Universiteit
Leiden
The Netherlands

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

Emig, K.L.

Citation
Emig, K. L. (2021, April 29). *From star-formation to recombination: expanding our view of the radio recombination line universe*. Retrieved from https://hdl.handle.net/1887/3160759

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3160759

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

The handle https://hdl.handle.net/1887/3160759 holds various files of this Leiden University dissertation.

Author: Emig, K.L.
Title: From star-formation to recombination: expanding our view of the radio recombination line universe
Issue Date: 2021-04-29

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

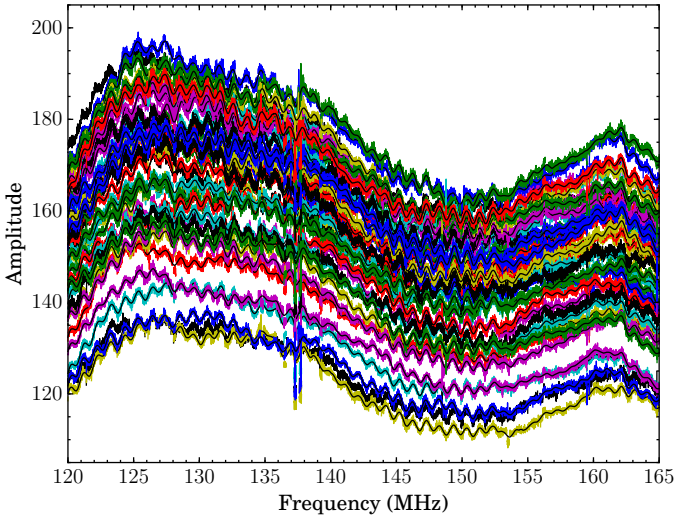

Figure 5.1: Bandpass solutions of the XX polarization towards the primary calibrator 3C196, in which stations are represented with different colors. These solutions demonstrate the global shape of the bandpass, as well a ∼1 MHz ripple which results from a standing wave. Unflagged RFI is still present 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

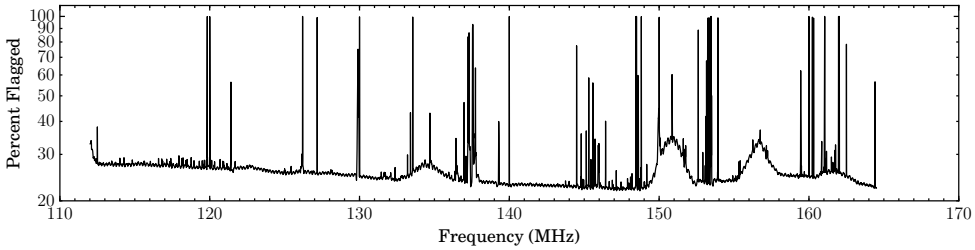

Figure 5.2: Percentage of flagged visibilities per channel in the calibrated target data. The total percentage here includes 10 remote stations and 4 core stations that have been flagged. Broad bumps centered at frequencies of about 135, 151, 157, and 162 MHz show broad-band RFI that results from intermodulation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

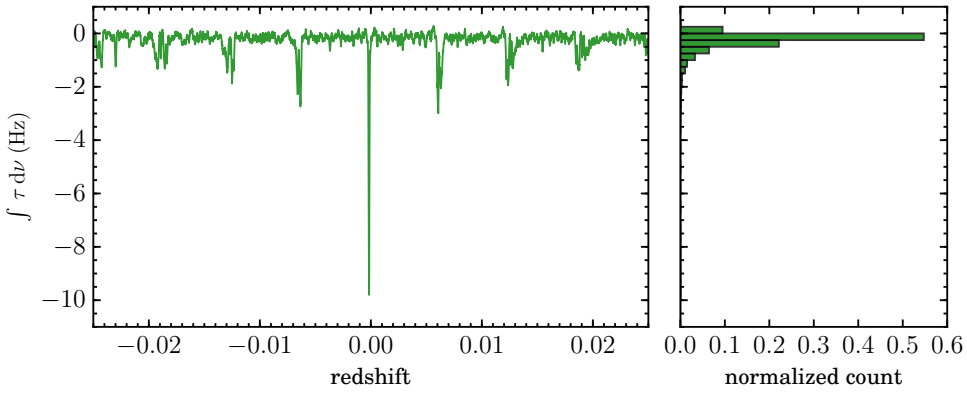

Figure 5.3: Cas A spectra has been stacked for Cα RRLs across a range of redshifts. The plot on the *left* shows the optical depth integrated at the central velocities (-7.7 km s^{-1} to +19.8 km s^{-1}, which encompasses the -47 and -38 km s^{-1} components) at each redshift. Here we clearly see an 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

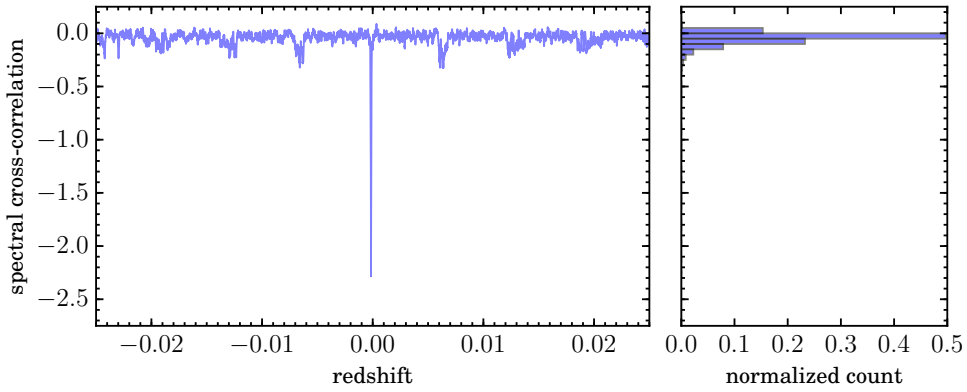

Figure 5.4: Cas A spectra have been cross-correlated for Cα RRLs with a template spectrum across a range of redshifts. The plot on the *left* shows the spectral cross-correlation value at each redshift. Again we clearly see an outlying peak at the redshift ($z = -0.000158$) of the -47 km s^{-1} component. 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

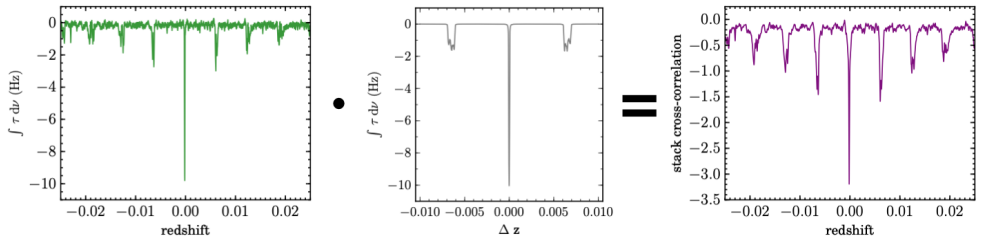

Figure 5.5: Here we demonstrate the stack cross-correlation. Cas A spectra have been stacked for Cα RRLs across a range of redshifts. The plot on the *left* shows the optical depth integrated at the central velocities at each redshift. The *middle* plot shows the integrated optical depth resulting from 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*.

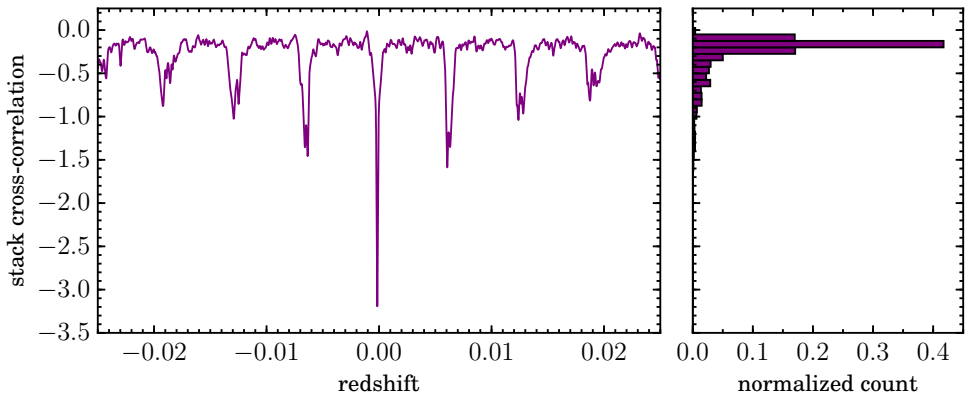

Figure 5.6: The integrated optical depth of the observed Cas A spectra is cross-correlated with the expected template spectrum. Here we clearly see an outlying peak at the redshift ($z = -0.000158$) of the -47 km s^{-1} component. The histogram on the *right* shows the binned distribution of integrated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

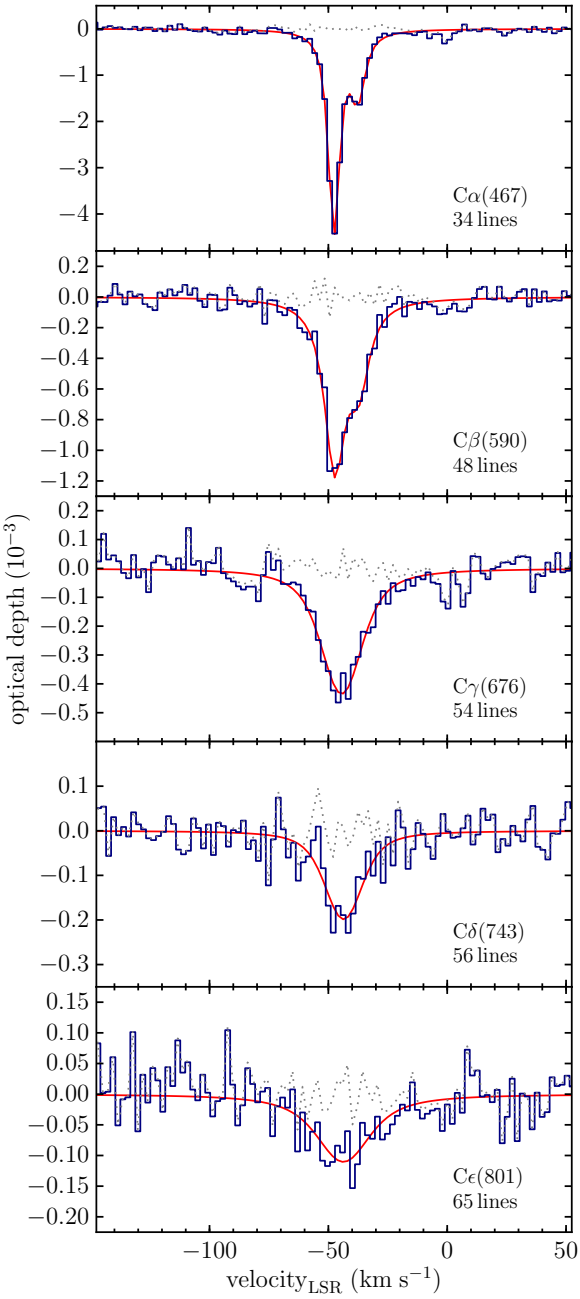

Figure 5.7: The stack averaged profiles for the carbon transitions associated with the $v_{\rm LSR} = -47$ km s^{-1} and $v_{\rm LSR} = -38$ km s^{-1} components in 55 MHz – 78 MHz observations of Cas A. These have all been significantly identified by our method.

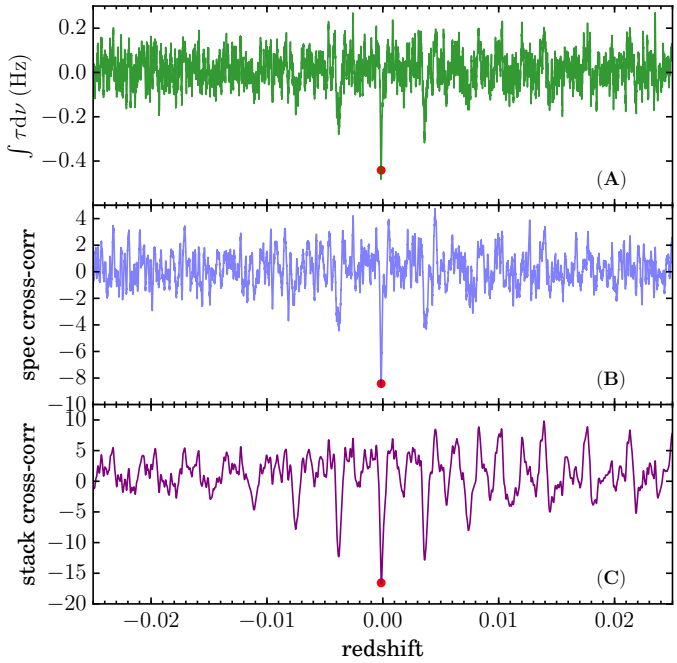

Figure 5.8: The three steps of our method applied to ϵ-transition stacking in the spectrum of Cas A: **(A)** the integrated optical depth at each redshift (see Section 5.3.1) where "mirrors" are seen at e.g. $z_m = \pm 0.004$, **(B)** the spectral cross-correlation (see Section 5.3.2), and **(C)** the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

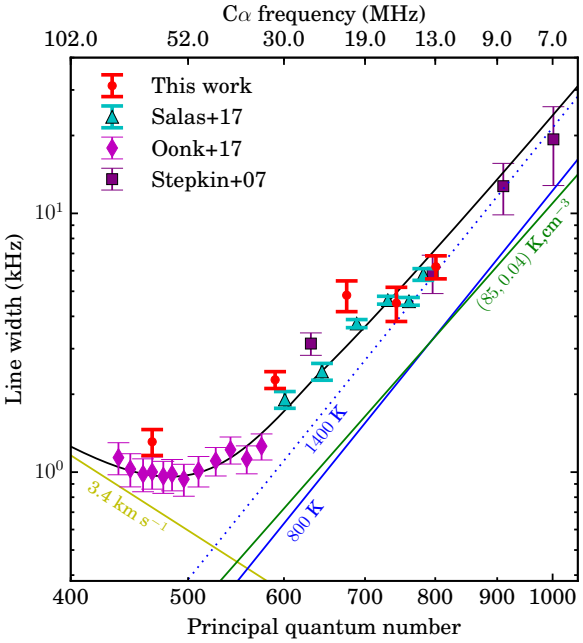

Figure 5.9: Line-width for the -47 km s^{-1} velocity component (or the sum of blended -47 and -38 km s^{-1} components) as a function of principal quantum number. The red points show the measured line-widths from this work (Table 5.1); the Cα(467) and Cβ(590) are derived only from the -47 km s^{-1} 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.

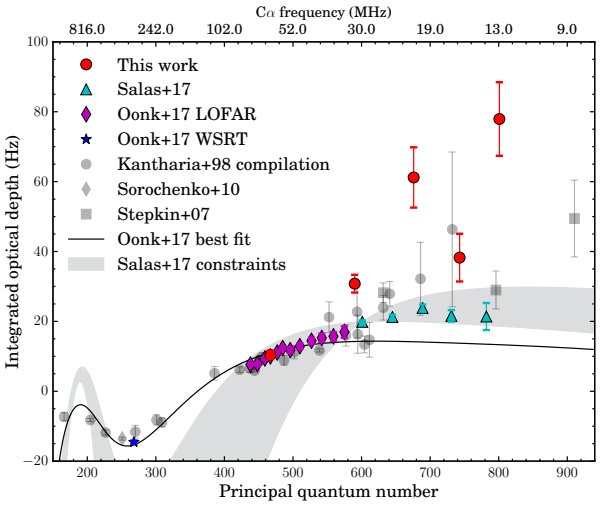

Figure 5.10: Integrated optical depth as a function of principal quantum number for the sum of the Perseus arm components at -47 and -38 km s^{-1}adapted from S17. The cyan triangles show the 10–33 MHz LOFAR observations of S17. The purple diamonds represent the 33–78 MHz LOFAR detections and the blue 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

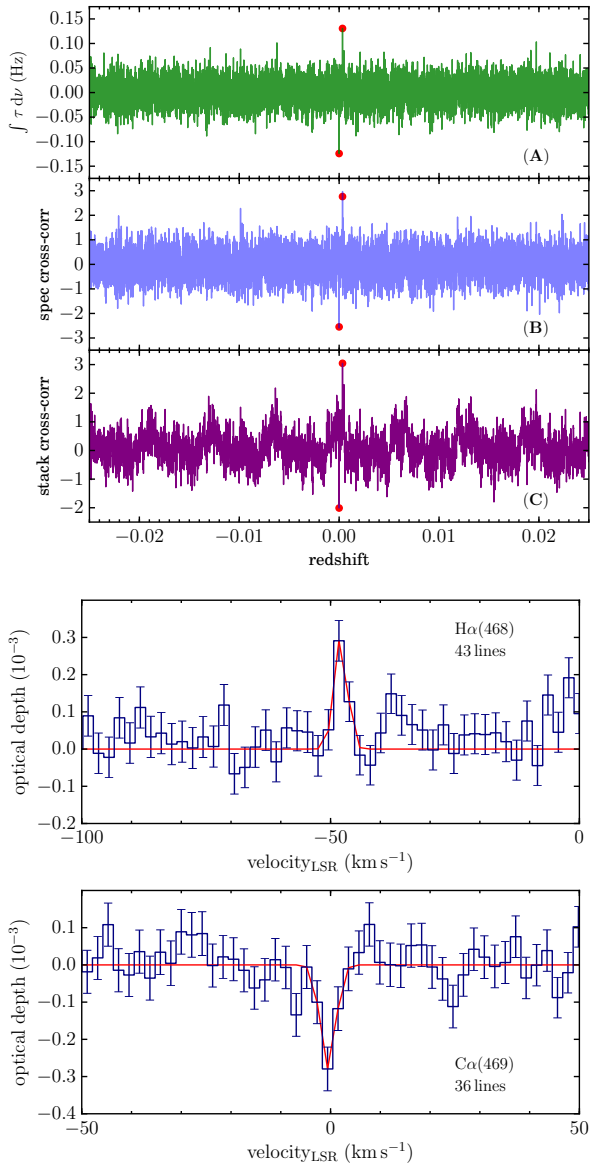

Figure 5.11: *Left:* Same as in Figure 5.8, except applied to Cα-transitions in Cas A spectra. Fits to the -47 km s^{-1} and -38 km s^{-1}components (and all higher order transitions) have been subtracted in the spectrum of Cas A. Residuals remaining from imperfect subtraction of the brightest α transitions 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

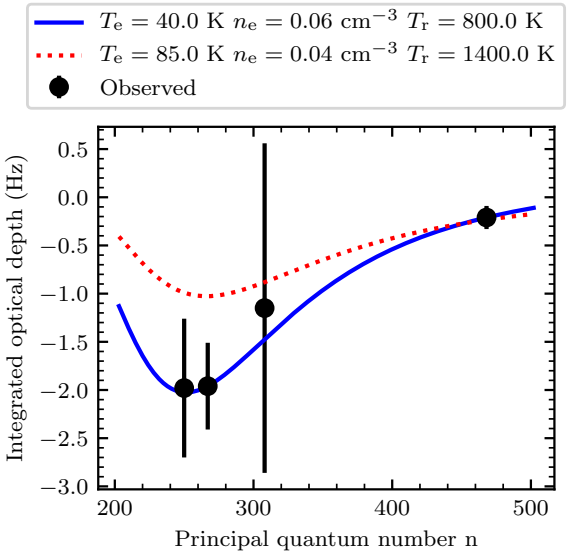

Figure 5.12: The integrated optical depth as a function of principal quantum number, n, for the Hydrogen RRLs detections associated with the -47 km s^{-1} velocity component towards Cas A. The figure shows the values from Table 5.2 using 3σ error bars. The blue, solid line shows the best fit model. The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

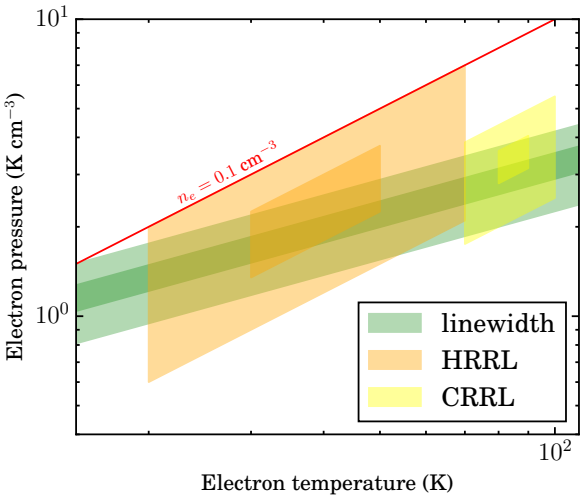

Figure 5.13: The electron pressure as a function of electron temperature of the -47 km s^{-1} velocity component towards Cas A. The red solid line shows the upper limit of a Pressure broadened line-width. Shaded regions represent 1σ and 3σ uncertainty. The green shaded regions show the line-width constraints 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.

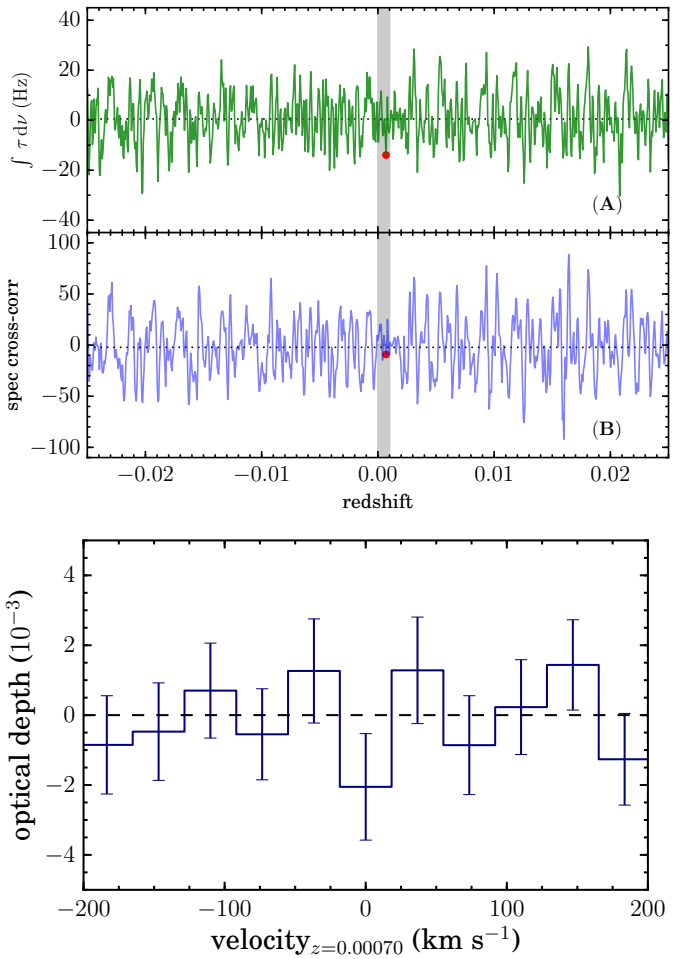

Figure 5.14: Our results for M 82. *Left:* Same as in Figure 5.8, except applied to the Cα-transitions of M 82. We did not proceed with the stack cross-correlation because no significant outlier was identified in (B). The shaded region in gray shows the redshift range probed by M14. The red data point 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

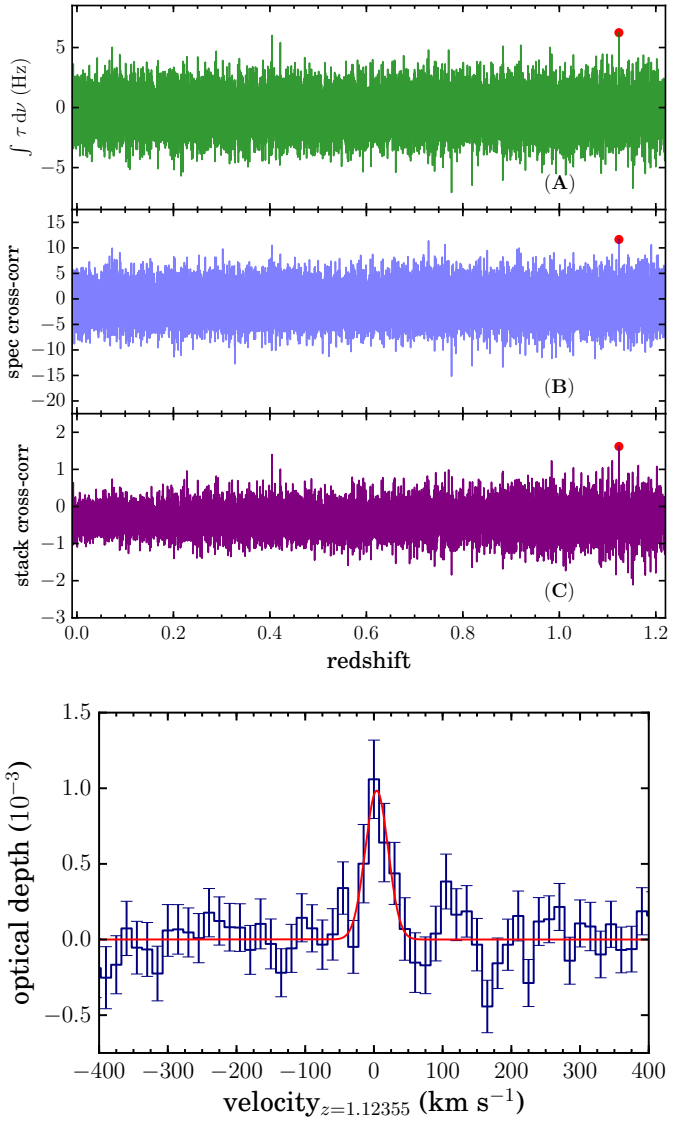

Figure 5.15: Our results for 3C 190. *Left:* Same as in Figure 5.8 except applied to the Cα-transitions of 3C 190. The red data point marks the significantly identified redshift of $z = 1.12355$. *Right:* The spectrum resulting from stacking Cα-transitions at $z = 1.12355$ in 3C 190, using a line-blanking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

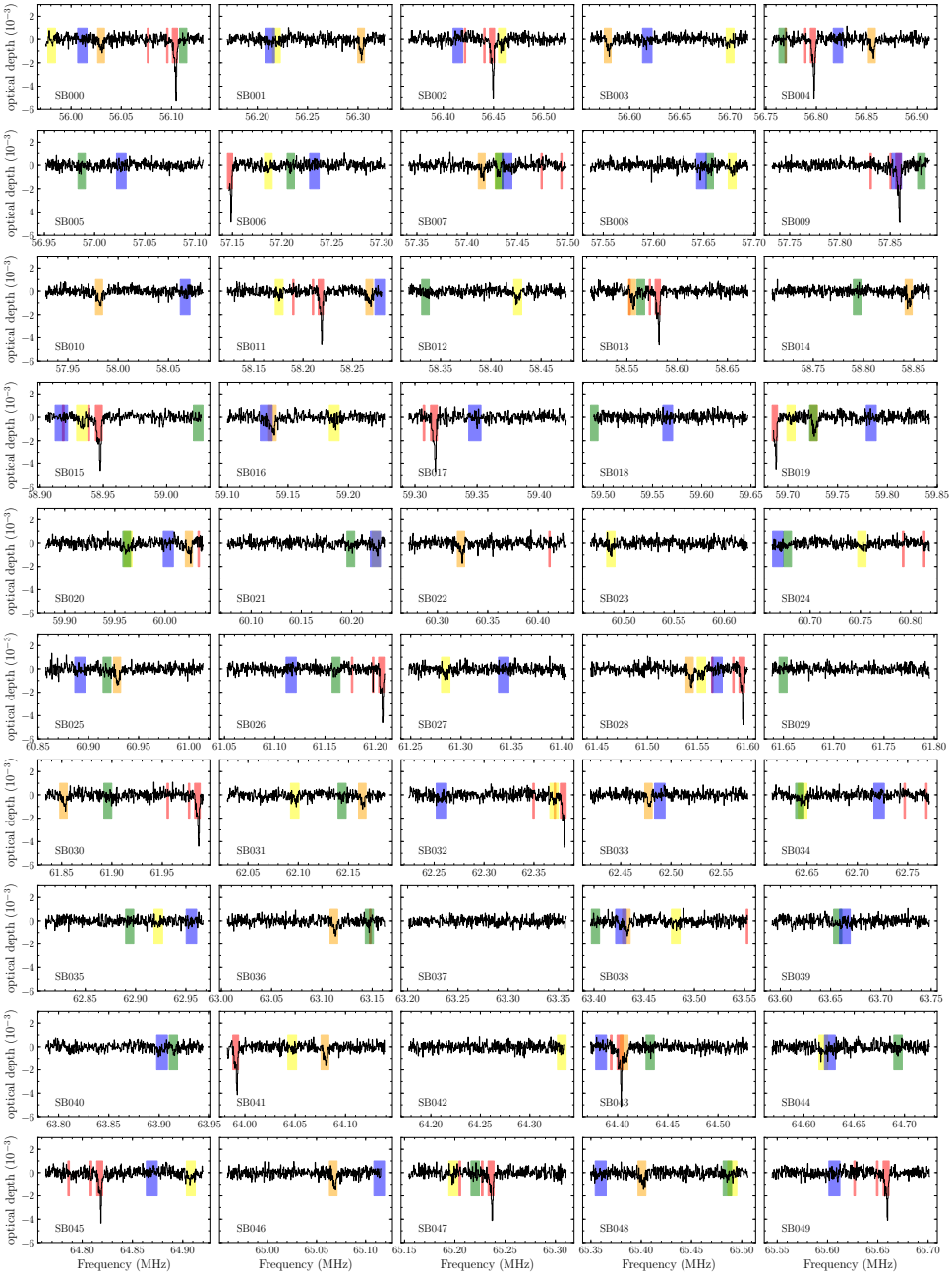

Figure 5.16: *(cont.)*

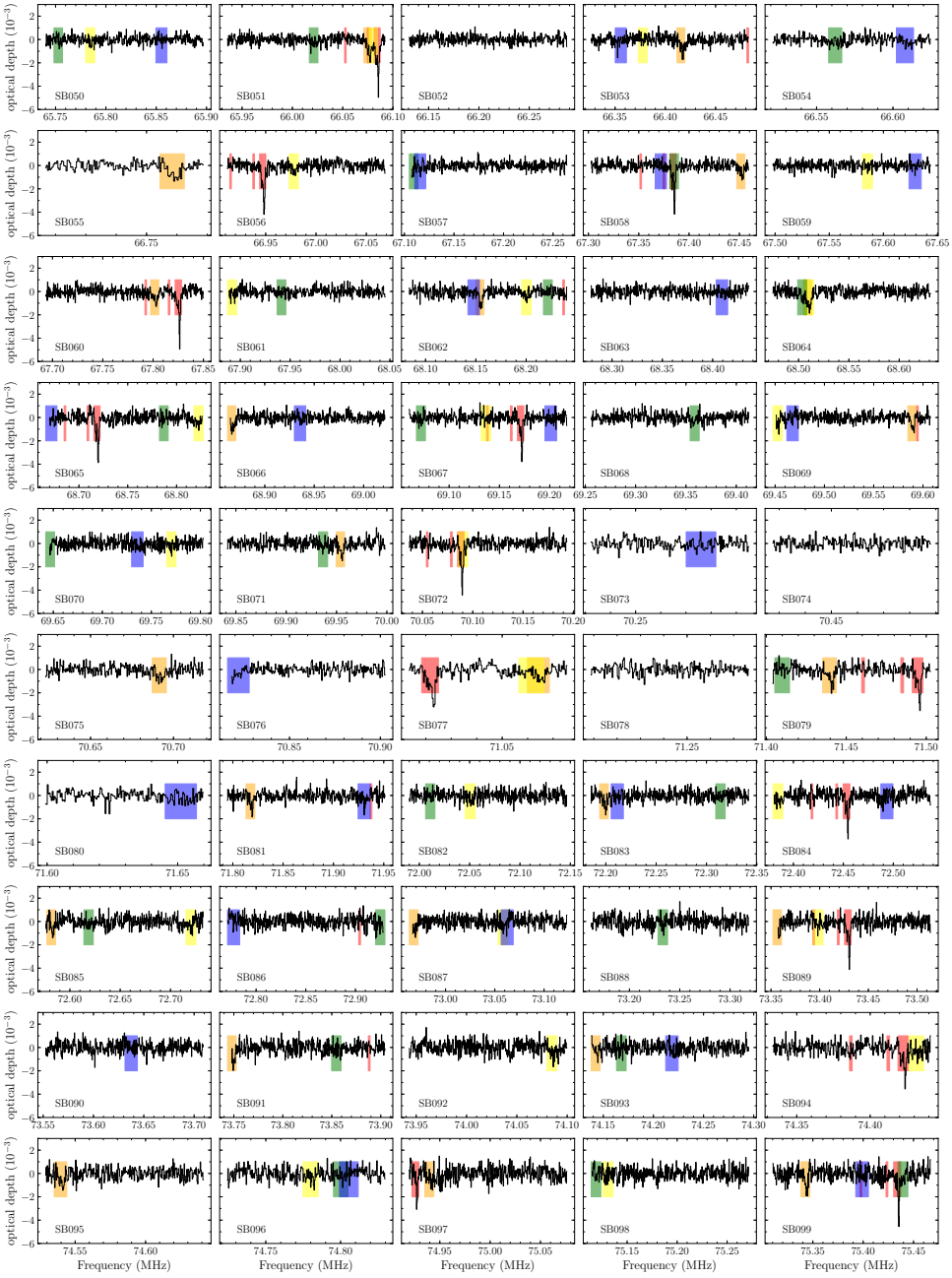

Figure 5.16: *(cont.)*

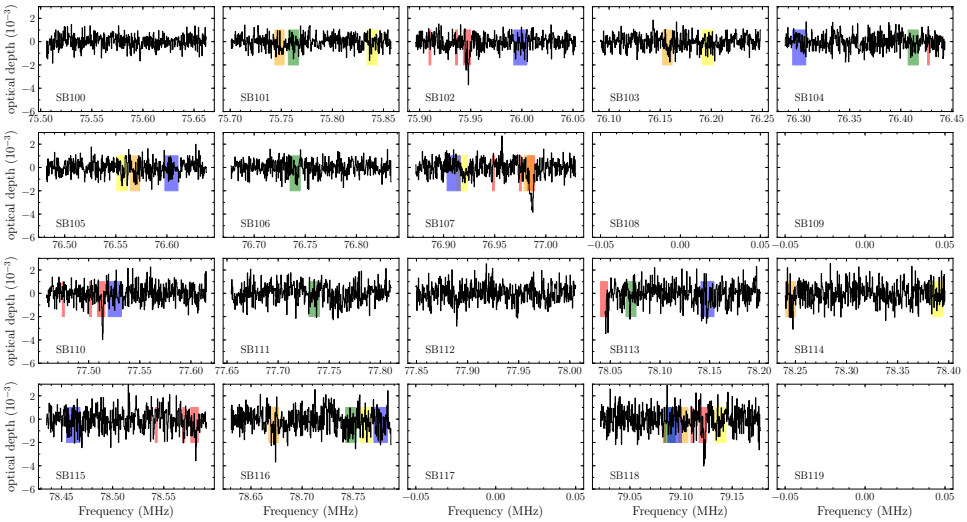

Figure 5.16: The subband spectra of Cas A, shown after continuum removal, as the optical depth as a function of frequency. The red shaded regions cover the location of α-transitions, orange regions cover β-transitions, yellow regions cover γ-transitions, green regions cover δ-transitions, and blue regions 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

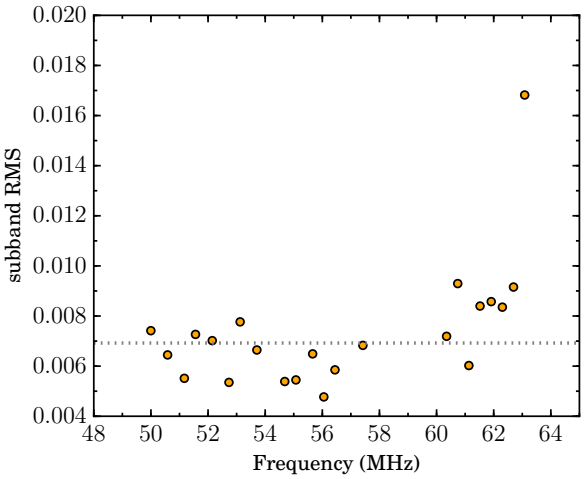

Figure 5.17: The RMS of each M 82 subband plotted against the average frequency of the subband. The gray dotted lines represents the median value of 6.9×10^{-3}. These subbands were flagged and processed following the criteria of M14. The RMS was determined from line-free channels only.

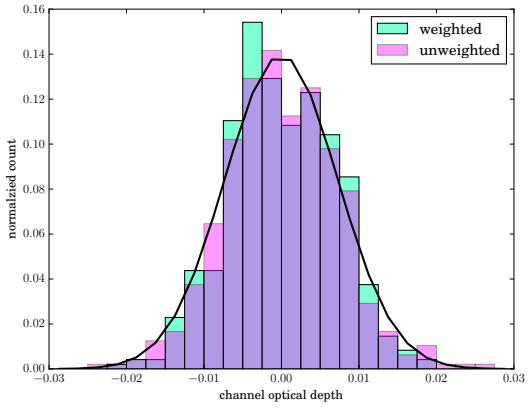

Figure 5.18: Histogram of the optical depth of line-free channels, shown before (pink) and after (green) weighting with their overlap in purple. The black solid line shows the Gaussian fit to the weighted subbands. Unweighted, the fit to the distribution has a mean of $\mu_{uw} = -1.4 \times 10^{-3}$ 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.

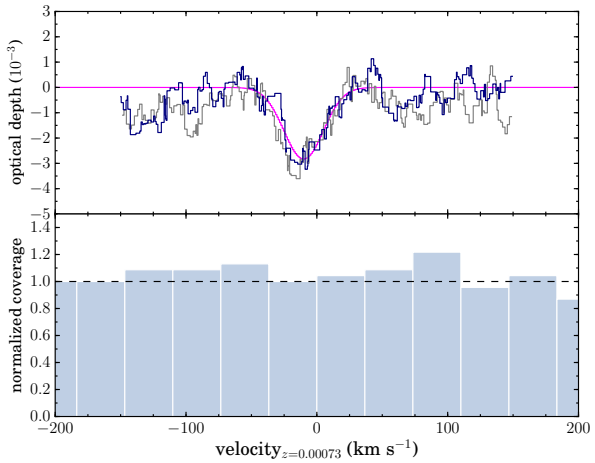

Figure 5.19: The top plot shows the optical depth as a function of velocity for the RRL spectrum of M 82 stacked at $z_{\rm M82} = 0.00073$ and smoothed with a Savitzky-Golay filter. In navy blue is the spectrum we reproduced following the criteria of M14; the magenta line shows the best fit Gaussian. 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.

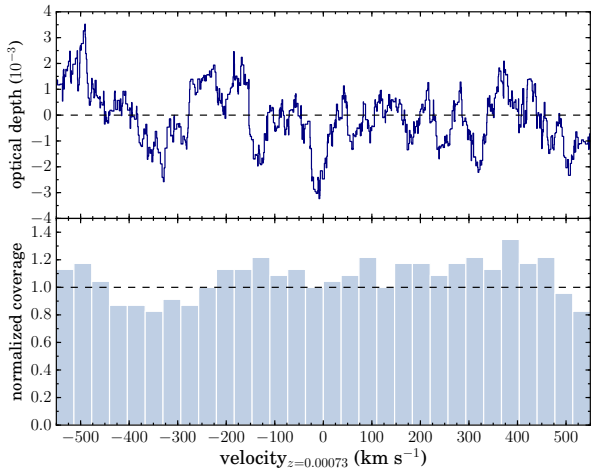

Figure 5.20: Optical depth as a function of velocity, same as in Figure 5.19, except we have extended the coverage to larger velocities by including subbands of M 82 flanking those which contain a line. The standard deviation in the line-free channels after correcting for non-uniform coverage is 1.3×10^{-3}.

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.

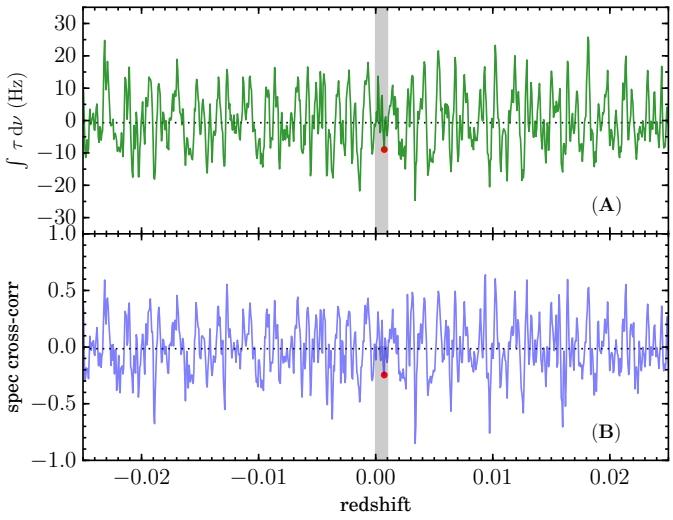

Figure 5.21: Same as in Figure 5.14 where we have implemented a stacking and spectral cross-correlation across a range of redshifts, except in this case the spectra have been combined and smoothed via a Savitzky-Golay filter. We mimicked the procedure M14 used to flag, process, and smooth the data of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