21 cm line of atomic hydrogen

J. W. V. Storey, M. C. B. Ashley, M. Naray, and J. P. Lloyd

School of Physics, University of New South Wales, Sydney 2052, Australia

(Received 24 January 1994; accepted 20 May 1994)

Discovered in 1951, the 21 cm hyperfine line of atomic hydrogen has become one of the best-studied spectral features in radioastronomy. Used to trace the distribution and velocity of atomic gas in our own and in other galaxies, the line has made an enormous contribution to our understanding of galactic structure. Now, 40 years after the line’s discovery, a low-cost receiving system can be assembled from items of consumer electronics to enable study of the line in an undergraduate laboratory.

I. INTRODUCTION

The 21 cm, or “spin-flip” line of atomic hydrogen arises from the transition between the $F=1$ and $F=0$ hyperfine structure levels of the ground state, $^{1}S_{1/2}$. The energy of the two hyperfine states differs slightly, owing to the interaction between the electron spin and the nuclear spin. The frequency of the transition can be calculated in terms of fundamental constants.\(^1\) Account must be taken of the anomalous electron spin $g$ factor (whose existence was first realized in 1947 when measurements of the 21 cm line frequency did not agree with theory; this led to the development of QED). When this is done, the largest uncertainty is in the value of the fine structure constant, $\alpha$. Comparison with the measured frequency of the line, which is known to exceptional precision, is in fact then used to derive the accepted value of $\alpha$.

The suggestion that the 21 cm line of atomic hydrogen might be detectable in interstellar space was first made by van de Hulst\(^2\) in 1945, in occupied Holland. Soon after the war, searches were initiated in both Holland and the U.S. Unfortunately for the Dutch team, their receiver caught fire and was destroyed, setting their work back by several months.

The line was first detected in the interstellar medium by Ewen and Purcell\(^3\) at Harvard University. Wishing to allow others the opportunity to confirm the discovery before publication, they immediately informed colleagues in Holland and Australia. Both groups made successful detections within a few weeks, with the Dutch paper\(^4\) appearing as a companion to the original Nature discovery paper, and the Australian detection as a footnote. This early history of the 21 cm line, which represents an interesting insight into how science was conducted in gentler times, is described, for example, by Kerr\(^5\) and Sullivan.\(^6\) The original horn antenna used by Ewen and Purcell is currently on display at the National Radio Astronomy Observatory, Greenbank, WVA.

The accepted laboratory frequency of the line, as measured with hydrogen masers, is

$$\nu = 1 \, 420 \, 405 \, 751.786(30) \, \text{Hz}.$$  

“Laboratory” is perhaps a misleading term in this context. The first precise measurements\(^7\) were actually carried out at the Seventh International Conference on Chronometry, Switzerland in 1964, where representatives from Varian were exhibiting two hydrogen masers, and representatives from Hewlett-Packard were exhibiting two cesium-beam frequency standards. After the conference the masers were taken to a Swiss national laboratory, where a further measurement\(^8\) against a cesium-beam standard confirmed the “Conference” measurement. (Both values are in agreement with the currently accepted value.) Again we see a fine example of scientific cooperation, this time between rival manufacturers of laboratory equipment.

The hydrogen maser is often used nowadays as a frequency standard, especially where exceptional short-term frequency stability is required. Laboratory masers have exhibited drifts of less than one part in $10^{13}$ over a period of 12 h.

Meanwhile, virtually every cm-wave radiotelescope in the world has made study of the line a major part of its research program. A very readable account of contemporary 21 cm studies is available by Dickey and Lockman.\(^9\)

Traditionally, undergraduate teaching laboratories have not made much use of radio astronomy in their programs. This is because construction of a radio telescope, particularly one suited for spectral line studies, has been a major electrical engineering undertaking. However, recent developments in consumer electronics have now made the task much more straightforward. Three factors in particular can be taken advantage of as follows.

1. Large (4 m diam) parabolic dishes intended for reception of direct-broadcast TV satellites are readily available at modest cost. Designed for use at 4 GHz and higher, these dishes have more than adequate surface accuracy for use at 21 cm.

2. Computer-controlled, general purpose communications receivers, with crystal-locked frequency synthesizers and a frequency coverage which includes 1.4 GHz, can be purchased inexpensively “off-the-shelf.”

3. By a happy coincidence, the intermediate frequency of TV satellite systems (i.e., the frequency to which the incoming microwave signal is converted before being sent down the cable to the receiver from the “block down converter” at the dish itself) is 950–1450 MHz. Moderately low-noise amplifiers covering this frequency range, and intended for use as line-boosting amplifiers, are thus readily available.

To put together a complete radiotelescope for study of the 21 cm line requires the above items plus a 1.4 GHz feed horn, a computer to scan the receiver and record the signal strength, and, if possible, a very low noise preamplifier. It is interesting to compare such a receiving system to that used by Ewen and Purcell; see Table I.

For a point source (or one small compared to the antenna beam), the integration time required to achieve a given signal/noise ratio goes as the square of the receiver temperature, inversely with the number of frequency channels available, and as the inverse square of the antenna effective aperture. We should thus require only $10^{-3}$ of the observing time needed by Ewen and Purcell to see the line—in other words, the line is immediately visible.
Table I. Comparison of the telescope used to originally discover the 21 cm line (Harvard 1951), with the present undergraduate experiment (UNSW 1993).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Harvard (1951)</th>
<th>UNSW (1993)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antenna</td>
<td>Horn</td>
<td>Parabolic dish</td>
</tr>
<tr>
<td>Effective aperture (m²)</td>
<td>~1</td>
<td>~5</td>
</tr>
<tr>
<td>Beamwidth (deg)</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>Receiver Temperature (K)</td>
<td>3400</td>
<td>50</td>
</tr>
<tr>
<td>Frequency resolution (kHz)</td>
<td>17</td>
<td>6, 15, 150</td>
</tr>
<tr>
<td>Number of channels</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

II. ANTENNA AND FEED HORN

The antenna dish itself is a 3.7-m-diam spun aluminium paraboloid, with a focal ratio \((f/d)\) of about 0.3, marketed in Australia for reception of the Australian television satellite Aussat at 12 GHz. We chose this particular dish because it works well at high frequency, and other radio astronomy experiments are planned, e.g., measurements of the surface temperature of the moon and of the brightness temperature of the sun at 12 GHz.\(^{11}\) (A similar experiment, involving measurement of the sun at 4 GHz, has been described previously.\(^{12,13}\) However, if operation at 21 cm only is required, a much cruder surface would suffice. A mesh surface, for example, is much lighter and has a lower wind loading than a solid surface. Antenna dishes made for reception of Intelsat at 4 GHz would be ideal. Alternatively, a dish can be constructed using aluminium tubing and wire mesh; see, e.g., the ARRL Antenna Handbook.\(^{13}\)

A diameter of approximately 4 m offers a reasonable beam size without being too unwieldy. At 21 cm the full width at half-maximum beamwidth of a 4 m dish is \(\lambda/d\approx3^\circ\), which is approximately the width of the Galactic Plane. Such a beam size is therefore almost optimum for detection and mapping of the plane itself. Since the earth turns on its axis once every 24 h, the sky moves past at a rate of 15 arcmin per min of time (at zero declination; slower by the cosine of the declination at other declinations). It thus takes about 12 min for a point source to move through the beam, and manual tracking of the telescope is therefore straightforward.

The parabolic dish reflects the incoming 21 cm radiation back up to a feedhorn placed at its focus [see Fig. 1(b)]. Typical designs for such a feedhorn are presented in the ARRL Antenna Handbook.\(^{13}\) A particularly simple and perfectly satisfactory approach is to use a short length of 6 in. (150-mm-diam) copper water pipe as a circular waveguide. Signal is coupled out of the waveguide by a quarter-wavelength probe, and the back of the waveguide is sealed with a shorting end cap. The beamwidth of the feed horn is determined by the diameter of the waveguide; this beamwidth determines the coupling between the parabolic reflector and the feedhorn and ultimately the beam profile of the complete antenna.\(^{10}\)

III. PREAMPLIFIER

To minimize signal loss in the transmission line from the antenna to the observing room, it is highly desirable to place the preamplifier as close to the feedhorn as possible—on it, for example. Commercial low-noise amplifiers can be purchased for a few hundred dollars, or one could be constructed using inexpensive microwave integrated circuits. Better performance, however, can be achieved with a purpose-built amplifier using GaAs field-effect transistors (FETs) or HEMTs. We built a three-stage amplifier using the circuit of Williams et al.,\(^{14}\) and achieved a noise temperature of about 50 K with MGF1402 FETS. With modern HEMTs it should be possible to reach 30 K, even with a room temperature system.

The coaxial cable provided by suppliers of domestic satellite TV systems is designed to carry the 950–1450 MHz intermediate frequency of such systems with relatively low loss (typically 20 dB/100 m). It is therefore perfectly suitable for our current purposes. However, if a long cable run is required between the antenna and the laboratory, it may be necessary to add an additional preamplifier—preferably at the antenna end of the cable. Suitable amplifiers, with a gain of 20 dB and covering the 950–1450 MHz intermediate frequency range used by satellite TV receivers, are available commercially and are inexpensive. The noise temperature is of order 1000 K, which is adequate for our purposes. Such amplifiers typically require their power supply voltage to be provided on the coaxial cable center conductor; this is straightforward to arrange with a small 24 V laboratory power supply and a bias \(T\).

In principle, it may be possible to detect the line without any preamplifier at all. However, the quoted sensitivity of the Icom receiver used corresponds to a noise temperature of about 1200 K, so much poorer signal-to-noise ratios would be obtained.

IV. RECEIVER

The receiver used in this experiment is an R-7000, manufactured by Icom. Other, similar receivers would no doubt be equally satisfactory. The receiver frequency is controlled by an internal digital frequency synthesizer, and can be set remotely via commands from a personal computer. All the oscillators in the synthesizer chain are crystal locked, giving a
frequency stability over the temperature range 0–50 °C of the order of ±10 parts per million, or ±14 kHz at 1.4 GHz. In an air-conditioned laboratory the drift is very much smaller than this (as can be verified by tuning the receiver in SSB mode to a uhf television station and listening to the beat note against the TV vision carrier), and is insignificant compared to the expected linewidths.

The spectrum of the hydrogen line is determined by scanning the receiver step by step across the line frequency, and measuring the signal strength at each frequency. To do this, the receiver must be opened up and a suitable point identified which carries a voltage proportional to signal strength. We did this by tapping into the AGC (automatic gain control) line of the receiver, buffering this voltage, then feeding it via a 12-bit A/D converter to the computer. (With a sufficiently low-noise preamplifier, it is also actually possible to see the hydrogen line as a deflection on the signal-strength meter of the receiver!)

The AGC voltage is a rather nonlinear function of signal strength. Optimum signal-to-noise ratio is obtained when a square-law detector is used. However, with certain simplifying assumptions about the character of the noise present within the receiver, it is possible to simulate the use of a square-law detector by simply converting the measured AGC voltage to an equivalent signal power before any averaging is carried out. This is done with the help of a look-up table, created with the aid of a calibrated signal generator.

The frequency resolution, or width of each received frequency “channel,” is determined by the intermediate frequency filters of the receiver. In the case of the Icom this filter width can be selected (under computer control) by selecting the receiving mode (i.e., SSB, AM, AM wide, or FM wide) of the receiver. Depending on the source, HI clouds within our own Galaxy have linewidths ranging from 10 to 600 km/s. The FM-wide filter is thus generally the most suitable, with a full width at −6 dB of 150 kHz (corresponding to a velocity resolution at 1.4 GHz of Δν = cΔν/ν = 32 km/s).

V. SOFTWARE

The control computer is an 80386-based personal computer operating under MS-DOS. A 12-bit A/D converter occupies one slot, and accepts the signal-strength voltage from the receiver. Control of the receiver frequency is via the serial port. For the Icom receiver, the RS-232 voltages must be converted to levels required by Icom’s proprietary “Communications Interface V” (CI-V), with a MAX-232 integrated circuit. A suitable bus arbitration protocol must also be implemented on the PC to allow the correct hand shaking and packet collision detection to occur.

The computer first selects the appropriate receiving mode (which determines the spectral resolution), then tunes the receiver to the first frequency “channel.” A delay is then implemented to allow time for the synthesizer to settle and for the voltage on the signal-strength (AGC) line to stabilize. The A/D converter is then sampled at a rate fast compared to the AGC time constant (typically about 200 ms). These data are averaged together to form the first data point. The sampling is then stopped, the receiver stepped to the next frequency channel and, after a suitable delay, sampling for the second data point commences. This procedure is repeated until all frequency channels (typically about 16) have been scanned, after which the whole process is repeated for as long as required. Meanwhile, the spectrum can be displayed on the screen with suitable scaling and error bars.

VI. SPECIAL PROBLEMS

While radioastronomical observatories are usually placed at remote sites chosen for their low level of radio interference, universities are typically placed in an urban environment. The University of New South Wales is one such example; it is located just 5 km from Sydney airport where a 2 million W radar operates at a frequency of 1.35 GHz. As expected, the radar obliterates reception at 1.4 GHz as the beam sweeps through our location, regardless of where the 3.7 m dish is pointing.

Fortunately, the radar beam sweeps quickly by, causing interference for only about 1 s in every 11.48 s scan of the sky. While it might be possible to construct a microwave filter (using either cavity or stripline techniques), the wide-band nature of radar signals and the sheer magnitude of the interference discourage this approach. Instead, we simply ignore the data collected during the brief time that the radar beam sweeps by. This is done in software by means of a “sigma clipping” technique, whereby data points which differ from their neighbors by more than a predetermined amount are discarded. Although this is an inefficient technique in terms of utilizing all the available information, it is an effective and easy-to-implement method of rejecting interference.

VII. CALIBRATION

The antenna can be calibrated by pointing it at a point-source signal of known flux, $S$. This could either be an astronomical source, an artificial satellite, or a local source. If a local source is used, it must be placed in the far field of the antenna, i.e., at a distance equal to or greater than about $2D^2/\lambda$ (where $D$ is the diameter of the dish), or about 150 m for a 4 m dish. This measurement is all that is required to relate the receiver output (in A/D units or “chart-recorder deflection”) to the received flux in W/m².

However, to properly characterize our antenna, we need also know its effective aperture, $A_e$, and the size and shape of the main beam. These measurements are also performed with a point source.

For a horn antenna, the effective aperture is readily calculated from the physical aperture of the mouth of the horn. (For this reason, horn antennas are often used for calibration purposes.) For other antennas, such as a parabolic dish, the calculation is less straightforward. $A_e$ is related to the physical aperture, $A_p$ ($A_p = \pi D^2/4$), by

$$A_e = \eta A_p,$$

where $\eta$ is the aperture efficiency of the telescope.

Measurement of the effective aperture of the telescope (or, equivalently, of the aperture efficiency, $\eta$) can be done by measuring the gain, $G$, of the dish. $G$ is simply the ratio of the signal received by the dish to that received by an antenna with an isotropic response. $A_e$ can then be derived from

$$A_e = \lambda^2G/4\pi k_0,$$

where $k_0$, the ohmic loss factor, can be assumed equal to unity.

Alternatively, for an antenna constructed as described in the text, it is probably sufficient to simply assume a value of approximately 50% for the aperture efficiency.
The signal received by a radiotelescope is usually described by the antenna temperature, $T_A$, \[^{10}\] ($T_A$ is simply equal to the power collected by the antenna divided by Boltzmann’s constant, $k$, provided only that $h \nu \ll kT_A$.) The signal received from a point source of brightness $S$ can thus be expressed in terms of antenna temperature via the expression

$$T_A = A_s S / 2k,$$

where $k$ is Boltzmann’s constant, and the factor of one-half arises because the receiver is sensitive to only one of the two possible polarizations. To convert antenna temperature into source brightness temperature, $T_B$, requires knowledge of the antenna pattern, $P(\theta, \phi)$. This can be determined by mapping a bright point source. The sun is a particularly convenient source: it is very bright (at least $10^5 \text{ K}$ at this frequency), it is easy to find, and it casts a nice shadow of the feedhorn onto the vertex of the dish to confirm the pointing. Figure 2 shows a drift scan of the sun, obtained by pointing the telescope a little ahead of the sun’s expected position in the sky, and allowing the sun to drift through the beam. The time taken for this to happen ($13 \pm 1 \text{ min}$ between half-power points) allows us to derive a telescope beamwidth (full width at half-maximum) of $3.25^\circ \pm 0.25^\circ$. Unfortunately the sun is less satisfactory as a flux calibrator, as its brightness can increase by more than an order of magnitude during periods of high solar activity.

Relating the observed antenna temperature, $T_A$, to the source brightness temperature distribution, $T_B$, can then be done by inverting the expression

$$T_A = A_e / \lambda^2 \int \int P(\theta, \phi) P(\theta, \phi) d\Omega,$$

which represents the convolution of the brightness distribution of the source with the normalized antenna pattern.

Since the 21 cm atomic hydrogen hyperfine radiation is that of a magnetic dipole with a dipole matrix element of 1 bohr magneton, the transition probability, or Einstein $A$ coefficient, can easily be derived as

$$A = 2.87 \times 10^{-15} \text{ s}^{-1}.$$

The mean time, $t$, for a hydrogen atom to spontaneously emit 21 cm radiation is therefore $t = 1/A = 10^7 \text{ yr}$. Since, at normal interstellar hydrogen densities, collisions between atoms occur much more frequently than this, the relative population of the hydrogen atoms in the upper and lower states of the transition can be described by a spin temperature, $T_s$, which is in general close to the kinetic temperature, $T_k$, of the gas.

With certain simplifying assumptions, we can then relate the observed line intensity to the column density, $N_H$, of hydrogen atoms per cm$^2$ along the line of sight. For example, if the emission is optically thin, the column density is independent of the spin temperature and is given by

$$N_H = 1.82 \times 10^{18} \int T_B d\nu,$$

where $\int T_B d\nu$ is the integrated line intensity, with $\nu$ in km/s, $T_B$ in K.

**VIII. RESULTS**

A complete survey of the sky between $+42^\circ$ and $-90^\circ$ was carried out in the early sixties with a $2.2^\circ$ beam.\[^{15,16}\] Because this beamsize is similar to ours, these papers form an excellent guide as to where the regions of strongest $\text{H I}$ emission are to be found.

Figure 3 shows the first profile recorded with our system. These data took less than a minute to acquire, despite the fact that our observing efficiency was low because of the algorithms used to avoid the radar interference (Sec. VI). The region observed ($\text{R.A.} = 5^\circ45', \text{Dec} = -25^\circ$) is in the Galactic plane, close to the area of peak emission seen by McGee and Murray\[^{17}\] in the Taurus–Orion neutral hydrogen complex. The antenna temperature of 60 K observed for this region by those authors can be used to give a rough indication of our sensitivity.

**IX. FUTURE POSSIBILITIES**

While step-by-step scanning of the receiver is a simple and effective way to plot out the spectral line shape (and indeed is the technique used by Ewen and Purcell), it is clearly inefficient to observe only one frequency channel at a time while ignoring all the rest. A frequency analyzer, or "backend", which simultaneously looks at $n$ spectral channels reduces by a factor of $n$ the time required to achieve a given signal-to-noise ratio. At a modern radio observatory,
three frequency analyzing techniques are in common use: multichannel filterbanks, digital autocorrelators, and acousto-optical spectrometers. Each could be reproduced for the undergraduate laboratory, and each contains interesting physics in its own right. At the University of New South Wales we are currently constructing a simple 16-channel filterbank, while examining ways of implementing the other two techniques as additional experiments.

At longer integration times the accumulated signal may become dominated by systematic effects. Alternatively, the signal-to-noise ratio may fail to improve as expected with integration time because of 1/f noise. Both problems could possibly be alleviated by use of frequency chopping (as was used, in fact, by both Ewen and Purcell, and Muller and Oort). Here, observations at a particular frequency are alternated with measurements at a nearby reference frequency, and the difference signal recorded. The frequency switching is carried out as fast as the system time-constants allow. This technique, although still sometimes used, has limited applicability on a “real” radiotelescope because standing waves between the feedhorn and dish vertex result in a residual baseline ripple with a periodicity comparable to the width of a typical spectral line. With a small dish such as we are using, however, the periodicity of the baseline ripple (equal to c divided by twice the dish focal length) is about 300 MHz, wide enough compared to expected line profiles to be of no consequence.

Although detailed studies have not yet been carried out, it should be possible to map the Galaxy and derive its rotation curve using the simple equipment described above. Preliminary maps have been made by drift scanning (i.e., setting the telescope to a given elevation angle and collecting data over 24 h as the sky moves past) to demonstrate the feasibility of this. In addition, it should be possible (at least from the southern hemisphere) to detect the Magellanic Clouds.

Future plans include an extension of the operating frequency coverage to include other astrophysically important lines. In particular, the methanol maser at 12.179 GHz is exceptionally bright and would seem to offer an excellent candidate for study, especially with the recent availability of low-noise phase-locked down converters at this frequency (intended for reception of satellite-broadcast audio channels).

X. CONCLUSIONS

A radioastronomical observatory capable of observing and studying the 21 cm hyperfine line of atomic hydrogen can be readily assembled for the undergraduate laboratory from various items of consumer electronics. Studies of the line involve the student in a wide range of interesting topics, from the quantum-mechanical calculation of the line frequency, to mapping of the Galactic plane, to microwave engineering. While experiments in optical astronomy have long been popular with students, radioastronomical experiments offer several logistical advantages to laboratory organizers. Two in particular stand out: radio astronomy can be conducted during the day and, at least at 21 cm, the experiments can be carried out regardless of the weather conditions.

ACKNOWLEDGMENTS

This work was supported by a General Development Grant of the University of New South Wales. The article was written while one of the authors (J.W.V.S.) was a guest at the Max Planck Institute for Extraterrestrial Physics, supported by a grant from the Alexander von Humboldt Stiftung. We particularly thank our colleagues at the Australia Telescope (CSIRO) for their practical assistance and for stimulating discussions. We also thank Tom Puzzer for building the preamplifier, and Matt DeMaere for help with the figures.