 485, 143
Reich, W. 1982, A&AS, 48, 219
Reipurth, B., & Schneider, N. 2008, in Handb Star Form Reg Vol I North Sky, Vol. 4 (ASP Monograph Publications), 36
Rekola, R., Richer, M. G., McCall, M. L., et al. 2005, MNRAS, 361, 330
Remazeilles, M., Dickinson, C., Banday, A. J., et al. 2015, MNRAS, 451, 4311
Reynolds, R. J. 1984, ApJ, 282, 191
Rico-Villas, F., Martín-Pintado, J., González-Alfonso, E., Martín, S., & Rivilla, V. M. 2020, MNRAS, 491, 4573
Robitaille, T., & Bressert, E. 2012, APLpy: Astronomical Plotting Library in Python
Rodriguez-Rico, C., Goss, W. M., Zhao, J.-H., Gomez, Y., & Anantharamaiah, K. R. 2006, ApJ, 644, 914
Rodriguez-Rico, C. A., Goss, W. M., Turner, J. L., & Gomez, Y. 2007, ApJ, 670, 295
Rodriguez-Rico, C. A., Goss, W. M., Viallefond, F., et al. 2005, ApJ, 633, 198
Rodriguez-Rico, C. A., Viallefond, F., Zhao, J. H., Goss, W. M., & Anantharamaiah, K. R. 2004, ApJ, 616, 783
Roelfsema, P. R., & Goss, W. M. 1992, A&ARv, 4, 161
Roshi, D. A., & Anantharamaiah, K. R. 1997, MNRAS, 292, 63
—. 2000, ApJ, 535, 231
—. 2001, JApA, 22, 81
Roshi, D. A., & Kantharia, N. G. 2011, MNRAS, 414, 519
Roshi, D. A., Kantharia, N. G., & Anantharamaiah, K. R. 2002, A&A, 391, 1097
Rossa, J., & Dettmar, R.-J. 2003a, A&A, 406, 493
—. 2003b, A&A, 406, 505
Roy, A. L., Goss, W. M., & Anantharamaiah, K. R. 2008, A&A, 483, 79
Roy, A. L., Goss, W. M., Mohan, N. R., & Anantharamaiah, K. R. 2005, A&A, 435, 831
Roy, A. L., Oosterloo, T., Goss, W. M., & Anantharamaiah, K. R. 2010, A&A, 517, A82
Rubin, R. H. 1968, ApJ, 154, 391
Rygl, K. L. J., Brunthaler, A., Sanna, A., et al. 2012, A&A, 539, 79

Ryon, J. E., Adamo, A., Bastian, N., et al. 2014, AJ, 148, 33
Ryon, J. E., Gallagher, J. S., Smith, L. J., et al. 2017, ApJ, 841, 92
Salas, P., Morabito, L., Salgado, F., Oonk, J. B. R., & Tielens, A. G. G. M. 2016, CRRLpy: First Pre-release
Salas, P., Oonk, J. B. R., Emig, K. L., et al. 2019, A&A, 626, 70
Salas, P., Oonk, J. B. R., van Weeren, R. J., et al. 2018, MNRAS, 2511, 2496
—. 2017, MNRAS, 467, 2274
Salgado, F., Morabito, L. K., Oonk, J. B. R., et al. 2017a, ApJ, 837, 141
—. 2017b, ApJ, 837, 142
Salgado, F., Berne, O., Adams, J. D., et al. 2012, ApJL, 749, L21
Salgado Cambiazo, F. J. 2015, PhD thesis, Leiden University
Sander, A., Shenar, T., Hainich, R., et al. 2015, A&A, 577, 13
Santoro, F., Rose, M., Morganti, R., et al. 2018, eprint arXiv:1806.09461
Sault, R. J. 1994, A&AS, 107, 55
Savitzky, A., & Golay, M. J. E. 1964, Anal Chem, 36, 1627
Schneider, N., Bontemps, S., Simon, R., et al. 2006, A&A, 458, 855
—. 2011, A&A, 529, 1
Schneider, N., Bontemps, S., Motte, F., et al. 2016, A&A, 591, A40
Seaquist, E. R., & Bell, M. B. 1977, A&A, 60, L1
Seaquist, E. R., Bell, M. B., & Bignell, R. C. 1985, ApJ, 294, 546
Seaquist, E. R., Carlstrom, J. E., Bryant, P. M., & Bell, M. B. 1996, ApJ, 465, 691
Seaquist, E. R., Kerton, C. R., & Bell, M. B. 1994, ApJ, 429, 612
Shaver, P. A. 1975a, A&A, 43, 465
—. 1975b, Pramana, 5, 1
—. 1976, A&A, 49, 1
—. 1978, A&A, 68, 97
Shaver, P. A., Churchwell, E., & Rots, A. H. 1977, A&A, 55, 435
Shaver, P. A., Churchwell, E., & Walmsley, C. M. 1978, A&A, 64, 1
Shaver, P. A., McGeem, R. X., Newton, M. L., et al. 1983, MNRAS, 204, 53
Shimwell, T. W., Röttgering, H. J. A., Best, P. N., et al. 2017, A&A, 598, 102
Shimwell, T. W., Tasse, C., Hardcastle, M. J., et al. 2019, A&A, 622, A1
Smith, H. E., & Spinrad, H. 1980, ApJ, 236, 419
Sofia, U. J., Lauroesch, J. T., Meyer, D. M., & Cartledge, S. I. B. 2004, ApJ, 605, 272
Sorochenko, R. L., & Smirnov, G. T. 2010, Astron Reports, 54, 776
Sousbie, T. 2011, MNRAS, 414, 350
Spencer, R. E., Schilizzi, R. T., Fanti, C., et al. 1991, MNRAS, 250, 225
Spinrad, H., Marr, J., Aguilar, L., & Djorgovski, S. 1985, PASP, 97, 932
Spoon, H. W. W., Koornneef, J., Moorwood, A. F. M., Lutz, D., & Tielens, A. G. G. M. 2000, A&A, 357, 898
Stanghellini, L., Magrini, L., & Casasola, V. 2015, ApJ, 812, 39
Stanimirovic, S. 2002, in Single-Dish Radio Astron Tech Appl, ed. S. Stanimirovic, D. Altschuler, P. Goldsmith, & C. Salter (Astronomical Society of the Pacific, vol. 278), 375–396
Stepkin, S. V., Konovalenko, A. A., Kantharia, N. G., & Udaya Shankar, N. 2007, MNRAS, 374, 852
Stockton, A., & Ridgway, S. E. 2001, ApJ, 554, 1012

Storey, P. J., & Hummer, D. G. 1995, MNRAS, 272, 41
Suri, S. T., Sanchez-Monge, A., Schilke, P., et al. 2019, A&A, 623, 142
Taylor, A. R., Gibson, S. J., Peracaula, M., et al. 2003, AJ, 125, 3145
Tenorio-Tagle, G. 1979, A&A, 71, 59
Terlevich, R., & Melnick, J. 1981, MNRAS, 195, 839
The Astropy Collaboration. 2018, AJ, 156, 123
Tielens, A. G. G. M. 2005, The Physics and Chemistry of the Interstellar Medium (Cambridge University Press), 495
Tingay, S. J., Goeke, R., Bowman, J. D., et al. 2013, Publ Astron Soc Aust, 30, e007
Todt, H., Sander, A., Hainich, R., et al. 2015, A&A, 579, 75
Tsai, C.-W., Turner, J. L., Beck, S. C., Meier, D. S., & Ho, P. T. P. 2009, AJ, 137, 4655
Tumlinson, J., Peeples, M. S., & Werk, J. K. 2017, ARAA, 55, 389
Tumlinson, J., Thom, C., Werk, J. K., et al. 2013, ApJ, 777, 59
Turner, J. L., Consiglio, S. M., Beck, S. C., et al. 2017, ApJ, 846, 73
Turtle, A. J., & Baldwin, J. E. 1962, MNRAS, 124, 459
Uyaniker, B., Fürst, E., Reich, W., Aschenbach, B., & Wielebinski, R. 2001, A&A, 371, 675
van de Hulst, H. C. 1945, Ned Tijdschr voor Natuurkd, 11, 210
van der Tol, S., Jeffs, B. D., & van der Veen, A.-J. 2007, IEEE Trans Signal Process, 55, 4497
van Diepen, G., & Dijkema, T. J. 2018, DPPP: Default Pre-Processing Pipeline
van Dishoeck, E. F., & Black, J. H. 1988, ApJ, 334, 771
van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, A&A, 556, 2
van Weeren, R. J., Williams, W. L., Hardcastle, M. J., et al. 2016, ApJSS, 223, 2
Villas, F. R., Martín-Pintado, J., González-Alfonso, E., et al. 2020, Submitt MNRAS, eprint arX, 14
Visser, R., van Dishoeck, E. F., & Black, J. H. 2009, A&A, 503, 323
Walch, S., Girichidis, P., Naab, T., et al. 2015, MNRAS, 454, 238
Walch, S. K., & Naab, T. 2015, MNRAS, 451, 2757
Walch, S. K., Whitworth, A. P., Bisbas, T., Wünsch, R., & Hubber, D. 2012, MNRAS, 427, 625
Watson, W. D., Western, L. R., & Christensen, R. B. 1980, ApJ, 240, 956
Weaver, R., Mccray, R., Castor, J., Shapiro, P., & Moore, R. 1977, ApJ, 218, 377
Wendker, H. J., Higgs, L. A., & Landecker, T. L. 1991, A&A, 241, 551
Westerhout, G. 1958, Bull Astron Institutes Netherlands, 14, 215
White, G. J., Abergel, A., Spencer, L., et al. 2010, A&A, 518, L114
Whitmore, B. C., Chandar, R., Schweizer, F., et al. 2010, AJ, 140, 75
Williams, W. L., Van Weeren, R. J., Rottgering, H. J. A., et al. 2016, MNRAS, 460, 2385
Wilson, C. D., Petitpas, G. R., Iono, D., et al. 2008, ApJSS, 178, 189
Wolfe, A. M., Gawiser, E., & Prochaska, J. X. 2005, ARAA, 43, 861
Wolfire, M. G., Hollenbach, D., & McKee, C. F. 2010, ApJ, 716, 1191
Wolfire, M. G., Hollenbach, D., McKee, C. F., Tielens, A. G. G. M., & Bakes, E. L. O. 1995, ApJ, 443, 152
Wolfire, M. G., McKee, C. F., Hollenbach, D., & Tielens, A. G. G. M. 2003, ApJ, 587, 278
Wright, N. J., Drake, J. J., Drew, J. E., et al. 2012, ApJL, 746, L21
Wright, N. J., Drake, J. J., Drew, J. E., & Vink, J. S. 2010, ApJ, 713, 871
Wright, N. J., Drew, J. E., & Mohr-Smith, M. 2015, MNRAS, 449, 741

Wright, N. J., Parker, R. J., Goodwin, S. P., & Drake, J. J. 2014, MNRAS, 438, 639
Xu, W. F., Gao, X. Y., Han, J. L., & Liu, F. S. 2013, A&A, 559, 81
Xu, X., Arav, N., Miller, T., & Benn, C. 2018, ApJ, 858, 39
Yorke, H. W., Tenorio-Tagle, G., Bodenheimer, P., & Rozyczka, M. 1989, A&A, 216, 207
Zari, E., Brown, A. G. A., de Bruijne, J., Manara, C. F., & de Zeeuw, P. T. 2017, A&A, 608, 148
Zhang, Q., & Fall, S. M. 1999, ApJL, 527, L81
Zhao, J., Anantharamaiah, K. R., Goss, W. M., & Viallefond, F. 1997, ApJ, 482, 186
Zhao, J.-H., Anantharamaiah, K. R., Goss, W. M., & Viallefond, F. 1996, ApJ, 472, 54
Zhu, H., Tian, W., Li, A., & Zhang, M. 2017, MNRAS, 471, 3494
Zuckerman, B., & Ball, J. A. 1974, ApJ, 190, 35

Samenvatting

Sterrenstelsels zijn grote, door zwaartekracht gebonden, verzamelingen van sterren (en alle planeten die hen omgeven), gas, stof, kosmische straling, massarijke zwarte gaten en donkere materie. Grote wolken van gas en stof, aangeduid als het interstellaire medium, vullen het merendeel van het volume van een sterrenstelsel. Daardoorheen verspreid zitten glimmende punten van sterrenlicht — sterren met een hele verscheidenheid aan verschillende grootten, massa's en helderheden. Hoogenergetische deeltjes, die zich met relativistische[3] snelheden voortbewegen, bekend als kosmische straling, vliegen door het sterrenstelsel dankzij magnetische velden die hen gevangen houden. Het centrum van de meeste sterrenstelsels bevat een superzwaar zwart gat. Behalve materie die met licht in het elektromagnetische spectrum interageert is er ook "donkere materie" die sterrenstelsels gravitationeel beïnvloedt.

Hoe deze basisingrediënten zich vormen, ontwikkelen en een sterrenstelsel beïnvloeden, stuurt modern astrofysisch onderzoek. Wanneer we kijken naar de gebieden in sterrenstelsels waar sterren gevormd worden, lijkt het bijna op een borrelde ketel. Als nieuwe sterren zich in (koude) interstellaire wolken vormen, vormen zij bellen van geïoniseerd en uitgestoten gas. Grote interstellaire wolken bestaan in verschillende karakteristieke fases (met uiteenlopende temperaturen, dichtheden en omvangen) — grotendeels weerspiegelen ze hiermee de globale invloed van sterren op het sterrenstelsel. De eigenschappen van deze wolken bepalen echter de omstandigheden voor nieuwe stervorming. Of een superzwaar zwart gat actief vrijelijk straling uitzendt of jets van stromende kosmische straling gevormd heeft, kan de samenstelling van een sterrenstelsel diepgaand beïnvloeden. Door de wisselwerking van deze fysische processen wordt een complex, wederzijds afhankelijk ecosysteem gecreëerd.

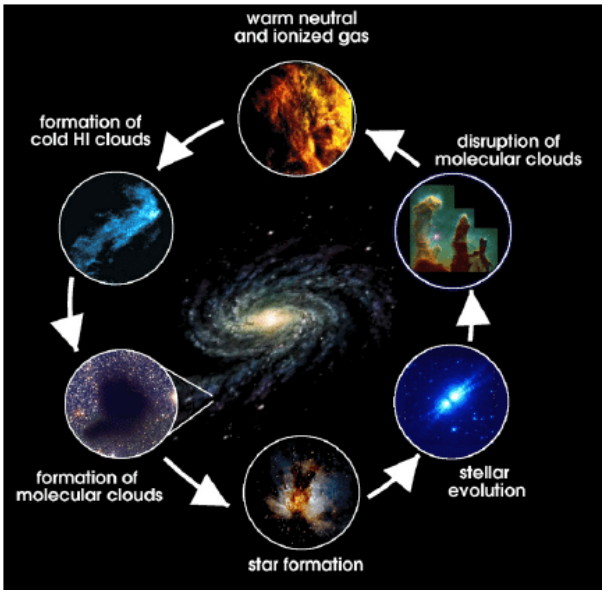
Deze ingewikkelde balans tussen deze processen bepaalt, hoe ingewikkeld ook, een aantal karakteristieke eigenschappen van een sterrenstelsel en zijn levensloop, die door sterrenkundigen vergaard worden.

Wat is de levenscyclus van een sterrenstelsel?

Figuur 5.22 vat de levenscyclus van sterren en interstellaire wolken samen voor een sterrenstelsel dat op onze Melkweg lijkt.

Gas dat koud is (10 K boven het absolute nulpunt!) en voornamelijk uit moleculen (niet uit atomen) bestaat vormt coherente gravitationeel gebonden wolken. Een tur-

[3] dat wil zeggen, met snelheden die een significant percentage van de snelheid van licht zijn



Figuur 5.22: De levenscyclus van interstellaire wolken

bulente gebeurtenis[4] in het sterrenstelsel veroorzaakt instabiliteiten en verdichtingen in een zogenaamde moleculaire reuzenwolk. Deze moleculaire reuzenwolk stort dan in, de verdichte klonten trekken in toenemende mate materiaal aan, en nieuwe sterren ontstaan.

Sterren fuseren elementen in hun kern — zo fuseren vier protonen, die elk ieder apart een waterstofkern zijn, tot een heliumkern. De sterren stralen licht uit en hebben een harde wind van kosmische straling aan hun oppervlak. Stellaire straling, in het bijzonder van de meest massarijke sterren, verhit het omliggende gas en stof, breekt moleculen af en stroopt elektronen van atomen af (ionisatie). Stellaire winden (en zelfs druk van de straling) duwen effectief tegen omliggende wolken aan, schokken en verhitten hen. Straling en winden stoppen instortend materiaal en vernietigen de restanten van hun geboorteplek — de moleculaire reuzenwolk — waardoor ze nieuwe stervorming tegenhouden.[5] Na een paar miljoen jaren zijn massarijke sterren niet langer omkleed in hun geboortemateriaal. Hun straling kan ver reiken, waardoor ze interstellaire wolken van het sterrenstelsel verhitten en uitgestrekte en doordringende componenten van het geïoniseerde interstellaire medium in stand houden.

Het leven van een ster eindigt wanneer hij niet langer atomaire kernen in zijn kern kan fuseren. Als sterren het einde van hun leven naderen, worden de buitenste lagen te zwaar om gedragen te kunnen worden. Sommige sterren stoten materiaal van hun

[4]Turbulentie kan ontstaan door een dichtbije exploderende ster of een grotere structuur in het sterrenstelsel zoals een spiraalarm of door een botsing met een ander sterrenstelsel.

[5]Tenminste, meestal wordt de wolk afgebroken en stervorming gestopt. Soms kunnen de schokken (van winden en straling) die door het medium bewegen ervoor zorgen dat sommige gebieden alsnog instorten en in deze dichte gebieden vormen wederom nieuwe sterren. Echter, onder de streep wordt meer stervorming voorkomen dan opgestart.

oppervlak af, materiaal dat gemaakt is van de kernen van "zware" elementen die in hun kern gefuseerd zijn. De meest massarijke sterren beëindigen hun leven met grote explosies genaamd supernova's. Door deze explosie stromen zware elementen nabije gaswolken in en verrijken hen. De schokgolf van de explosie dijt naar buiten uit, waarbij hij het medium verhit en kinetische energie omzet in een buitenwaartse druk.

Als sterren echter sterven, beginnen ook de interstellaire wolken die ze verwarmd hebben af te koelen. Soms hebben het gas en het stof in de omgeving van een ster of een groep van sterren zo'n grote zet meegekregen dat ze ver van de schijf van het sterrenstelsel vandaan vliegen en, ver van alles verwijderd, worden zij niet meer verhit. In elk geval koelen de gaswolken af. Moleculen beginnen zich te vormen, die afscherming van de wolk van straling vormen, en het materiaal koelt verder af en condenseert in een nieuwe moleculaire reuzenwolk. De voorwaarden voor nieuwe stervorming zijn bepaald, en de kringloop herhaalt zich.

Hoe ontstaan en ontwikkelen sterrenstelsels zich?

Het ontstaan en de ontwikkeling van sterrenstelsels is nauw verweven met de cyclische terugkoppelingsprocessen tussen sterren en interstellaire materie, wat sterrenkundigen stellaire feedback noemen. Afhankelijk van de nettohoeveelheid aan verhitting en koeling door de gezamenlijke eigenschappen van de sterren in een sterrenstelsel bereikt het interstellaire medium een evenwicht, waarin het gelaagde wolken met karakteristieke eigenschappen vormt. De omstandigheden in deze wolken reguleren het aantal sterren dat gevormd wordt. Echter, zoals we boven opmerkten, beïnvloeden sterren de wolken ook op kleinere (niet-globale) schalen door hen te verstrooien, en ze kunnen ook als de turbulente bemiddelaar optreden die ervoor zorgt dat andere moleculaire reuzenwolken instorten.

Het is zeer uitdagend om de wisselwerking tussen de verschillende betrokkene processen te begrijpen, want we kunnen de gebeurtenissen niet in echte tijd nabootsen. Ze voltrekken zich typisch in duizenden tot miljarden jaren; we krijgen slechts momentopnamen van de processen met onze huidige waarnemingen, en we proberen verschillende stadia van die processen aan elkaar te verbinden. Het is nog niet helemaal mogelijk om gesimuleerde films te creëren van de volledige collectieve processen door de fysische grondbeginselen die hen aansturen in te bouwen (hoewel veel voortgang geboekt is in het begrijpen van stukken daarvan!). Complexe interacties werken door van de kleinste schalen — met chemische netwerken in moleculaire wolken en invloedrijke processen van individuele sterren en zonnestelsels — helemaal tot de schaal van een geheel sterrenstelsel.

In de levensloop, die ik boven beschreven heb, is de relatie voornamelijk beschreven als die tussen een enkele ster en nabije wolken. Echter, sterren worden geboren in groepen op nagenoeg hetzelfde tijdstip en sommigen bevinden zich dicht genoeg bij elkaar zodat hun gezamenlijke bezigheid een eigen leven kan gaan leiden. Een typische moleculaire reuzenwolk heeft een massa van een miljoen zonnemassa's (10^6 $M_\odot$); dat wil zeggen, als het stervormingsproces 100% efficiënt is (en als slechts één soort ster gevormd wordt), zouden een miljoen kopieën van onze zon gevormd kunnen worden in de wolk. De turbulente gebeurtenissen die zich aan de wolk voltrekken, bewerkstel-

ligen dat de hele moleculaire reuzenwolk instort, en laat zo een stellaire wieg ontstaan van sterren met dezelfde leeftijd en plaats in het sterrenstelsel.

Ondanks dat moleculaire reuzenwolken gravitationeel gebonden zijn, geldt dit niet noodzakelijkerwijs voor de sterren die erin gevormd worden. Echter, we zien groepen van sterren die gevormd zijn en die gravitationeel gebonden zijn — we noemen deze groepen sterrenhopen. De meest massarijke en compacte sterrenhopen zijn vaak het meest efficiënt in het creëren van sterren. Waar in een sterrenstelsel deze sterren gevormd worden maakt ook verschil. In de schijf van een spiraalvormig sterrenstelsel, bijvoorbeeld, vormen zich nieuwe sterren typisch in de spiraalarmen. Massarijke en compacte sterrenhopen worden vaak dichter bij het centrum van een sterrenstelsel gevonden of grifweg als twee sterrenstelsels samensmelten.

Tot dusver heb ik de innige verbinding tussen sterren (hun locaties) en het interstellaire medium beschreven, en dit is waarop ik me hoofdzakelijk richt in mijn proefschrift. Hier is een lijst van andere factoren die ook belangrijk zijn voor de ontwikkeling van sterrenstelsels, maar die ik niet uitgebreid zal bespreken: (i) intergalactisch gas dat op sterrenstelsels accreteert en nieuw materiaal in het levenscyclusproces injecteert, (ii) de omgeving van een sterrenstelsel en hoe vaak het wisselwerkt met andere sterrenstelsels, (iii) een actief superzwaar zwart gat in het centrum van een sterrenstelsel dat interstellaire materie verslindt en waarvan straling en/of relativistische jets van heet plasma de kringloop van fasen tussen de wolken beïnvloedt, en (iv) de aanvankelijke verdichting in de ruimte na de oerknal, die de formatie van een sterrenstelsel op gang heeft gebracht, zou een hoofddeterminant van de eigenschappen van een sterrenstelsel kunnen zijn.

Hoe bestuderen sterrenkundigen interstellaire wolken en de effecten van sterren?

Om de wisselwerking tussen sterren en het interstellaire medium te begrijpen middels waarnemingen, voeren sterrenkundigen gedetailleerde hoge-resolutiestudies van individuele sterren en wolkenklonten uit, gebruiken ze grote steekproeven en overzichtsonderzoeken van astronomische objecten om statistisch inzicht te verkrijgen in fysische mechanismen, en maken ze schattingen op galactische schaal en hoe hun eigenschappen ontwikkelen in kosmische tijd. Zo worden dieptestudies gewoonlijk uitgevoerd in onze eigen Melkweg. De nabijheid staat toe dat we hen goed zien en we hebben een grote verscheidenheid aan manieren om het gas te detecteren en de invloed van fysische processen af te leiden, wat niet mogelijk zou zijn in sterrenstelsels die verder weg zijn en daardoor zwakkere signalen hebben. Echter, het is vaak nuttig om een vogelperspectief in te nemen en nabije sterrenstelsels in uiteenlopende oriëntaties, die verschillende omgevingen kunnen herbergen, te observeren. Omdat licht met een eindige snelheid voortbeweegt ($c = \nu\lambda = 1$ miljard kilometer per uur)[6], werd het licht dat ons nu bereikt van de meest afgelegen sterrenstelsel uitgezonden

[6]In de vergelijking $c = \nu\lambda$ is ν de frequentie van het licht en λ de golflengte; de snelheid van licht is constant (in een vacuüm), dus als de golflengte kleiner wordt, wordt de frequentie groter, en andersom, als de golflengte langer wordt, wordt de frequentie korter en het licht heeft lagere energie (omdat energie met de frequentie schaalt).

toen die sterrenstelsels jong waren. Zodanig (en omdat er aanwijzingen zijn dat het heelal homogeen en isotroop is) kunnen we sterrenstelsels observeren als functie van de leeftijd van het heelal om te zien hoe de eigenschappen van sterrenstelsels zich ontwikkelen in kosmische tijd.

Een schoonheid van de sterrenkunde is dat we het werkelijk eenvoudige concept gebruiken van het waarnemen van licht over het electromagnetische spectrum met verschillende golflengtes (frequenties, energie) om de fysica en de wonderen van ons heelal af te leiden.[7] Van het simpele concept om tweedimensionale plaatjes op verschillende golflengtes te maken, kunnen we de aanwezigheid van astronomische objecten, de fysische eigenschappen van deze objecten, hun driedimensionale verdeling en bewegingen construeren, hoe objecten elkaar beïnvloeden en hoe ze zich ontwikkelen in de tijd (vier dimensies).

De hoofdwegen om interstellaire wolken door continuümlicht te bestuderen zijn: infraroodlicht (golflengtes van $\lambda \sim 0.001 - 0.5$ mm) van stof met uiteenlopende temperaturen; het licht van vrije elektronen op radiogolflengtes ($\lambda \sim 5$ mm – 30 m); Röntgenstraling dat in een continuüm door heet gas uitgezonden wordt; en gammastraling, fotonen met de hoogste energie (en onvoorstelbaar kleine golflengtes, $\lambda < 10^{-11}$ mm) die uitgezonden worden als hoogenergetische deeltjes met interstellaire wolken botsen.

Naast continuümemissie kunnen bestanddelen van sterrenstelsels ook waargenomen worden door middel van spectroscopische signalen. Een wolk kan licht met een specifieke frequentie uitzenden (of absorberen) ten gevolge van overgangen van elektronen, vibraties en rotaties in atomen en/of moleculen. We weten bijvoorbeeld van het bestaan van grote wolken van atomaire waterstof door een spinomkering van een elektron in een waterstofatoom — wat zich bij een golflengte van 21 cm in het radioregime van het spectrum voltrekt. De meest gebruikelijke manier waarmee we moleculaire reuzenwolken kunnen zien is door het waarnemen van rotaties van koolstofmonoxidemoleculen (CO-moleculen) als spectraallijnen. Warm geïoniseerd gas dichtbij massarijke sterren, in de buitengebieden van een galactische schijf of in de buurt van massarijke zwarte gaten, emitteren spectraalovergangen van waterstof bij optische/zichtbare golflengten (zoals ook sommige overgangen van elementen als zuurstof en zwafel). Op lange infraroodgolflengten emitteren wolken een overschot van emissie van een overgang van koolstof. Met verbeterde technologie kunnen we steeds zwakkere signalen ontdekken en kunnen we telescopen bouwen die op golflengten van het electromagnetische spectrum waarnemen die voorheen te gecompliceerd waren om na te jagen (terwijl veel makkelijkere dingen nog steeds onontdekt bleven!).

Zwakke spectrale signalen die ik waarneem (en voor het eerst in verre sterrenstelsels) worden **radiorecombinatielijnen** genoemd. Radiorecombinatielijnen worden uitgezonden wanneer een geïonseerd atoom (in het geval van waterstof een proton) en een elektron recombineren, en de elektronen recombineren naar een hoog energieniveau en vervolgens naar lagere energieniveaus vallen. De elektronenovergangen die tussen de hoge toestanden plaatsvinden — de radiorecombinatielijnen — hebben zeer lage energie en zijn daarom op radiogolflengtes waar te nemen. De intensiteit van

[7] Uitzonderingen daarvan omsluiten de directe detectie van kosmische straling (hoogenergetische deeltjes) vanuit de ruimte of nieuwe detectoren van zwaartekrachtsgolven. Deze waarnemingen zijn zo uniek waardevol omdat ze het heelal om een geheel andere manier verkennen dan traditionele sterrenkunde.

de recombinatielijnen is direct evenredig met het invallende ioniserende stralingsveld. Terwijl optische en infraroodlijnen verzwakt of verstrooid kunnen worden door stof, komen deze spectraallijnen onverzwakt uit het medium. Alhoewel zwak en moeilijk te detecteren, zijn ze bijzonder nuttig om de aanwezigheid van massarijke sterren af te leiden in ondoorzichtige zichtlijnen (bijvoorbeeld in het vlak van de Melkweg) en in dichte omgevingen van jonge stervormingsgebieden.

In gas met lage dichtheid kan iets gebeuren terwijl de elektronen door de energieniveaus van het atoom vallen: botsingen met een ander vrij elektron of radiostraling kan met het atoom wisselwerken. De fysische eigenschappen (temperatuur en dichtheid) van het diffuse gas zijn zeer gevoelig voor hoe effectief deze twee processen zijn en bij welke energieniveaus zij plaatsvinden. We nemen de radiorecombinatielijnen op verschillende frequenties waar, meten hoe de intensiteit verandert, en gebruiken dat om de temperatuur, dichtheid en omvang van de gaswolken te achterhalen. Als we de fysische toestand meten, kunnen we de invloed van de verschillende mechanismes kwantificeren, zoals verhitting/koeling, duwing, verspreiding of opschudding van het gas. Omdat we geen manier hebben om dat te doen met eerder gebruikte spectraal- en continuümssignalen leveren waarnemingen van radiorecombinatielijnen diepe inzichten op.

Ik gebruik een nieuw geconstrueerde radiotelescoop, de Low Frequency Array, **LOFAR**, om continuümemissie en radiorecombinatielijnemissie waar te nemen. LOFAR is gecentreerd in en wordt bestuurd vanuit Nederland. LOFAR is ontworpen om radiogolflengten van 1 – 30 m of frequenties van 10 – 250 MHz waar te nemen, met een gat bij 90 – 100 MHz dat de FM-banden van de radio die we in onze auto's oppikken omspant. Dit zijn de laagste frequenties waarmee de buitenaardse hemel vanaf de grond waargenomen kan worden, met een kantelwaarde bij ongeveer 10 MHz. Lagere frequenties worden slechts door de atmosfeer teruggekaatst. Naarmate radiogolven dichter bij de kantelfrequentie komen, worden ze steeds meer verstoord (gebroken en gedraaid in polarisatie) als ze door de atmosfeer gaan. De laag die hiervoor verantwoordelijk is wordt ionosfeer genoemd en de verstorende effecten variëren in tijd en plaats in het waargenomen gezichtsveld.

LOFAR is een (van een paar) nieuwe generatie radiointerferometers en maakt voor het eerst waarnemingen op deze frequenties met hoge resolutie en hoge gevoeligheid mogelijk. Interferometers bevatten meerdere antennes die gezamenlijk gebruikt worden om een astronomisch object waar te nemen. In plaats van één gigantische telescoop wordt een virtuele telescoop opgericht door de signalen van de antennes te harmoniseren. LOFAR gebruikt heel simpele antennes — er zijn twee types voor de twee verschillende golflengtebereiken. In principe zijn het vier stukken metaal (of draad) met een specifieke lengte, zodat, als radiogolven met de gewenste lengte hen bereiken, het radiolicht een stroom in de antennes induceert. Sterrenkundige synthetiseren het uitgelezen signaal van de stroom om een beeld te maken. De antennes van LOFAR zijn goedkoop en kunnen veelvuldig worden neergezet om gevoeligheid en verzamelvlak op te bouwen. Traditioneel werden bestuurbare schotels gebruikt om op een specifieke plek op de hemel uit te richten. De LOFAR-telescoop wordt digitaal uitgericht door de verwachte vertraging over de telescoop voor een specifieke locatie op de hemel in te prenten. Dit antenneontwerp maakt het ook mogelijk om grote stukken hemel gelijktijdig waar te nemen — zo kan bijvoorbeeld een oppervlakte aan

de hemel waargenomen worden die gelijk is aan 38 manen (de grootste schijnbare grootte van de maan) bij de hoogste frequenties en een oppervlakte van groter dan 2800 manen bij de laagste frequenties!

Een andere telescoop die ik gebruik wordt de Atacama Large Millimeter/Submillimeter Array, **ALMA**, genoemd, gelegen in Chili. Ook recent gebouwd (tenminste was het dat aan het begin van mijn promotietraject) in 2013 heeft het voor het eerst zeer gevoelige waarnemingen met hoge resolutie bij millimetergolflengtes mogelijk gemaakt. Er zijn veel moleculaire en atomaire overgangen die bij deze golflengten waargenomen kunnen worden, hetgeen nieuwe manieren oplevert om astrofysische eigenschappen middels astrochemie te achterhalen. ALMA is ook een interferometer, maar het bevat radioschotels met instrumenten die licht met golflengten van 0.8 – 3.5 mm opvangen. Het is de grootste sterrenwacht die op deze golflengten opereert — de toegevoegde antennes bereiken een betere resolutie en sensitiviteit.

Wat heb ik onderzocht in dit proefschrift?

In mijn onderzoek tijdens mijn promotietraject heb ik interstellaire wolken bestudeerd op schalen van sterrenstelsels tot schalen kleiner dan wolken. Ik neem radiocontinuümemissie waar van het ijle geïoniseerde gas in een stervormingsgebied in onze Melkweg en bestudeer de invloed van massarijke sterren door hun vervoer van ioniserende fotonen. Ik onthul de aanwezigheid van een populatie van zeldzame supersterrenhopen en karakteriseer hun eigenschappen. Ik verken de mogelijkheid om radiorecombinatielijnen te gebruiken voor de studie van sterrenstelsels op kosmische afstanden en ik ontwikkel gereedschap en strategieën die nodig zijn om lage-frequentiewaarnemingen te verwerken. Terwijl ik deze inspanningen leid werd het onderzoek uitgevoerd met de hulp en bijdragen van een schare van medewerkers en teamleden en door de geleiding van mijn begeleiders.

In hoofdstuk 1 geef ik een inleiding (in groter detail, voor sterrenkundigen) om de concepten die we bestuderen in te kaderen. Specifieker zijn de vragen die ik in de daaropvolgende hoofdstukken wil aangaan:

- *Hoofdstuk 2:* Hoe beïnvloedt het ijle geïoniseerde gas de evolutie van het massarijke, galactische stervormingsgebied Cygnus X? Zijn dezelfde vingerafdrukken aanwezig in studies van het ijle gas in onze Melkweg?

- *Hoofdstuk 3:* Wat zijn de eigenschappen van stervorming (sterrenhopen) in de centrale *starburst* van het sterrenstelsel NGC 4945?

- *Hoofdstuk 4:* Kan het ISM worden verkend buiten het locale universum door waarnemingen van radiorecombinatielijnen? Wat zijn de ISM eigenschappen van een dwergachtig sterrenstelsel bij $z = 1.1$?

- *Hoofdstuk 5:* Welke technieken zijn het meest geschikt om zwakke radiorecombinatielijnen (bij onbekende roodverschuiving) in extragalactische bronnen te detecteren?

Vooruitzichten

Zo veel wacht erop om ontdekt te worden door op het onderzoek in dit proefschrift voort te bouwen. De analyse die we in *Hoofdstuk 2* presenteren — een analyse van het thermische radiocontinuüm van een stervormingsgebied in onze Melkweg — is de eerste van zijn aard die de LOFAR-telescoop gebruikt; ongeveer de helft van het galactische vlak kan door LOFAR waargenomen worden en verdere waarnemingen zullen onze kennis over de aanwezigheid en de oorsprong van deze overvloedige gasfase van het interstellaire medium statistiek verbeteren. Bovendien legt deze arbeid de grondslag voor complete LOFAR-waarnemingen van radiorecombinatielijnen van diffuse moleculaire wolken en geïoniseerd gas in het gebied van Cygnus X. Door de aanwezigheid van een zeldzame populatie van supersterrenhopen te onthullen en hun fundamentele eigenschappen te karakteriseren in *Hoofdstuk 3* levert deze arbeid een waardevolle dataset op om stellaire terugkoppeling en de invloed van sterrenhopen op het medium te bestuderen, evenals de vroege formatiefasen in de ontwikkeling van deze typen van sterrenhopen, hoeveel sterrenhopen met een gegeven massa ontstaan en wat hun hiërarchie beïnvloedt, hoe eigenschappen van sterrenhopen zich verhouden met hun interstellaire medium, en een breder perspectief is dat deze waarnemingen gebruikt kunnen worden samen met supersterrenhopen in andere nabije sterrenstelsels om mogelijkerwijs typen van eigenschappen van stervorming aan te wijzen die vaker voorkwamen toen sterrenstelsels in het heelal gezamenlijk de meeste sterren vormden. Waarnemingen met ALMA om deze sterrenhopen te ontdekken en de inzichten in ons heelal die zij opleveren zijn pas begonnen. Een hoofdboodschap van dit proefschrift, en *Hoofdstukken 4 en 5* in het bijzonder, is dat de fysische eigenschappen in het diffuse interstellaire medium verkend kunnen worden door middel van waarnemingen van radiorecombinatielijnen in en/of tegen radio-heldere sterrenstelsels tot op enorme afstanden. Dit opent de deur naar onderzoeken met bestaande faciliteiten zoals LOFAR, VLA, GMRT, WSRT en ASKAP. Voortgaande studies die naar de 21-centimeteremissie van atomaire waterstof in populaties van sterrenstelsels zoeken zijn prachtige complementen en kunnen gebruikt worden voor verdere detecties van radiorecombinatielijnen. De langere-termijn-toekomst is tamelijk glorend omdat een nieuwe telescoop, die nu net gebouwd wordt, de Square Kilometer Array (SKA), ons in staat zal stellen om aanzienlijk gevoeligere waarnemingen bij lage radiofrequenties uit te voeren en de (extragalactische) studies van radiorecombinatielijnen zal revolutioneren door de detectie van tienduizenden bronnen mogelijk te maken.

Summary

Galaxies are large gravitationally-bound collections of stars (all the planets that surround them), gas, dust, cosmic rays, black holes, and dark matter. Large clouds of gas and dust, referred to as the interstellar medium, fill most of a galaxy's volume. Intermixed are shining points of stellar light — stars with a whole slew of different sizes, masses and brightnesses. High energy particles moving at relativistic[8] speeds, known as cosmic rays, zip around galaxies due to influences by magnetic fields. The center of most galaxies contains a super massive black hole. In addition to matter which interacts with light in the electromagnetic spectrum, there is also "dark matter" which influences galaxies gravitationally.

Understanding how these basic components coalesce, evolve, and influence a galaxy guides modern astrophysical research. When we look in regions of galaxies where stars are forming, it resembles a bubbling cauldron — for example, see the Cygnus X star-forming in Figure 5.23. As new stars form in (cold) interstellar clouds, they create pockets of ionized[9] and blown out gas. These bubbles expand and often spill out from a galaxy's disk. Large interstellar clouds exist in different characteristic phases (with various temperatures, densities and sizes) — largely reflecting the global galaxy-wide influence of stars. Yet the properties of clouds set the conditions for new star formation. If the supermassive black hole has an accretion disk that is actively emitting copious radiation or has formed jets of streaming cosmic rays, that may also profoundly influence the make up of a galaxy. A complex, interdependent ecosystem is created by the interplay of these physical processes.

Although intricate, the balance acting between these processes does set up some characteristics features of a galaxy and its life cycle that astronomers have come to gather.

What is the lifecycle of interstellar clouds?

Figure 5.24 summarizes the lifecycle of stars and interstellar clouds in a galaxy that is much like our Milky Way.

Gas that is cold (10 K above absolute zero!) and that largely consists of molecules (rather than single atoms) forms coherent gravitationally-bound clouds. A turbulent event[10] in the galaxy causes instabilities and over-densities in the so-called giant

[8] moving at velocities that are a significant fraction of the speed of light

[9] Ionized gas is made up of atoms that have electron(s) removed.

[10] Turbulence can come from a nearby exploding star or larger structure in the galaxy like a spiral

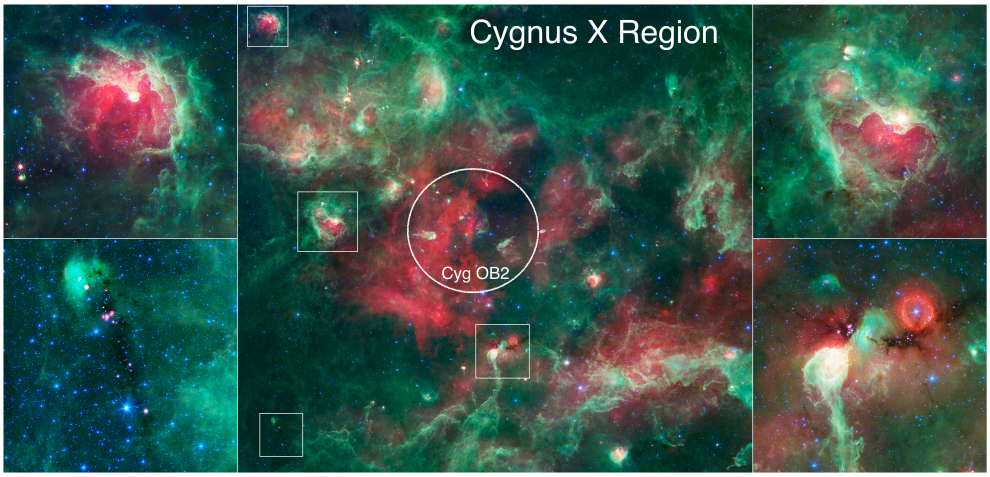


Figure 5.23: The Cygnus X star-forming region, a bubbling cauldron of star birth surrounding the Cyg OB2 association — observed with the *Spitzer* Space Telescope: 3.6 μm in blue, 4.5 μm in blue-green, 8.0 μm in green, and 24 μm in red. Irradiated surfaces of molecular clouds appear bright in green-white. Massive stars that have blown bubbles and cavities in the dust and gas are evident where green-white emission surrounds a bubble of glowing red emission from warm dust (ionized gas). The brightest, yellow-white regions are warm centers of star formation. Tendrils of dust appear green, and stars generally appear as blue point sources. The pillar-like and elongated features directed away from Cyg OB2 have been shaped by the stellar radiation and winds from this massive ($M_\star \approx 2 \times 10^4$ M$_\odot$) association. The boxed zoom-in regions show massive star(s) (formation) in AFGL 2636 (upper left) and DR22 (upper right). The lower left and right (DR15) images show clouds that are so thick to be absorbed at the mid-infrared wavelengths of Spitzer. Young stars, visible as red points, are buried in the dark clouds. Image credit: NASA / JPL-Caltech / Harvard-Smithsonian.

molecular cloud. At this point, the giant molecular cloud collapses as the over dense pockets attract an ever-growing amount of material and new stars form.

Stars fuse elements in their core — for example, four protons (i.e., four hydrogen nuclei) fuse to create a helium nucleus. Stars radiate light and have a gusting wind of cosmic rays at their surface. Stellar radiation, especially from the most massive stars, heats up the surrounding gas and dust, dissociating molecules and stripping atoms of their electrons (ionization). Stellar winds (and even pressure from radiation) effectively push on surrounding clouds, shocking and heating them. Radiation and winds halt collapsing material and destroy the remains of their natal giant molecular cloud, thereby preventing new stars from forming.[11] After a few million years, massive stars are no longer enshrouded in their natal material. At this point in time, their radiation is able to reach large distances, heating the interstellar clouds of the galaxy and maintaining extended and pervasive components of ionized gas.

The life of a star ends when it can no longer fuse atomic nuclei in its core. As stars approach the end of their lives, the outer layers become too heavy to support. Some stars burp out and slough off material from their surface, material made up of the nuclei of "heavy" elements that were fused in their core. The most massive stars end their lives with great explosions called supernovas. With that explosion, heavy elements stream into and enrich nearby gas clouds. The shock wave from the explosion expands outwards, also heating the medium and imparting kinetic energy

arm or due to a collision with another galaxy.

[11] At least, breaking up the cloud and halting star formation happens most of the time. In a smaller fraction of instances, as shock waves (from the winds and radiation) move through the medium, they can cause some regions to collapse in on themselves and in those dense regions new stars actually form. However, the net effect is that more star formation is stopped than gets initiated.

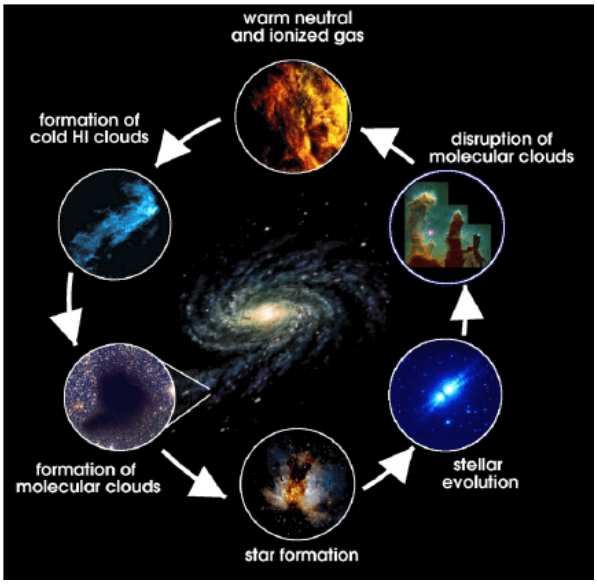


Figure 5.24: The lifecycle of interstellar clouds.

into an outward push.

As stars die off, the interstellar clouds which they have warmed start to cool off too. Sometimes the gas and dust surrounding massive stars or groups of stars has been given such a big push that the clouds are sent speeding away from the galaxy's disk, and far away, they are no longer heated. In any case, the gas clouds cool down. Molecules start to form, the molecules help the cloud to shield itself from radiation, and the material further cools and condenses forming a giant molecular cloud. The conditions for new star formation are set, and the cycle repeats.

How do galaxies evolve?

The origin and evolution of galaxies are closely tied to the cyclic feedback processes between stars and interstellar matter, what astronomers call stellar feedback. Depending on the net amount of heating and cooling that occurs, the interstellar medium equilibrates, forming stratified clouds of characteristic properties. Cloud conditions regulate the number of stars that form. But as we mentioned above, stars also influence the clouds on smaller (non-global) scales by dispersing them, and they may also act as turbulent agents.

Understanding stellar feedback, i.e., the interplay of the different processes involved, is quite challenging because we can not play out the events in real time. They typically occur over thousands to billions of years; so we just get snapshots of the process with our current observations, and we try to piece together different stages of the processes. Creating simulated movies of the entire collective process by incorporating the basic physics that govern it is not fully possible yet (though great progress has been made to understand pieces of it!). Complex interactions transpire on the smallest scales — with chemical networks inside molecular clouds and influential processes occurring from individual stars and solar systems — all the way up to the scale of an entire galaxy.

In the lifecycle I described above, the relationship is mainly framed around the stellar feedback from a single star and a nearby cloud. However, stars are born in groups at similar times and some reside close enough to each other such that their concerted efforts create compounding effects. A typical giant molecular cloud has a mass of about a million times the mass of our sun; if the star formation process is 100% efficient (and if only one type of star is formed), a million of our suns could form from one cloud. Turbulent events that act on clouds usually cause the entire giant molecular cloud to collapse, thus producing a stellar nursery of stars of similar ages and locations in the galaxy.

Even though giant molecular clouds are gravitationally bound, the stars that form within it won't necessarily be bound to one another. However, we do see groups of stars that have formed and are gravitationally bound — we call those star clusters. The most massive and compact star clusters tend to be the most efficient at creating stars. Where in a galaxy these stars form also makes a difference. Within the disk of a spiral galaxy, for example, new stars typically form in spiral arms. Massive and compact star clusters tend to form close to the center of galaxies (as can be seen in Figure 5.25) or promptly when two galaxies merge.

So far, I have described the intimate connection between stars (their locations),

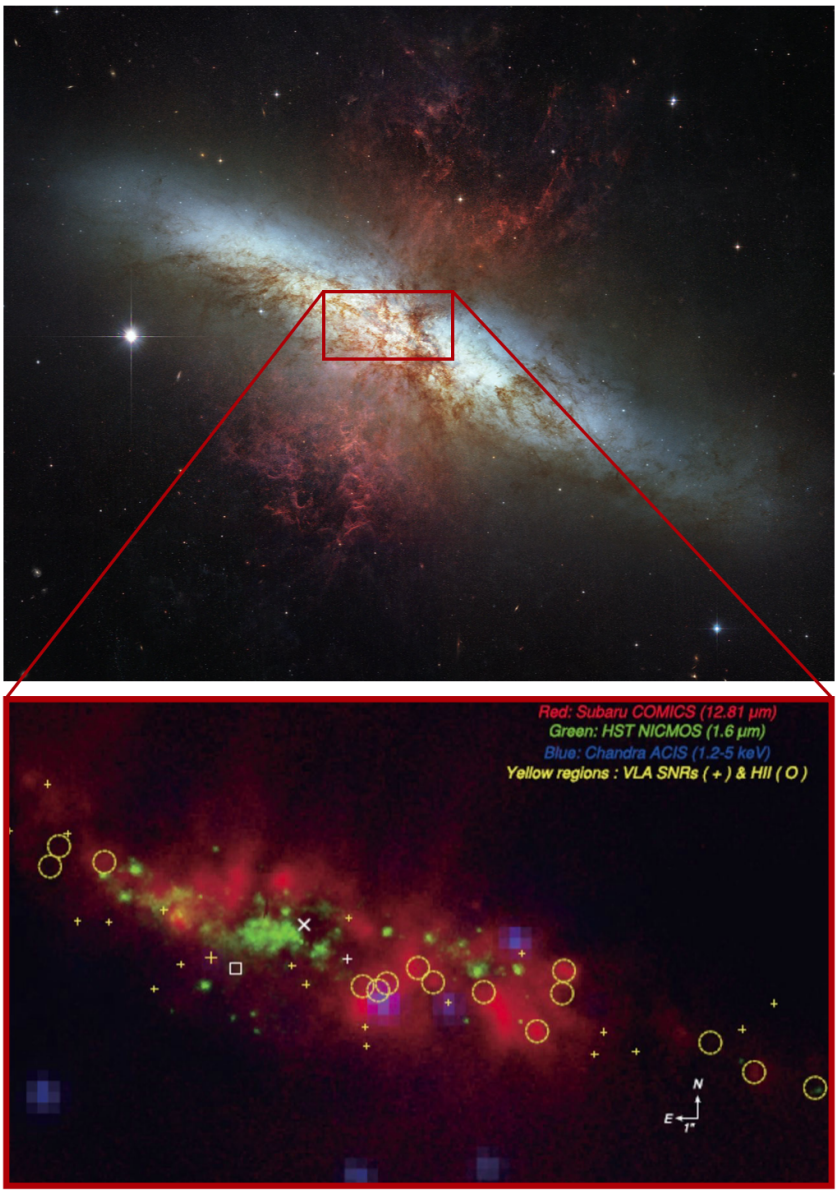


Figure 5.25: In the top panel, the disk of the nearby (3.5 Mpc) edge-on galaxy, M82, glows in white-blue and the prominent galactic winds from the nucleus are traced by the PAH emission in red. Super star clusters (yellow circles) in the nucleus (in the bottom panel) drive the outflow. Image credits: NASA / ESA / The Hubble Heritage Team.

stellar feedback, and the interstellar medium, and this is primarily what I focus on in my thesis. The following are additional factors which are important to the evolution of galaxies but which I do not discuss at length: (i) intergalactic gas that accretes onto galaxies and injects new material into the life-cycle process, (ii) a galaxy's environment and how often it interacts with other galaxies, (iii) an active supermassive blackhole at the center of a galaxy that is gobbling up interstellar matter and whose radiation and/or relativistic jets of hot plasma impact the cycling of intercloud phases, and (iv) the initial over density in space after the big bang which seeded the formation of a galaxy might be a major determinant in a galaxy's properties.

How do astronomers study interstellar clouds and stellar feedback?

To understand the interstellar medium and its interplay with stars using observations, astronomers make detailed, high-resolution studies of individual stars and cloud clumps, use large samples and surveys of astronomical objects to statistically assess physical mechanisms, and make galaxy-wide assessments on how their properties track over cosmic time. In-depth studies are commonly performed within our own Galaxy for example. The proximity allows us to see objects in the Galaxy close up at high resolution, and we have a multitude of ways to detect the gas and infer the influence of physical processes that would not be possible in galaxies that are further away and thus have fainter signals. However, it is often useful to take a bird's-eye-view and observe nearby galaxies in a variety of orientations, which may harbor different environments. Because light travels at a finite velocity ($c = \nu\lambda = 0.7$ billion miles-per-hour)[12], the light that is just reaching us now from the most distant galaxies was emitted when those galaxies were younger. In this way (and since there is evidence that the universe is homogeneous and isotropic), we can observe galaxies as a function of the age of the universe to see how the bulk properties of galaxies evolved over cosmic time.

A beauty of astronomy is that we use the really simple concept of observing light along the electromagnetic spectrum with different wavelengths (frequencies, or energy) to infer the physics and wonders of our Universe.[13] From the simple concept of making 2D pictures at different wavelengths, we infer the presence of astronomical objects, the physical properties of the objects, their 3-dimensional distributions and movements, how objects influence each other, and how they evolve over time (adding a fourth 4th dimension to the 3D picture!).

The main ways we study interstellar clouds through continuum light are: infrared light (wavelengths of $\lambda \sim 0.001 - 0.5$ mm) from dust at a range of temperatures, the emission from free electrons that is radiated at radio wavelengths ($\lambda \sim 5$ mm – 30 m) when the electrons pass close to one another, hot gas emits a continuum of X-ray

[12] In the equation $c = \nu\lambda$, ν is the frequency of light and λ is the wavelength; the speed of light is constant (in a vacuum), so as the wavelength gets smaller the frequency goes up and vice versa, as the wavelength is longer the frequency is shorter and the light has lower energy (since energy is proportional to frequency).

[13] Exceptions to this include directly detecting cosmic rays (high energy particles) from space or new gravitational wave detectors. These observations are so uniquely valuable because they explore the universe in an entirely different way than traditional astronomy.

light, and as high-energy particles collide with interstellar clouds gamma-ray emission — the highest energy photons (with inconceivably small wavelengths, $\lambda < 10^{-11}$ mm) — is emitted.

In addition to continuum emission, another way to observe the constituents of galaxies are through spectral lines. Spectral lines occur when atoms and molecules produce (or absorb) light at specific frequencies. This process takes place whenever the energy of an atom or molecule changes, which takes place relatively often in interstellar clouds. These energy changes might be due to vibrations and rotations of molecules or changes in the configuration of the electrons in an atom. For example, we know of the existence of large clouds of atomic hydrogen through a spin flip of an electron in a hydrogen atom, which occurs at a wavelength of 21 cm in the radio regime of the spectrum. The most common way that we are able to see giant molecular clouds is through rotations of carbon monoxide (CO) molecules that we observe through spectral lines. Warm ionized gas that is found near massive stars, on the outskirts of a galactic disk, and in the vicinity of massive black holes, emits spectral transitions of hydrogen at optical/visible wavelengths. At long infrared wavelengths, interstellar clouds emit a surplus of emission from a transition in carbon. As technology improves, we uncover fainter and fainter signals and construct telescopes that observe at wavelengths of the electromagnetic spectrum that were previously too complicated or costly to pursue.

Faint spectral signatures that I observe (and have observed for the first time in a distant galaxy, Chapter 4) are called **radio recombination lines** (RRLs). Radio recombination lines occur when an ionized atom (in the case of hydrogen, a proton) and an electron recombine. In some cases, the electron ends up in a high energy state after recombination, and from there it cascades down to lower energy levels. The steps down this cascade have very low and well-defined energies and are thus observable at specific wavelengths in the radio regime of the spectrum. The former makes them appear as "lines" in a spectrum and the latter puts the "radio" in radio recombination lines. While optical and infrared lines may be attenuated and scattered by dust, these spectral lines at radio wavelengths make it out of the medium unattenuated. Although they are faint and harder to detect, they are especially useful to infer the presence of massive stars towards obscured lines of sight (e.g., along the plane of the Galaxy) and in the dense environments of young star-forming regions.

In low-density gas, as a freshly recombined electron cascades down the energy levels of the atom, radio light and collisions with free electrons influence the cascading process. How effective those two processes can be during the cascade process and at which energy levels of the atom they occur are very sensitive to physical properties (temperature and density) of the diffuse gas. We observe the radio recombination lines at different frequencies, measure how the intensity changes, and use that to determine the temperature, density, and size of the gas clouds. When we measure the physical conditions, we can then quantify the influence of the different mechanisms which may be heating/cooling, pushing, diffusing, or stirring up the gas. Measuring these physical conditions using other means is usually not as direct, so observations of radio recombination lines provide a powerful complementary method for studying the interstellar medium.

I use a newly-constructed radio telescope, the **Low Frequency Array** (LOFAR),

to observe continuum emission and radio recombination line emission. LOFAR is centered in and operated from the Netherlands, but it also has stations in nine European countries. LOFAR was designed to observe at radio wavelengths of 1 – 30 m (3 – 100 feet!) or frequencies of 10 – 250 MHz, with a gap between 90 – 110 MHz straddling the FM Radio that we tune to in our cars and homes. These are the lowest frequencies that make it in and out of the atmosphere, as radio waves longer than about 30 m (10 MHz) are only reflected off the atmosphere. Radio waves increasingly closer to the cutoff frequency get distorted (refracted and rotated in polarization) as they pass through the atmosphere. Specifically, the layer responsible for this is called the ionosphere, and the distorting effects vary in time and location across the sky.

LOFAR is (one of a few of) a new generation radio interferometer and is enabling observations at these frequencies with high resolution and high sensitivity for the first time. Interferometers contain multiple antennas that can be used in conjunction to observe an astronomical object. Instead of having one gigantic telescope, a virtual telescope is created by harmonizing the signals from the antennas. LOFAR makes use of very simple antennas — there are actually two types for the two different wavelength ranges. Essentially, they are four pieces of metal (or wire) made with a specific length such that when radio waves of the desired length reach it, the radio light induces current in the antennas. Astronomers synthesize the read-out of the current to create images. LOFAR's antennas are inexpensive, and many of them can be installed to build up sensitivity and collecting area. Traditionally, steerable dishes have been used to point at a specific location in the sky. The LOFAR telescope is digitally pointed by imprinting the expected time delay observed across the telescope for a specific location in the sky. This antenna design also enables large patches of the sky to be observed simultaneously — for example, an area on the sky is observed simultaneously that equals 38 moons (the largest apparent size of moon) at the highest frequencies and an area of more than 2800 moons at the lowest frequencies!

Another telescope I use is called the **Atacama Large Millimeter/Submillimeter Array** (ALMA), located in Chile. Also recently constructed (at least it was at the start of my PhD program) in 2013, it has enabled high resolution and very sensitive observations at mm (and sub-mm) wavelengths for the first time. There are many molecular and atomic transitions that can be observed at these wavelengths, providing new ways to probe astrophysical properties through astrochemistry. ALMA is also an interferometer, but it is comprised of radio dishes with instruments that capture light with wavelengths of 0.8 – 3.5 mm. It is the largest observatory of its kind operating at these wavelengths — additional antennas achieve a better resolution and sensitivity.

What did I research in this thesis?

In my research throughout the PhD program, I investigated interstellar clouds on the scales of galaxies down to sub-cloud scales. I observed radio continuum emission from low-density ionized gas in a star forming region in our Galaxy, Cygnus X (remember the bubbling cauldron from Figure 5.23?) and studied the influence of massive stars through their transport of ionizing photons. I revealed the presence of a population of

rare super-star clusters in nearby galaxy and characterized their properties. I explored using radio recombination lines to observe galaxies at cosmological distances (very far away and young), and I developed the tools and strategies needed to process new low-frequency observations. While I lead these efforts, this research was carried out with the help and input from a host of collaborators and team members and through the guidance of my advisors.

In Chapter 1 of this thesis, I give an introduction (in more detail, for astronomers) to frame the concepts that we investigated. The research questions that are addressed in additional chapters:

- *Chapter 2:* How does low-density ionized gas affect the evolution of the massive, galactic star-forming region, Cygnus X? Are the same fingerprints present in surveys of low-density ionized gas in our Galaxy?
- *Chapter 3:* What are the properties of star formation (star clusters) in the central starburst of the galaxy NGC 4945?
- *Chapter 4:* Can the ISM be explored outside of the local universe through radio recombination line observations? What are the properties of the interstellar medium in a dwarf-like galaxy at $z = 1.1$?
- *Chapter 5:* What techniques are best suited to detect faint radio recombination lines (at an unknown redshift) in extragalactic sources?

What are the next steps?

So much is waiting to be discovered by building off of the research in this thesis. The analysis we present in *Chapter 2* — the thermal radio continuum from a star forming region in our Galaxy — is the first of its kind using the LOFAR telescope; roughly half of the galactic plane is observable by LOFAR and additional observations will statistically inform on the presence and origins of this plentiful gas phase of the interstellar medium. Furthermore, this work lays the foundation for LOFAR observations of radio recombination lines from diffuse molecular clouds and ionized gas in the Cygnus X region. By uncovering the presence of a rare population of super star clusters and characterizing their basic properties in *Chapter 3*, this work provides a valuable data set to investigate (i) stellar feedback and the influence of the star clusters on the medium, (ii) the early forming stages in the evolution of these types of star clusters, (iii) how many clusters of a given mass form and what influences that hierarchy, (iv) how cluster properties relate to their interstellar medium, and (v) a broader view is that these observations can be used in conjunction with the super star clusters in other nearby galaxies to potentially inform on the types of star forming conditions that were more common when galaxies in the universe were collectively peaking in the amount of stars forming. Observations with ALMA are only just starting to uncover these star clusters and the insights into our Universe that they provide. A major take away from this thesis, and *Chapters 4 and 5* in particular, is that the physical conditions in the diffuse interstellar medium can be explored through observations of radio recombination lines in and/or against radio-bright galaxies out to vast distances. This opens the door to investigations with existing facilities like the LOFAR,

VLA, GMRT, WSRT, MeerKAT, and ASKAP. On-going surveys that are searching for atomic hydrogen 21 cm emission in populations of galaxies are great complements and can be used to make more detections of radio recombination lines. The longer arcing future is rather bright because a new telescope which is just breaking ground on construction, the Square Kilometer Array (SKA), will enable significantly-more-sensitive observations at low radio frequencies and will revolutionize (extragalactic) radio recombination lines studies, allowing for the detection of (probe of) tens of thousands of sources.

List of Publications

First Author

Emig K.L., White G., Salas P., et al. 2021, *Low-Frequency Observations of Diffuse Ionized Gas in Cygnus X*, in preparation.

Emig K.L., Bolatto A., Leroy A., et al. 2020, *Super Star Clusters in the Central Starburst of NGC 4945*, ApJ, 903, 50.

Emig K.L., Salas P., de Gasperin F., et al. 2020, *Searching for the largest bound atoms in space*, A&A, 634, 138.

Emig K.L., Salas P., de Gasperin F., et al. 2019, *The first detection of radio recombination lines at cosmological distances*, A&A, 622, 7.

Emig K.L., Lunardini C., and Windhorst R. 2015, *Do high energy astrophysical neutrinos trace star formation?*, JCAP, 12, 29.

Contributing Author

Levy R., Bolatto A., Leroy A., et al. (including **Emig K.L.**) 2021, *Outflows from Super Star Clusters in the Central Starburst of NGC253*, accepted ApJ (arxiv: 2011.05334)

de Gasperin F., Williams W., Best P., et al. (including **Emig K.L.**) 2021, *The LOFAR LBA Sky Survey I. survey description and preliminary data release*, accepted A&A (arxiv: 2102.09238)

Mandal S., Prandoni I., Hardcastle M., et al. (including **Emig K.L.**) 2021, *Extremely deep 150 MHz source counts from the LoTSS Deep Fields*, accepted A&A (arxiv: 2011.08829)

Lunardini C., Vance G., **Emig K.L.**, et al. 2019, *Are starburst galaxies a common source of high energy neutrinos and cosmic rays?*, JCAP, 063, 319

Salas P., Oonk J., **Emig K.L.**, et al. 2019, *Carbon radio recombination lines from gigahertz to megahertz frequencies towards Orion A*, A&A, 626, 70

de Gaserpin F., Dijkema T., Drabent A., et al. (including **Emig K.L.**) 2019, *Systematic effects in LOFAR data: A unified calibration strategy*, A&A, 622, 5

Shimwell T., Tasse C., Hardcastle M., et al. (including **Emig K.L.**) 2019, *The LOFAR Two-meter Sky Survey. II. First data release*, A&A, 622, 1

Salas P., Oonk J., van Weeren R., et al. (including **Emig K.L.**) 2018, *Mapping low-frequency carbon radio recombination lines towards Cassiopeia A at 340, 148, 54, and 43 MHz*, MNRAS, 475, 2496

Salas P., Oonk J., van Weeren R., et al. (including **Emig K.L.**) 2017, *LOFAR observations of decameter carbon radio recombination lines towards Cassiopeia A*, MNRAS, 467, 2274

Driver S., Andrews S., Davies L., et al. (including **Emig K.L.**) 2016, *Measurements of Extragalactic Background Light from Far UV to Far IR from Deep Ground- and Space-based Galaxy Counts*, ApJ, 827, 108

Trujillo C., Ball J., Boccas M., et al. (including **Emig K.L.**) 2013, *Altair at Gemini North: Full Sky Coverage Laser AO Correction at Visible Wavelengths*, Proceedings of the Third AO4ELT Conference, 51

Cenko S., Li W., Filippenko A., et al. (including **Emig K.L.**) 2012, *Supernova 2012Z in NGC 1309 = Psn J03220535-1523156*, CBET, 3014, 1

Curriculum Vitae

In the fullest sense of the words curriculum vitae, this is a story of my academic life. I was born on the 21st of January 1989 in Norristown, Pennsylvania in the United States of America to Bonnie Daya Emig and Ronald Emig. As a child, I was curious and fascinated to learn — learn about all types of things. How things work. Why they work the way they do. One of my grand visions: image learning and discovering the wonders of the world for your entire life! (Only later did I realize that this is an inherent part of being human.) Early on I showed an aptitude for math and eventually science as it was introduced. I loved working with numbers and the language of mathematics. Knowing the qualities and strengths that I possess now and thinking back to myself as a child, I realize how the basic concept behind the scientific method – hypothesizing, testing, assessing the extent and limitations of results – came so naturally to my mind. I was drawn to, what I perceive as, this "best way" to learn about the physical world. I must confess though, I was a bad student! The older I got, the worse it got. Instead of doing homework, you could more easily find me barefoot, climbing a tree, training in competitive gymnastics, and as I got older, coaching gymnastics, working, or being a mischievous teenager.

After high school I enrolled in the local Montgomery County Community College. It was at "MC3" that I took an introductory astronomy course with Dr. Peter Bachmann. Wow! How awe-inspiring is the beauty of the cosmos? Instead of discovering the wonders of the world, I could discover the wonders of the universe. And I could combine mathematics with scientific rigor to do it! I was hooked. Through the introductory course and the astronomical journals that Dr. Bachmann shared with me, I learned about Mauna Kea and some of the world's best observatories. I was inspired to study astronomy in that environment, and I valued and had a desire to learn about and experience the culture and traditions of native Hawaiians. In deciding to pursue astronomy, I also pledged to myself that I would be an advocate for feminism through leading by example and directly contributing (to enrich women in society) as an involved scientist.

At the University of Hawai'i at Hilo, I completed double majors in astronomy and physics and a minor in mathematics. I became very focused on school work and immersing myself in physics and astronomy, meanwhile I coached and judged gymnastics on nights and weekends. The astronomy community in Hilo and on Mauna Kea was a fantastic environment for a young astronomer in the making. I assisted in observations with the Caltech Submillimeter Observatory and the Sub-millimeter Array and toured many other facilities. I regularly contributed in public outreach activities

connected with the university and the observatories – in elementary schools, at university events, at the shopping mall, at the Visitor Information Station on Mauna Kea, and by setting up solar telescopes at beaches. With the Pacific International Space Center for Exploration Systems (PICSES), I worked with NASA and the Canadian Space Agency on simulated lunar rover missions on Mauna Kea. I gained research experience working with Prof. Marianne Takamiya on star-forming galaxies and as a part of this, observed at the Subaru Telescope. Through the Pacific Undergraduate Research Experience in Mathematics working with Profs. Rebecca and Luis Garcia, Efren Ruiz, and Roberto Pelayo, I investigated the abelian sandpile groups of book graphs. At the CTIO research experience for undergraduates in La Serena, Chile, I worked with Catherine Kaleida on simulated star clusters in Hubble Space Telescope data. During my undergraduate, I also organized an independent study on particle physics, I participated in the William Putnam mathematical competition, and I was awarded the Daniel K. Inouye Scholarship in Astronomy. I was also mentored by Prof. Philippe Binder. I graduated in May 2012, also gaining much appreciation and perspective during this time as a minority in society.

In August 2012, I started an internship at the Gemini (North) Observatory in Hilo. With Dr. Rachel Mason, I investigated low-luminosity AGN and I also helped to recommission the GNIRS instrument. I stayed at Gemini until May 2013. In that time I also worked on an instrument performance monitoring system with Drs. Michael Pohlen and Andre-Nicolas Chene, and I attended two astronomical conferences.

By August 2013, I started a graduate program in astrophysics at Arizona State University in the group of Prof. Rogier Windhorst, Dr. Seth Cohen, and Dr. Rolf Jansen. I was ecstatic to lead and publish my first academic article with Profs. Cecilia Lunardini and Rogier Windhorst on possible counterparts of high energy neutrino events from the IceCube Neutrino Observatory. Another research project at ASU led me to Germany to work (primarily) with Prof. Marcus Brüggen and Dr. Francesco de Gasperin on interferometric radio observations of merging galaxy clusters taken with the Low Frequency Array. I quickly took to calibrating radio data. With this unique and valuable experience and the bright and pioneering future I saw in the radio astronomy community, I was motivated to continue working with LOFAR. I earned a Master of Science in astrophysics in August 2015. During that time, I also was awarded the Dean's Fellowship, assisted in teaching four bachelor courses, secured funding to attend multiple conferences, co-founded a rocketry club, observed at the Kitt Peak National Observatory, served as a Graduate Council Representative peer-elected from more than 100 graduate students, and received honorable mention as a National Science Foundation Graduate Research Fellow.

My journey to earn a PhD from the Leiden Observatory with Profs. Alexander Tielens and Huub Röttgering started in September 2015. I spent the majority of my time and efforts as a PhD researcher leading the development of data processing strategies for extragalactic spectroscopic observations with LOFAR, developing techniques to search for redshifted radio recombination lines, and contributing as a team member to galactic observations of carbon radio recombination lines from diffuse molecular gas and to the LOFAR Surveys Key Science Project. My time as a visiting researcher at the University of Maryland in Fall of 2019 with Prof. Alberto Bolatto was also highly influential to me. I broadened my research base and entered

into a fruitful collaboration.

Throughout my PhD, I developed the tools to navigate academia as an independent researcher, an effective team member, an expert in scientific and technical capabilities, and a public and professional educator. I have published three first author papers in high-profile astronomical journals. I have been the PI and Co-I of dozens of observing proposals for telescopes, helped to supervise two master student projects, assisted in teaching a master course on the interstellar medium for three years, presented research in numerous international conferences, public events, and as a colloquium speaker, participated in and helped to organize technical and scientific workshops, organized public outreach events, authored a children's article, organized the weekly department borrel, and served as a graduate student ambassador. And just when I thought I was in the homestretch, I saw through the completion of two thesis chapters during the global pandemic of 2020. An important aspect of my time in Leiden is that I have also learned the experience of a foreigner in an ambiguous and sometimes insecure environment. As I have done during my PhD, I continue to educate myself about and be a proponent of equitable and oppression-free practices in (academic) organizations.

Starting in December 2020, I am honored to be a postdoctorate Jansky Fellow of the National Radio Astronomy Observatory based in Charlottesville, Virginia, USA. I look forward to honing my leadership skills and developing radio recombination lines as a powerful tool to study the interstellar medium in external galaxies. I continue to be a member of the LOFAR Surveys Key Science Project and Square Kilometer Array working groups. Throughout my academic journey, I have also persevered through multiple occasions of harassment and bullying. I thank the many individuals who have helped me through those experiences and grow in positive ways from them.

Acknowledgements

My PhD advisors have had a profound impact on the content of this thesis and my experience in becoming a doctor of astronomy. Xander, you have been such a great role model; thank you for sharing with me the multi-faceted qualities and undertakings of a great leader and professor. I am so grateful that I have been able to learn from you all these years. Thank you for valuing and supporting me and my work, especially through harder times. I will miss our weekly discussions. Alberto, since the moment I popped-in and surprised you at your office door, you have warmly welcomed me into your research groups. It is such a pleasure to work with you and the curiosity, versatility, comprehensiveness, and caringness that I see shining through you and the environment that you help to create. One of the important lessons that I have learned from you is how (and why) to hone my idealisms and keep the bigger picture in focus – a huge thank you for that. Huub, thank you for bringing me into the LOFAR group and enabling me to pursue my interests in radio astronomy. What I admire and what I try to carry through in my approach to research is the grand vision and impact-driven approach with which you lead.

There are a number of others in Leiden who have taught, mentored, and inspired me over the years; thank you Reinout van Weeren, Francesco de Gasperin, Wendy Williams, and Huib Intema. I have also had the pleasure of working with a number of amazing astronomers over the years who have directly shaped and contributed to the academic that I am today. With much gratitude, I am ever mindful of Glenn White, Peter Teuben, Rogier Windhorst, Cecilia Lunardini, Chris Groppi, Marcus Brüggen, Rachel Mason, Michael Pohlen, André-Nicolas Chené, Catherine Kaleida, Marianne Takamiya, and Philippe Binder. Remembering the partnership with German LOFAR and ASU that first introduced me to LOFAR and sparked my journey to do a PhD in Leiden, a special thank you to Marcus Brüggen, Evan Scannapieco, Chris Groppi, Judd Bowman, Ralf Dettmar, and Dominik Bomans for the time and efforts you have invested into a very successful initiative.

The Leiden RRL group: we have embarked on a pioneering RRL journey together! What a ride it has been. Thanks Pedro, Carmen, and Raymond, for sharing all that you have through out the years, especially sweets during group meetings ;-) It has been a blast to learn and share experiences with the many members of the LOFAR Group in Leiden — especially Aayush, Alex, Andrea, Duy, Edwin, Fra, Frits, Gabriella, George, Huib, Jit, Joe, Josh, Julius, Ken, Martijn, Michiel, Reinout, Tim, and Wendy. Maolin, you are missed and cherished dearly. I have also enjoyed participating as a member of the LOFAR Survey Key Science Project. Thank you all for your collective efforts over

the years. Special thanks to those whom I have come to know: Amanda, Annalisa, Carole, Federica, Franco, Gianfranco, Glenn, Marcus, Mark, Pratik, and Raffaella. I would also like to acknowledge those in ASTRON who have made the research in this thesis possible, especially the LOFAR telescope operators and support scientists as well as developers, Tammo Jan Dijkema and André Offringa. The ISM group in Leiden has expanded my ISM knowledge, and I am thankful for our meetings together – Alessandra, Andrew, Cameron, Cornelia, Dario, Daniel, Helgi, Jordy, Kirstin, Liz, Marcelo, Michal, Morgan, Pablo, Sanjana, and Sascha. The Starburst Gang – Adam, Alberto, Becca, Betsy, Fabian, Maria Jesus and Nico — you expand my research horizons with every meeting, thank you!

Daily (and long-term) runnings of the Leiden observatory and crucial aspects to completing this thesis have been possible thanks to Aart Vos, Alexandra Schouten-Voskamp, Caroline de Bruin, Els Heijsman, Eric van der Kraan, Erik Deul, Evelijn Gerstel, David Jansen, Leonardo Lenoci, Liesbeth van der Veld, Marjan Balkestein, Monica Lamers, and Yvonne Kluijt — thanks for the care and help all of these years. I would also like to make a special tribute to my co-Borrelinas: Eleonora, Francisca, and Maria Cristina, to all the past and future borrel committees, and to the attendees and enjoyers of the Friday borrels — what a necessary and welcomed way to gather with colleagues! Thank you to those who I have organized with and who have appreciated the efforts of L.A.D. F. Kaiser, the Kaiser Spring Lectures, and the PhD ambassadors. I also appreciate the discussions and guidance by Benne Holwerda, Elena Rossi, Jackie Hodge, Leo Burtscher, and Matthew Kenworthy. To the amazing female scientists at the Sterrewacht, thank you for being such great role models!

To friends and colleagues in Leiden, I cherish the connections we have made. Thanks for the memorable times: Aayush, Alberto, Alex, Amanda, Andrea, Andres, Anette, Anna, Aurora, Bas, Christos, Clément, Cornelia, Cristina & Claudio, Dario, Dilovan, Eleonora, Emanuele, Francisca, Fraser, Frits, Gabby & Marco, Gabriella, Heather, Hiddo, Jit, Jordy, Jorge, Josh & Aleks, Lammim & Ruslan, Mantas, Maria Cristina, Marina, Mark, Matus, Nikita & Tahani, Omar, Pablo, Pooneh, Pratik, Sanjana, Santiago & Ylva, Stefano, Tiago, Tommaso, Turgay, Valeria & Luke, Vincenzo, and Yapeng. Thanks to everyone at Taekwondo. Best of wishes to all of those in my PhD cohort!

It was a pleasure working and spending (lunch)time with the folks at the University of Maryland, especially Alice, Andy, Becca, Elizabeth, Jialu, Laura, Lee, Marc, Mark, Maitraiyee, Peter, Ramsey, Stuart, Vicente, and the Thursday group of PhDs. Thank you to the Department of Astronomy and the Laboratory for Millimeter-wave Astronomy at the University of Maryland and to the Green Bank Observatory for hosting me during part of this thesis.

A big thank you to all of my immediate and extended family. You have been so supportive throughout my PhD – checking in on the research and progress, being there emotionally, and discussing about both what I work on and space related topics. A las familias Goetschmann, Muñoz, Ortiz, y Salas, estoy muy agradecida de todo el amor y cariño que me han entregado. My parents, Bonnie Daya Emig and Ronald Emig, thank you for the unwavering support in my endeavors and for helping me be the person I am today. Zada Gray you are a shining light and I thank you for bringing (us all) so much joy and happiness. Pedro, my friend, life partner, and

collaborator, thank you for being the person you are and for the countless ways you have contributed to the successful outcome of this thesis. A special thank you to Dad, Javi, John, Aunt Linda, Mom, Skylar, and Uncle Tim for directly contributing to the summary of this thesis!

I am thankful for the financial support from the Netherlands Organization for Scientific Research (NWO) through TOP grant 614.001.351 as well as from the Leids Kerkhoven-Bosscha Fonds and the OBELICS scholarship for extending international travel grants to me throughout my thesis work.