A COURSE IN RADIO ASTRONOMY

by

H.C. VAN DE HULST

LEIDEN, 1951

" A Course in Radio Astronomy"

by H.C. van de Hulst

FOREWORD

These notes are based on a lecture course first given at Leiden Observatory in Fall, 1950, then in extended form at Harvard Observatory in Spring, 1951. The lectures were intended to make graduate students in astronomy and physics appreciate the problems and possibilities of radio astronomy and to give them a survey of its results. Observations were presented fairly completely, where possible in a condensed form, Of theoretical developments those were selected that seemed most instructive or most likely to survive.

The notes were written as assignments by the Harvard students, R.B.Dunn, F.Q.Orrall, H.J.Smith, A.Wyller, and H.Zirin; their names are mentioned at the chapter headings. Some chapters were thoroughly corrected or rewritten while others are presented with few changes. Thus no homogeneity or completeness could be achieved. This holds in particular for the references.

In a growing field like this the most important results often are not yet published. We were happy to be able to include many unpublished results. We have always tried to mark such results as unpublished, to give due credit to the authors, and to announce data as preliminary when they were not based on a manuscript submitted for publication.

The final editing was done jointly by Mr. Frank. J. Kerr of the Radiophysics Laboratory of the C.S.I.R.O. at Sydney, Australia and the author during the last weeks of our stay at Harvard Observatory. Mr. Kerr also contributed Chapter 9 on radar methods. Some footnotes and further references have been added in proof in August 1951. The result is a "semi-hasty"

publication that undoubtedly falls short of established standards in style, documentation and presentation. We hope that these defects will be excused by the students and active research workers for whom these notes are intended.

In addition to those mentioned above I wish to extend my thanks to the Astronomy Department of Harvard University for its help in many respects, to Mr. B.Davis, who did most of the typing, and to the Leiden students H. van Woerden and G.Westerhout, who helped with the big task of proofreading, correction and the drawing of figures. My far greater thanks are due to the dozens of friends and colleagues who by their conversations, letters and publications made me find my way into this new and exciting field of science.

Sterrewacht, Leiden
September 5, 1951

H.C.van de Hulst

A limited number of copies are available at cost price (about 7 guilders or $ 1.90) to students and research workers on direct request to Leiden Observatory.
Please send money only after copy has been received.

Published and mimeographed in the Netherlands

Leiden 1951

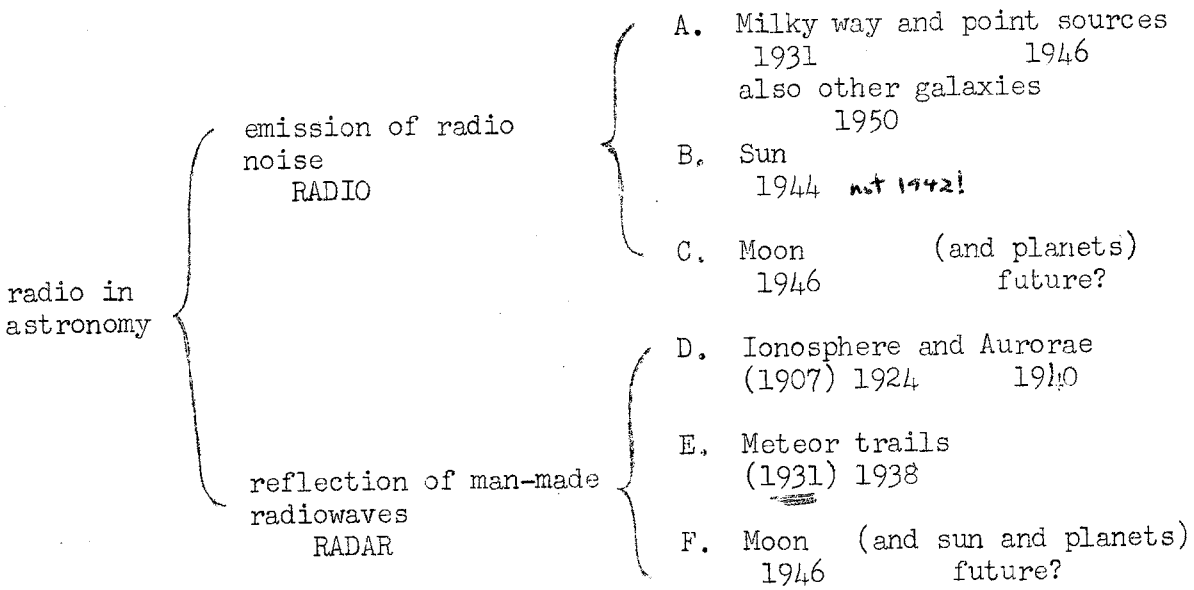
"A Course in Radio Astronomy"

CONTENTS

Chapter 1. INTRODUCTION (H. C. van de Hulst)

This course cannot be classed as an observational or theoretical course. Fortunately we do not have to. The effort to understand, that is science, always is theory in its final stage but draws heavily on observation throughout the winding ways to reach this aim. My emphasis will be on the theoretical astrophysics in radio astronomy. We shall forego the enormous problems that make observational work in this field quite an art. But we have to learn to appreciate the limitations and possibilities of the instrument in order to interpret the observations correctly.

Classification of subjects:



This table gives a survey of the problems we might cover. By choice we shall only briefly discuss the Radar or echo methods, of which D and E are geophysical subjects with astronomical implications. It is advisable not to use the word Radar (made from: radio direction and

range) for the subjects A, B, and C. Some confusion may arise from the fact that quite a few antennas now in use as radio telescopes were formerly used as radar antennas.

Importance of these investigations

Technical and economical:--These aspects I discuss first, not because they are most important on an absolute scale of values but because they can be formulated relatively easily.

The ionosphere (D) ranks first, because of its importance for world-wide radio communication. It was a big surprise when transatlantic communication with short waves was first established. Heaviside in 1907 assumed the existence of a reflecting layer. Appleton in 1924 demonstrated its existence by vertical reflections. Most important in our context is the F-layer, with average height about 300 km, and frequency cut-off for vertical incidence about 10 Mc/sec (30 meter), subject to daily, seasonal and irregular fluctuations. Moon echoes (F) are a supplementary tool to probe the ionosphere. The point sources (A) and to some extent the enhanced solar noise (B) show twinkling caused by and providing information on the ionosphere. The ionosphere is caused by ultraviolet radiation from the sun. Thus solar radiation is the prime factor entering in the complex photochemistry of the upper air. Among all signs of activity on the sun ultraviolet and corpuscular emission is one aspect and radioemission (B) another, the study of which may help to give more clarity or better predictions of activity. Also meteor trails (E) are an effective tool for studying the upper air.

Less important is solar and galactic noise as a source of interference with radio reception. Yet the discoveries bear out that they have some importance. Galactic noise was discovered in basic research on the lowest noise level

that could be attained in a receiver. Solar noise was discovered in daily radar operation during the war, when distant planes could occasionally not be detected because of solar noise.

Biological:--Several groups are investigating the effect of meter waves on plant growth. One might speculate on the possibility that the waves from solar outbursts, that are very weak even at maximum strength, may have some direct influence on plants and animals.

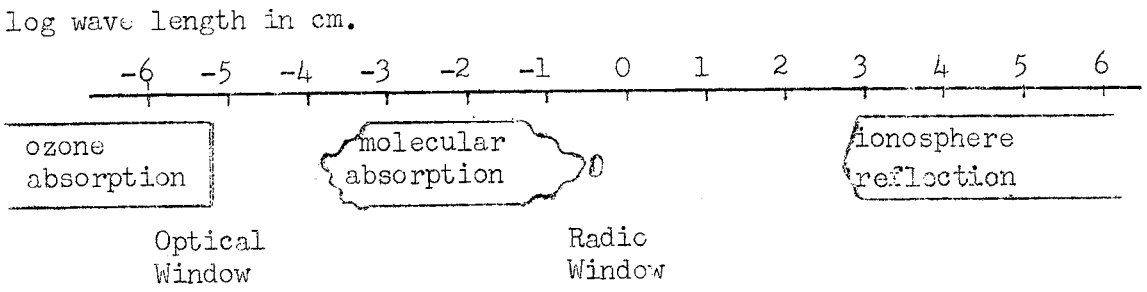
Philosophy and pure science:--Civilization may be seen as the road to airplanes, radios and what next. It may also be seen as an adventure of ideas (Whitehead). It certainly is a big adventure to try to understand the things in the universe and people of all ages have been attracted to astronomy. Quite irrespective of the results it is fascinating to see the irregular paths along which science finds its way. Here is the history of science in the making: a very young branch of science comes rapidly to full growth and is confronted with one of the oldest fields: astronomy. We do not doubt that it is the same sun and the same galaxy that we now observe with radiotelescopes. But well-known subjects may suddenly seem strange, new aspects appear and the language becomes temporarily confused. Who understands these points will also be warned against simplistic quoting of years and dates. They mark moments that a certain surprise came or a certain aim was achieved; they are not discontinuities in the history of science.

Chapter 2. FUNDAMENTALS OF OBSERVATIONAL METHODS (F. Q. Orrall)

2.1 Radiation

Most of the empirical knowledge of astronomy which we can obtain, comes from the measurement of radiation. The most complete observation of radiation which we might describe would be one giving the intensity and polarization from any point in the sky at any wave length, at any time. The various sections of observational astronomy give different emphasis to the properties of radiation measured; for example, positional astronomy is concerned with the point in the sky from which the radiation comes, spectroscopy with the wave length of the radiation, and photometry with its intensity. Since a strict separation of the sections of astronomy does not exist, we shall compare optical astronomy with radio astronomy in the light of their ability to measure the properties of radiation and see where the emphasis of radio astronomy will be.

Wave length and frequency. We cannot detect radiation from outer space of all wave lengths because of absorption or reflection in the earth's atmosphere. We have, however, two "windows" in the earth's atmosphere, through which radiation from outside the earth can reach us. One of these is the one with which optical astronomy is concerned. This extends from 3,000 Å to 30,000 Å (±). Radiation of wave lengths shorter than 3,000 Å is absorbed by ozone in the atmosphere, and that in the far infra-red by molecular absorption. The other "window" is open to certain radiation of radio wave length. This extends from 1 cm to about 15 metres (±). Radiation of wave lengths greater than about 15 metres is reflected by the ionosphere. The cut-offs at 1 cm. and 3.000 Å are not sharp, but first appear as separate absorption bands. These "windows" are shown graphically in the following figure:

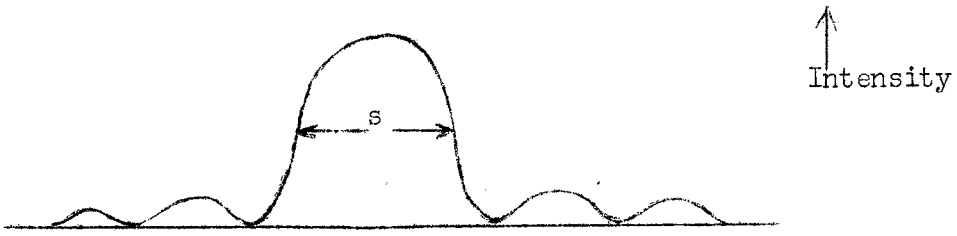


Sharp determination or selection of a given wave length is often troublesome in optical astronomy, but it is easy in radio astronomy. We can build circuits resonant to any radio frequency. In radio astronomy, we often use frequency instead of wave length, since frequency does not change in a medium. It is convenient to use the megacycle per second as the unit of frequency, and the meter or centimeter as the unit of wave length. In vacuum, we have the relation between frequency ν and wave length λ.

$$\lambda \text{ (in metres)} \cdot \nu \text{ (in megacycles/second)} = 300.$$

$$1 \text{ megacycle/second} = 10^6 \text{ cycle/second} = 10^6 \text{ sec}^{-1}$$

Position in the Sky. The resolving power of radio telescopes is very poor compared with that of optical telescopes, as would be expected from the longer wave length employed. We know that the image of a point source formed by an optical telescope is not a point, but rather consists of a central disk called the spurious disk, surrounded by diffraction rings, and that about 86% of the light of the source goes into this central disk. If we plot the intensity of the light in the image as ordinate and the distance from the center in radians as abscissa, we obtain a curve such as this:



where s is the width of the central disk as measured between points where the intensity is one-half the central intensity. If s is given in radians and if a is the aperture of the telescope in the units in which λ is measured, then s is given by:

$$s = 1.03\frac{\lambda}{a}$$

for the case when the aperture is circular. We see that the quantity s gives us a measure of the resolving power of the telescope, since the spurious disk of the images of two point sources separated by a distance s will already have begun to overlap. In radio astronomy we may liken the main lobe of the antenna pattern to the central disk, and the side lobes to the diffraction rings. If we put wave lengths representative of optical astronomy and radio astronomy into the above formula, along with practical values of the aperture we find that s for optical telescopes may be as small as a fraction of a second of arc, while for radio telescopes s may be measured in degrees or in tens of degrees. We see then, that it is difficult for a radio telescope to locate the point in the sky from which radiation is coming, with great accuracy. As in optical astronomy we use other methods to increase resolving power, such as interferometers and eclipses. These will be discussed in more detail later.

Time. This is an important element in radio astronomy measurements, for we need a reasonable time to suppress chance fluctuations, as in using a photo cell. Also, there are many transient phenomena that change very quickly and which we do not desire to suppress. We shall also consider this in greater detail.

Intensity. If we have a plane polarized [electro-magnetic] wave from a distant source falling upon a point in space, there will be an electric field $\vec{E}$ and a magnetic field $\vec{H}$ set up at that point. The absolute value of the Poynting vector $\vec{S}$ associated with that point will determine the energy carried by the wave at that point. In order to avoid the confusion in electro-magnetic theory caused by the variety of systems of units used, we shall review the expression for the Poynting vector in two systems of units in common use. In the practical system we have replaced $\sqrt{\mu_o/\varepsilon_o}$, or the "impedance of free space," by its value $2\pi \cdot 60$ ohm.

Theoretical System (Gauss)	Practical System (Giorgi)
E cgs Units	E volts/meter
H cgs Units (Gauss)	H amperes/meter
S ergs/sec cm^2	S watts/$meter^2$
then	
$S = \frac{c}{4\pi}\vec{E}x\vec{H}$	$S = \vec{E} \times \vec{H}$
but for a plane wave in vacuum	
$H = E$	$H = \sqrt{\frac{\varepsilon_o}{\mu_o}}E = \frac{E}{2\pi \cdot 60 \text{ ohm}}$
whence	
$S = \frac{cE^2}{4\pi}$	$S = \frac{E^2}{2\pi \cdot 60 \text{ ohm}}$

If the wave is periodic as of the form $E = E_o \cos \omega t$ then the time average is given by

$$\bar{S} = \frac{cE_o^2}{8\pi} \qquad \bar{S} = \frac{E_o^2}{4\pi \cdot 60 \text{ ohm}}$$

There is no need to assume such a periodic wave, If the radiation has the form of noise, we can take a time average and get:

$$\bar{S} = \frac{c}{4\pi}\overline{E^2} \qquad \bar{S} = \frac{\overline{E^2}}{2\pi \cdot 60 \text{ ohm}}$$

We thus obtain the energy in the entire spectrum, or the flux density in the entire spectrum. This quantity S (we usually omit the bar denoting time average) we recognize to be identical with the solar constant, in the case of the sun:

$$S = 1.94 \text{ cal cm}^{-2} \text{ minute}^{-1} = 0.13 \times 10^7 \text{ erg cm}^{-2}\text{sec}^{-1} = 1300 \text{ watt m}^{-2}$$

Polarization. Before we can go on to consider in detail the analysis of noise, we shall consider polarization. We inquire: What replaces S, the flux density, and $\overline{E^2}$ of our previous discussion when we consider a wave of arbitrary polarization? The relation $S = \frac{c}{4\pi}\overline{E^2}$ remains, where clearly $E^2 = E_x^2 + E_y^2$ where x and y are two arbitrary reference axes. Besides this, there are three other analagous parameters, comprising the four Stokes Parameters. These are best expressed in complex notation. Let A be the amplitude of the wave in the x or the y direction as specified and let an asterisk denote the complex conjugate. The Stokes parameters then become for a Maxwell wave:

$$\begin{aligned} I &= A_x A_x^* + A_y A_y^* \\ Q &= A_x A_x^* - A_y A_y^* \\ U &= A_x A_y^* + A_y A_x^* \\ V &= i(A_x A_y^* - A_y A_x^*) \end{aligned}$$

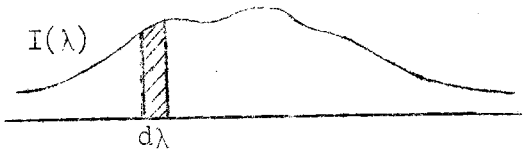
As an example, these parameters take on these values for variously polarized light

	I	Q	U	V	
Unpolarized	1	0	0	0	
Right-hand Circularly Polarized	1	0	0	1	
Plane Polarized in x direction	1	1	0	0	
Plane Polarized in y direction	1	-1	0	1	
Plane Polarized at 45° to x and y	1	0	1	0	et cetera

We shall not often need this formal representation.

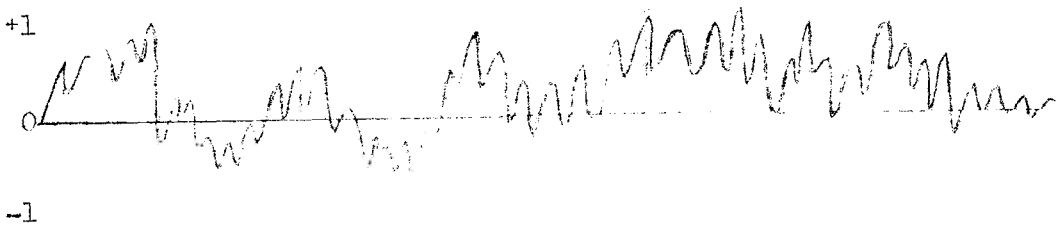
2.2 Noise and its measurement

Spectral intensity of a noise. In optical astronomy we can obtain a curve of the intensity as a function of wave length or frequency by means of the spectrograph. If we then have $I(\lambda)$ or $I(\nu)$, we know that the intensity within a band of wave length $d\lambda$ (or frequency $d\nu$), will be given by $I(\lambda)d\lambda$ (or $I(\nu)d\nu$).



The spectrograph thus performs an analysis of the light entering the slit. In order to devise an instrument that will analyse the radio-frequency components of the incident radiation in a similar way, we have to know exactly what analysis this is. The following formulation holds as well for optical as for radio astronomy.

The light arriving from a certain direction has an electric field strength that is a function of time, f(t), thus:



For convenience we discuss only one polarized component. This function is all that is given, so it contains information on the intensity distribution over all frequencies. But it cannot be directly registered. If f(t) were periodic,

$$f(t) = \cos 2\pi\nu t,$$

the radiation would be monochromatic with frequency ν. But strict periodicity does not exist, so always the line is broadened. More usually in astronomy and virtually always in radio astronomy f(t) is an irregular function or "noise." The spectrum then contains all frequencies and is a continuous spectrum.

Let us omit any constant part so that $\overline{f(t)} = 0$.

We then define the total Intensity I_{Tot} (total) thus:

$$I_{Tot} = \overline{\{f(t)\}^2}$$

where we are omitting any constant factors depending on units.

Now, suppose we consider an interval of time τ from t = 0 to t = τ and perform a Fourier analysis; then:

$$A_\tau(\nu) = \int_0^\tau f(t)\, e^{2\pi i\nu t} dt$$

If we perform an analysis in the interval t = τ to t = 2τ, we obtain some amplitude $A_\tau'(\nu)$, and for the interval t = 0 to t = 2τ clearly

$$A_{2\tau}(\nu) = A_\tau'(\nu) + A_\tau(\nu).$$

If we continue to add the amplitude A_τ, A_τ', A_τ'',.....,A_τ^n as the interval of time increases without limit, we do not arrive at a fixed value

of $A(\nu)$ since the amplitudes are complex numbers with virtually unrelated phases. We see, then, that Fourier-analysis of an infinitely long noise-train is meaningless.

We may, however, consider the mean squares, where the mean implies an average from samples taken far apart; then:

$$\overline{|A_{2\tau}(\nu)|^2} = 2\overline{|A_\tau(\nu)|^2}$$

So it is not A itself but the modulus of its square which is proportional to τ. Moreover, we can show by a well-known theorem of statistics that the relative dispersion of the actual values of $|A(\nu)|^2$ decreases as $\tau^{-\frac{1}{2}}$. We see then that if we take τ really large, the values of $|A_\tau(\nu)|^2$ will have a small relative dispersion and will be almost strictly proportional to τ, thus the limit

$$\lim_{\tau\to\infty}\frac{2}{\tau}|A_\tau(\nu)|^2 \qquad \text{exists}$$

and we define

$$I(\nu) = \lim_{\tau\to\infty}\frac{2}{\tau}|A_\tau(\nu)|^2$$

as the spectral intensity of the noise function. We shall see that the factor 2 is needed for normalization.

Total Intensity Theorem. What we want, however, is a definition that gives:

$$I_{Tot} = \int_0^\infty I(\nu)d\nu\ .$$

We have defined the total intensity I_{Tot} and the spectral intensity $I(\nu)$. We shall show that the above relation follows from these definitions. We shall need Parseval's Theorem to do this. This reads:

$$\text{if } F(\nu) = \int_{-\infty}^{+\infty} f(t)e^{2\pi i\nu t}dt,$$

then by Fourier inversion:

$$f(t) = \int_{-\infty}^{+\infty} F(\nu)e^{-2\pi i\nu t}d\nu$$

and Parseval's Theorem states:

$$\int_{-\infty}^{+\infty} f(t)f^*(t)dt = \int_{-\infty}^{+\infty} F(\nu)F^*(\nu)d\nu$$

We then apply this theorem to

$f(t)$ = the chopped off real noise function $f(t)$ of sample duration τ, and to $F(\nu) = A_\tau(\nu)$; $F(-\nu) = A_\tau^*(\nu)$

then

$$\int_0^\tau |f(t)|^2 dt = 2\int_0^\infty A_\tau(\nu)A_\tau^*(\nu)d\nu$$

Now if we divide by τ and take $\lim_{\tau\to\infty}$, we obtain the required relation.

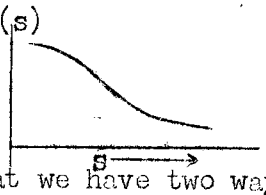
Correlation-Function Theorem. We could get at this another way. Let $g(s)$ be a function of the correlation interval s defined thus

$$g(s) = \overline{f(t)f(t+s)} \quad \text{where } f(t), g(s) \text{ are real;}$$

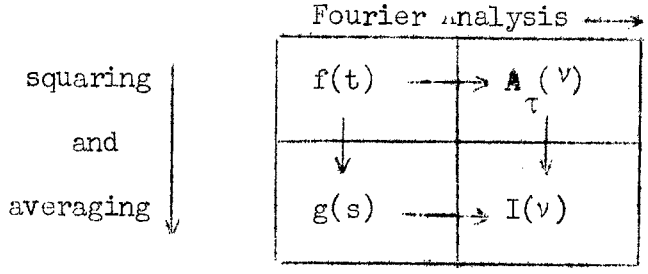
then $I(\nu)$ is given by: (theorem stated without proof)

$$I(\nu) = 4\int_0^\infty g(s)\cos 2\pi\nu s\,ds$$

The "correlation function" $g(s)$ will go toward zero quickly, being of the form shown.



We see then, that we have two ways of obtaining $I(\nu)$ from $f(t)$, which we might illustrate in this form:



Also, by Fourier Inversion we have

$$g(s) = \int_0^\infty I(\nu) \cos 2\pi \nu s d\nu$$

We see that for s = 0 we have the Total Intensity Theorem, whereupon we might consider the Correlation Theorem an extension of the Total Intensity Theorem.

Application to Radio Noise. The mathematical treatment given above is important, for we shall see that the instrumental technique follows the mathematical process exactly. Also, the time τ of the analyzed sample is relatively short, so that intensity fluctuations are often important. We shall consider these two points at greater length.

A radio telescope consists of an antenna which intercepts the radiation, a receiver composed of an amplifier and a detector, and some sort of registering meter. The receiver is usually of conventional superheterodyne design, often a modified radar receiver. The antenna does some of the selecting of frequencies, but not much. The varying field strength due to the radiation at the antenna corresponds to the f(t) of our analysis. The antenna is linked to the amplifier, usually by means of wave guides for radiation shorter than 20 cm., or by coaxial cable for radiation of longer wave length. The selection of frequency is done in the amplifier, which selectively amplifies frequencies in a narrow band $\Delta\nu$, around some frequency ν, about a millionfold in typical cases. This corresponds to the computation of the Fourier-amplitude in our analysis. The "squaring" is done in the quadratic or "square-law" detector.

Theoretically an infinite time is needed to record spectral intensity, since we defined $I(\nu)$ in the limit $\tau \to \infty$. If we analyze a number of short

samples of duration τ, the values of $I(\nu)$ may differ from one sample to another, so that the meter will record different values of the "intensity" for different samples. These are called "intensity fluctuations."

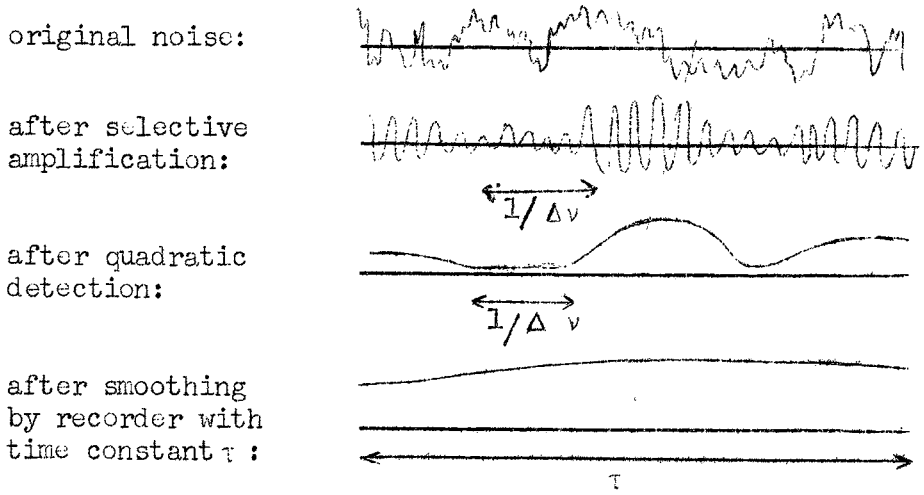
We can reduce these fluctuations, if we desire, by increasing the time constant of the recording meter. Or we might simply average the registered intensity over a time τ.

This is exactly the same procedure which we followed in the mathematical formulation. Let τ be the time constant of the recording meter. Then the relative fluctuation of the recorded intensity will be proportional to $\tau^{-\frac{1}{2}}$. The further factor determining the size of these fluctuations is the frequency range $\Delta\nu$ of the accepted radiation, i.e. the band width of the amplifier. We have (with omission of a factor of the order of 1):

$$\frac{\sigma_I}{I} = (\Delta\nu \cdot \tau)^{\frac{1}{2}}$$

where σ_I is the root mean square deviation of the measured intensity from its average (true) value.

It is instructive to see from the following illustration what the physical reason for this relationship is.



(Rayleigh kind)

The clue is that the current after selective amplification has a fixed phase and amplitude only for a time of the order of $1/\Delta\nu$. The detected current still varies in strength all the way from zero over a wide range, so its relative dispersion σ_I/I is of the order of 1. Smoothing over a time τ that is much longer than the time in which this current fluctuates reduces the relative dispersion to $(\Delta\nu\cdot\tau)^{\frac{1}{2}}$. We can reduce the fluctuation by increasing $\Delta\nu$ or τ. We shall consider some limitations placed on the values of these quantities presently.

Internal and External Noise. What we hope to measure at the recording meter, is not the intensity of the radiation directly, for this depends also on certain properties of the antenna which we shall consider later. We hope, instead, to measure the total power absorbed by the receiver. What we actually measure, however, is the power absorbed from the antenna plus the power generated within the receiver itself.

This power is generated by thermal motions of electrons in resistors and tubes, and it gives random noise of the same form as that we are trying to measure. The noise arising in each stage of the amplifier will be amplified in each succeeding stage. So the noise arising in the first stage, which is amplified just as much as the external noise, is most bothersome.

In practice, then, the power P which we measure is the sum of the power absorbed from the antenna P_{ext} and the power generated by the equipment P_{int}. We often find that the power which we hope to measure, P_{ext}, is much smaller than the internal noise power P_{int}.* For example, suppose that the antenna noise is negligible. The output meter reading will fluctuate in a random way about the reading corresponding to the value P_{int}. The size of these

* If we wish to detect a small value of P_{ext}, it must be at least as large as the fluctuations in P_{int}.

fluctuations will be

$$\Delta P = \frac{P_{int}}{\sqrt{\Delta\nu\ \tau}}$$

If we now connect the antenna, the meter reading will fluctuate about a new, higher value corresponding to the power $P_{int} + P_{ext}$, but we shall not be able to detect the increase unless the increase is larger than the fluctuations, that is:

$$P_{ext} > \frac{P_{int}}{\sqrt{\Delta\nu\ \tau}}$$

We see, then, that the greatest sensitivity will be obtained by using an amplifier of large band width, followed by an output indicator having a large time constant. (Since the reciprocal of the time constant acts as a band width, one often calls $1/\tau$ the output band width and $\Delta\nu$ the input band width.)

There are, however, certain fundamental limitations which are imposed on the values of $\Delta\nu$ and τ. If we desire to know how intensity varies with frequency we must use a narrow band width. If we are using an interferometer, then the radiation must be nearly "monochromatic" if we are to obtain many fringes. This means that $\Delta\nu$ must be smaller than about $1/100\,\nu$ or $1/1000_{\nu}$. There is another limit which is imposed on the band width by the nature of the amplifier circuits used: wide-band amplifiers require low input and output impedances, while high-gain amplifiers require high input and output impedances. Consequently, a wide band can be realised only with a small gain in each stage. This in turn means that the internal noise is relatively high, because effectively more than just the first stage contribute to it. There are also limitations imposed by the output band width (that is, by the time constant of the recording meter). If we want to observe any transient

or variable phenomena, then τ must be shorter than the time of variation. We might also want to observe rapid changes in power caused by rotation of the antenna, or we might want to observe very short pulses. There is another more practical and difficult problem: we cannot make τ larger than the time during which the factor of amplification remains constant. In practice it is difficult to keep the gain (amplification factor) constant. For example, Bolton found with his equipment, that in order to keep P_{int} constant to one part in 1000, he had to keep the plate voltage constant to one part in 3000, and the heater voltage at one part in 1500. One solution is continuous calibration, but first, let us consider calibration in general.

Calibration. In practice, equipment is calibrated as follows: A reading R_1 is taken with the receiver connected to the antenna, whose power is to be measured. The antenna is then disconnected and two known noise sources of power, P_1 and P_2, are connected in turn, and the readings of the output meter, R_2 and R_3, are taken. We know that in each case the output reading will be equal to the sum of the internal and external noise power times some constant factor x. We thus have three equations:

$$x\cdot(P_{ext} + P_{int}) = R_1$$
$$x\cdot(P_1 + P_{int}) = R_2$$
$$x\cdot(P_2 + P_{int}) = R_3$$

and we solve for the unknowns P_{int}, P_{ext} and x. Signal generators are not accurate enough to be used in the calibration. Also, it is better to use a power source which generates noise rather than a signal. For noise generators we may employ a resistor at some temperature T, or we may use a diode with

the filament at some temperature, T. These must be provided with some circuit to match them to the receiver. That is the impedance of the noise source must match that of the receiver. Matching will be discussed in more detail later.

The method of calibration just described leads us to a method of defining the noise factor N, which is the measure of excellence of a receiver as regards sensitivity. Let the antenna be replaced by a matched resistor at room temperature T; this produces the (available) noise power $P_o = kT\Delta\nu$ (see later). If now a signal with power S is introduced in this resistance, the signal to noise ratio in the input is S/P_o. After amplification, with gain G, the signal to noise ratio has become $GS/(GP_o + GP_{int})$. The noise factor N is now defined by

$$N = \frac{\text{signal/noise ratio in input}}{\text{signal/noise ratio in output}} = \frac{S/P_o}{S/(P_o + P_{int})} = \frac{P_o + P_{int}}{P_o}.$$

Consequently

$$P_{int} = (N-1)P_o.$$

A common method of measuring N is to measure first the noise output with just the resistor as input and then to add by a noise generator a known noise power P_g to the input until the output noise is doubled. The equations then are

$$x(P_o + P_{int}) = R$$

$$x(P_o + P_g + P_{int}) = 2R,$$

from which we find

$$N = P_g/P_o.$$

Technical details must be omitted.

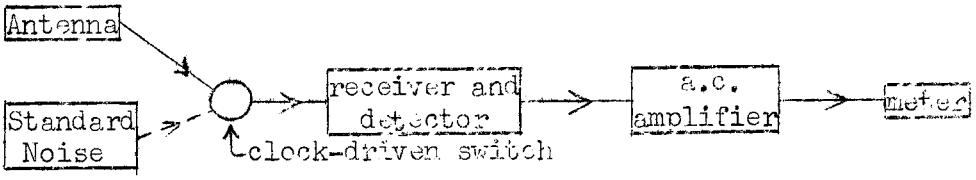
We have then, the following definition of noise factor N: The noise factor N is some number, characteristic of the receiver, such that N-1 is

the ratio of the actual internal noise of a receiver, to the theoretical minimum at that temperature. Often the noise factor is expressed in decibels: $n = 10 \times \log_{10} N$.

An idea of the sensitivity of receivers in use at the present time can be gained from the following values giving noise factors which can be obtained without undue difficulty at various frequencies:—

Frequency (Mc/s)	Noise Factor		~T_{sys}
	decibels	numerical ratio	
10	2	1.6	170 K
100	6	4	870
1000	10	10	2600
10000	15	32	9000

Continuous Calibration. The difficulties arising from variation of amplifier gain can be reduced by using a method of continuous calibration. The most commonly used method, often called the "chopping" method, was originated by Dicke (1946).



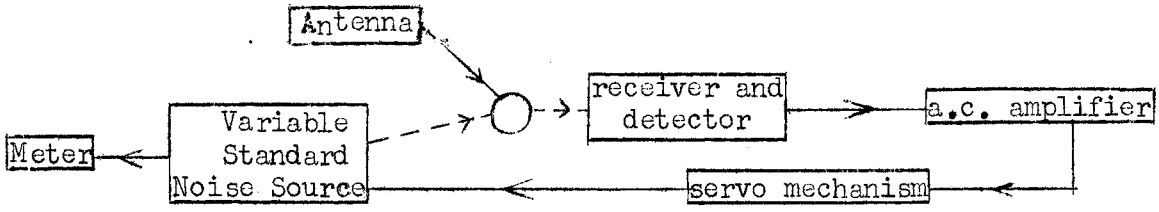
In this, the antenna and a standard noise source are connected alternately to the receiver, the switching being done at the rapid rate of about 30 c/s. The difference between the noise powers produced by the antenna and the standard source then appears as a 30c/s modulation of the receiver output. Thus the greatest sensitivity to the incident power can be obtained by using an additional detector which is sensitive only to the 30c/s component of the

output of the quadratic detector, eliminating the slower, undesired, variations. The fluctuations of the output power are governed by the bandwidth $\Delta\nu$, and the meter time constant τ, just as in the simpler technique.

While this system reduces the effect of receiver variations, calibration is still necessary, since the output depends on the linearity of the receiver, and the law of the detector.

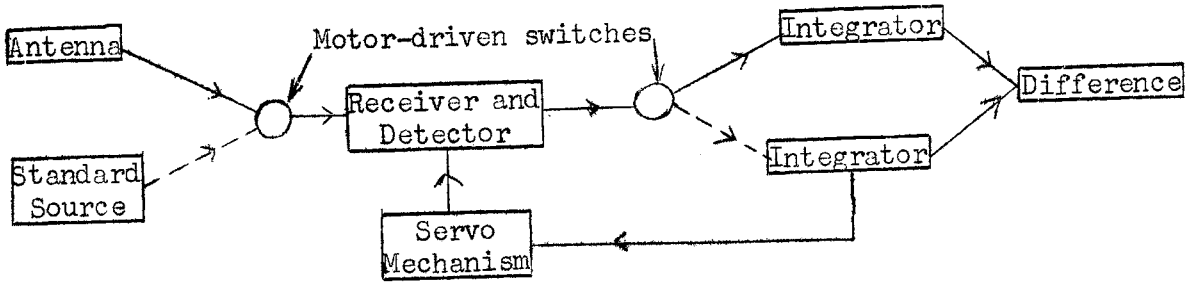
Some variants of Dicke's method have already been used in radio astronomy, and it is likely that others will come into use in the future, as the need for precise measurement of very small noise powers increases. For example, Ewen (1951), in his detection of the 21 cm line of galactic hydrogen, has used a system in which the frequency is switched back and forth through a small amount at a rapid rate.

Ryle and Vonberg (1948) have developed another method of measuring small noise powers which does not depend on knowledge of the law of the detector or the gain or the frequency response of the receiver. Also, it can operate over a wide range of input powers without the necessity of changing the gain and calibrating. In this method, a variable standard noise source is employed. The receiver is switched alternately between the antenna and noise source, and a servo mechanism varies the noise power of the source so that it is kept in equilibrium with the antenna. We can measure the power delivered by the antenna then, by measuring that of the standard source,



and no further calibration is needed. There is, however, this limitation: the power delivered by the antenna must be as high as the minimum noise we can generate.

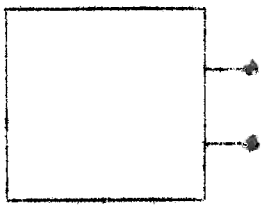
Steinberg has devised a method of continuous calibration which actually tries to stabilize the gain of the receiver. A constant standard noise source is employed and again the receiver is switched from the antenna to the standard. When the receiver is connected to the standard, a servo mechanism checks the output, and if it is not at a certain constant value, it adjusts the gain of the receiver until it is at this value. The final reading, then, is the difference in output between the antenna and standard source.



The development of such systems should make it possible to push the value of τ to minutes or even hours and thus overcome the difficulty of obtaining large values of $\tau \Delta\nu$, and we could then measure very weak noise powers. Of course, such systems have no advantage when we desire a short time constant to measure transient phenomena.

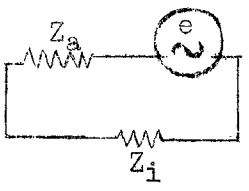
There still remain other limitations on the intensities we can measure, for variations in the ionosphere produce effects similar to bad "seeing" which we have to contend with in optical astronomy.

Available Power and Matching. In what has preceded, we have not defined just exactly what we shall understand by the power absorbed by our radio telescope. We shall now investigate this more closely.



Let us consider a two terminal network such as that shown on the right. If we measure the voltage V and the current I across the terminals, with proper regard to phase, then the impedance of the circuit is given by $Z = V/I$ and the power dissipated in the circuit by $P = \frac{1}{2}I^2R$, where R is the resistance, that is, the real part of Z. Now if we put an antenna in space we may consider it as a two terminal network and measure V and I. The resulting impedance is called the radiation impedance Z_a. The real part of this impedance is called the radiation resistance R_a. This is an equivalent resistance, that is, a conception to indicate the radiation properties of an antenna, drawn from an analogy with ohmic power losses. In other words, the radiation resistance is the equivalent resistance which would dissipate the power the antenna radiates, with a current flowing in it equal to the antenna current. This resistance varies slightly with the way the antenna is mounted, being affected by reflectors placed near it, height above the ground, conductivity of the ground, etc.

Now let us connect the antenna to the receiver. We symbolize this by the diagram on the right, in which the antenna is regarded as generating, as a result of the incident radiation, an e.m.f. $e = e_o \cos \omega t$ (where the subscript zero indicates the maximum value). The current I flowing in this circuit will be given by $I = \frac{e}{Z_i + Z_a}$. The power absorbed in Z_i will

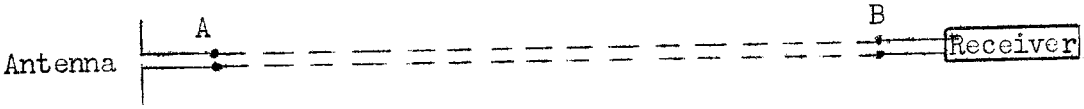


then be given by $P = \frac{\frac{1}{2}e_o^2}{|Z_i + Z_a|^2} R_i$ where the factor of $\frac{1}{2}$ enters by way of a time average.

Now if we hold Z_a and e_o constant, it can be readily shown that the maximum power will be absorbed in Z when $Z_i = Z_a^*$. If we write $Z_i = R_i + iX_i$ and $Z_a = R_a + iX_a$ we see that $Z_i = Z_a^*$ implies that $R_i = R_a$ and $X_i = -X_a$. We see then that the maximum power will be absorbed in Z_i when $Z_i + Z_a = 2R_a$. When Z_i is so chosen, the receiver is said to match the antenna. The power available to the receiver will then be $\overline{P} = \frac{\frac{1}{2}e_o^2}{4R_a}$.

Note that an equal amount of power is absorbed in the antenna. This power is re-radiated by the antenna into space.

We must also consider the transmission line which links the antenna to the receiver.



In the simplest (idealized) case there is no reflection at A and B. This requires that $Z_A = Z_o = Z_B$, where Z_o is the characteristic impedance of the line. There are no ohmic losses in the line if Z_o is real. Then matching requires that also Z_A and Z_B be real. Mismatch at B would cause reflection of incoming waves, i.e. standing waves in the transmission line.

In conclusion, then, we call the "absorbed power" the power which is collected by the antenna and made available to the receiver. The P_{ext} which we considered earlier is identical to this absorbed power.

2.3 Antennas, general properties

We have considered the properties of radiation and noise and how the noise power from the antenna is measured. We now consider the amount of energy actually radiated from the astronomical source and the antenna which intercepts this energy.

Astronomical Source. Any source may be considered to have some angular extent, and we can find a small solid angle over which it is uniformly bright. For that part of the source then, the brightness B has the definition

Brightness B = energy per unit time per unit area per unit solid angle per unit band width

The units are: ergs/sec cm^2 sterad cycles/sec.

or: watts/$meter^2$ sterad cycles/sec.

In optical astronomy this is more usually called the "surface brightness" of the source or "specific intensity" of the radiation emitted by the source. Brightness is the advisable term.

We have also the flux density S of the radiation from any finite source.

Flux Density S = power per unit area per unit band width

The units are: ergs/cm^2 cycles/second

or: watts/$meter^2$ cycles/second

We might call the flux density the "integrated brightness."

We observe that B is the more fundamental property since it tells us directly a meaningful property of the astronomical object whereas S varies inversely as the square of the distance of the observer from the object. Evidently then, the flux density is the brightness integrated over the solid angle to the object, thus

$$S = \int B d\Omega .$$

It will be noted that both B and S refer to power or energy. They include the energy in the entire beam, regardless of its polarization and correspond to the first Stokes Parameter (I). A fuller representation would give the three other parameters, and specify the state and degree of polarization of each point of the source.

Reciprocity Theorem. Helmholtz proved by means of an extension of Green's Theorem, that "if in a space filled with air which is partly bounded by extended fixed bodies, sound waves are being excited at any point A, the resulting velocity-potential at a second point B is the same both in magnitude and phase, as it would have been at A, had B been the source of the sound." This theorem was extended to infinite regions by Rayleigh. He greatly generalized this concept of reciprocity and extended it to studies in electricity and the theory of heat radiation and conduction.

This principle was extended to the radiation of antenna systems by Carson and is known as the Rayleigh-Carson Reciprocity Theorem and may be stated thus: if an electromotive force E, inserted in antenna A, causes a current I to flow at a certain point in a second antenna B, then the voltage E applied at this point in antenna B will produce the same current I (both in magnitude and phase) at the point in antenna A where the voltage E was originally applied. The theorem is restricted in the generality given above in that it fails to be true in an ionized medium in a magnetic field. However, what is significant for our purposes is that it follows from the above that we can study the properties of a receiving antenna as though it were a radiator. In practice, it is a far simpler problem to consider or design an antenna as a radiator than as an absorber of radiation. For simplicity then, we shall consider transmitting antennas.

Antennas. Let a distant source be located in a certain direction described by the coordinates (θ, φ) and let the source have the state of polarization for which the antenna is most sensitive and let the receiver be matched with the antenna so that the power absorbed is a maximum, then the "effective area" or "absorption cross section" of the antenna is defined by

$$A(\theta\ \varphi) = \frac{P_{abs}}{S\ \Delta\nu}$$

where P_{abs} is the power absorbed in the receiver and $S\ \Delta\nu$ is the flux density over a small band width. If the source is emitting natural radiation (randomly polarized), then we can decompose it into two equal parts, one for which the antenna is most sensitive, and one for which it is insensitive so that

$$P_{abs} = \tfrac{1}{2}A(\theta,\varphi)\cdot S\ \Delta\nu$$

If we desired a full representation for radiation of arbitrary polarization, we should again require the Stokes Parameters. If we were to consider the power absorbed by two sources, then clearly we would have

$$P_{abs} = \tfrac{1}{2}\ A(\theta_1\ \varphi_1)S_1\ \Delta\nu + \tfrac{1}{2}\ A(\theta_2\ \varphi_2)S_2\ \Delta\nu$$

In general then, for incident radiation of random polarization from an extended source, we introduce the brightness and integrate over the solid angle described by the source; then:

$$P_{abs} = \tfrac{1}{2}\int A(\theta\ \varphi)B(\theta\ \varphi)\ d\Omega\ \ \Delta\nu$$

We try to design antennas in such a way that $A(\theta\varphi)$ is larger in one direction than in others. If we plot a diagram of A vs direction, we obtain what is called the "antenna pattern" or "directional diagram" such as is

shown at right. The portion of the diagram where A is larger is called the "main lobe" or the main beam. The smaller lobes are called "side lobes." If the main beam is narrow it is called a "pencil beam." In such a beam the maximum value of $A(\theta\ \varphi)$ is denoted by A_o. We now define the average antenna cross-section $\bar{A}$ thus:

$$\bar{A} = \frac{1}{4\pi}\int A(\theta\varphi)\,d\Omega$$

and then we define the gain $G(\theta\ \varphi)$ in an arbitrary direction thus:

$$G(\theta\ \varphi) = \frac{A(\theta\ \varphi)}{\bar{A}}$$

and the gain in the forward direction (that is, in the maximum direction) as

$$G_o = \frac{A_o}{\bar{A}}$$

We note that by the Reciprocity Theorem, the power $P(\theta\ \varphi)$, which is radiated if the antenna is used as a transmitting antenna, has the same angular distribution. So $A(\theta\ \varphi)$, $G(\theta\ \varphi)$ and $P(\theta\ \varphi)$ are strictly proportional.

We shall now write down a relation which is fundamental for all antennas under matched conditions. The derivation follows from thermodynamics and will be given later: (p.42)

$$\bar{A} = \frac{\lambda^2}{4\pi}$$

or in a more familiar form, since $G_o = A_o/\bar{A}$

$$A_o = G_o \cdot \frac{\lambda^2}{4\pi}$$

We usually omit the subscript zero when we speak of the gain of an antenna, for we usually consider the gain as that in the optimum direction.

What we have called the gain in the above discussion should be strictly called the directivity D of the antenna. It is related to the gain by the relation

$$G = LD$$

where L is a loss factor due to losses in the transmission line. We usually assume, however, that we have so designed the transmission line that losses are negligible and thus L = 1. Then the values of G and D are the same.

There are several other quantities which serve the same purpose as the gain. We define the "effective solid angle" Ω_A of the antenna thus:

$$\Omega_A = \frac{4\pi}{G}$$

It is dangerous, however, to compute G and (so Ω_A) by looking only at the main lobe, for side lobes would depress G and thus make Ω_A larger than expected. We prefer to use G and not Ω_A.

We also define the "half-widths" of the main beam, s_a, s_b. The half-width is the angle between the points on the opposite sides of the main lobe where $A = \frac{1}{2}A_0$. Since the main beam almost always has an elliptical cross section, we usually find that $s_a \neq s_b$. There must be a rough relation between s_a, s_b and Ω_A. For instance, if A were constant inside an elliptical solid angle and zero outside, then:

$$\Omega_A = \frac{\pi}{4} s_a s_b = \frac{4\pi}{G}$$

whence

$$G = \frac{16}{s_a s_b}$$

However, more generally, there are side lobes and the gain is smaller than this. In practical antennas, an average value is

i.e., main beam efficiency = $\frac{10}{12} = 0.82$

$$G = \frac{10}{s_a s_b} \qquad \text{with } s_a,\ s_b \text{ in radians}$$

or

$$G = \frac{35000}{s_a s_b} \qquad \text{with } s_a,\ s_b \text{ in degrees.}$$

Special cases. In summary then, we say that for the power absorbed by an antenna exposed to randomly polarized radiation:

$$P_{abs} = \tfrac{1}{2}\int A(\theta\ \varphi)B(\theta\ \varphi)\,\Delta\nu\,.\,d\Omega$$

Finding the brightness distribution $B(\theta, \varphi)$ thus involves in general the solution of an integral equation. This equation is similar to the "equations of insufficient resolving power" occurring in spectroscopy, stellar statistics, etcetera. It here occurs in two dimensions on a sphere.

In two special, limiting cases, this equation becomes much simpler. First, consider an extended, homogenous source which fills the whole pattern. Then the brightness B will be constant, whence:

$$P_{abs} = \tfrac{1}{2}B\ \Delta\nu \int A(\theta\ \varphi)\,d\Omega$$

but $$\int A\,d\Omega = 4\pi\,\bar{A} = \lambda^2,$$

whence $$P_{abs} = \frac{\lambda^2}{2} B\,\Delta\nu$$

Next, we consider a very small source, so small that it is observed by the tip of the main lobe, whence $A = A_o$ = constant and

$$P_{abs} = \tfrac{1}{2}A_o\ \Delta\nu \int B(\theta\ \varphi)\,d\Omega$$

but $$\int B\,d\Omega = S$$

whence $$P_{abs} = \tfrac{1}{2}A_o S\ \Delta\nu = \frac{\lambda^2}{8\pi} G_o S\ \Delta\nu.$$

The examination of these special cases shows us that if the antenna is directive enough to make the source appear "extended," it does not help to make

it bigger, for P_{abs} does not contain the gain. We cannot then hope to improve the ratio P_{ext} / P_{int} except by reducing P_{int}, that is, by improving the receiver. For small sources, however, P_{abs} goes up with the absorption area A_o of the antenna.

References

R. H. Dicke; "The measurement of thermal radiation at microwave frequencies," Rev. Sci. Instr. 17, 268, 1946.

H. E. Ewen; to be published.

M. Ryle and D. D. Vonberg; "An investigation of radio frequency radiation from the Sun," Proc. Roy. Soc., A, 193, 98, 1948.

Suggested Reading

1.) R. Furth & D.K.C. McDonald; "Statistical analysis of spontaneous electrical fluctuations," Proc. Phys. Soc. London 59, 375, 1947.

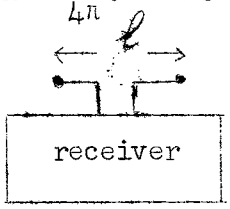
2.) S. O. Rice; "Mathematical analysis of random noise," Bell System Technical Journal 23, 282-332, 1944 and 24, 46-156, 1945.

3.) S. Chandrasekhar; "Stochastic problems in physics and astronomy," Rev. of Mod. Phys., 15, 1943.

4.) J. L. Pawsey and L. L. McCready, "The measurement of solar radio frequency radiation," C.S.I.R. Sydney, Report No. RPR 74, 1948.

2.4 Antennas, special types

Small dipole. This case will be worked out in some detail, to illustrate the use of the reciprocity principle, and to demonstrate the universal relation $\overline{A} = \frac{\lambda^2}{4\pi}$. Consider a dipole whose length is small compared with the wavelength, so that there will be no phase variation along the conductors. Suppose also that the dipole has very large end capacity, so that it may be regarded as two oscillating point charges.



In the emission case, let the oscillating current be $i_o \sin \omega t$, and the oscillating charges be $+q_o \cos \omega t$ and $-q_o \cos \omega t$. The electric moment will then be $p_o \cos \omega t$, where $p_o = q_o \ell = \frac{i_o \ell}{\omega}$, and the total power emitted by the dipole,

$$P = \frac{4 p_o^2}{3c^3} = \frac{\pi^2 i_o^2 \ell^2}{3c^3}$$

or in MKS units,

$$P = p_o^2 \omega^2 (\tfrac{2\pi}{\lambda})^2 \cdot 10 \text{ ohm} \quad (\text{since } \sqrt{\tfrac{\mu}{\varepsilon}} = 120 \text{ ohm in these units}).$$

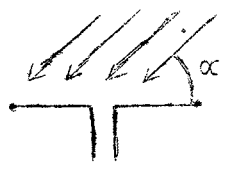
But P can also be put equal to $\frac{1}{2} i_o^2 R_r$, where R_r is the radiation resistance, thus

$$R_r = (\tfrac{2\pi\ell}{\lambda})^2 \cdot 20 \text{ ohm}$$

The antenna impedance is in this case mainly reactive, however, only a small part being real (R_r).

The emission pattern is proportional to $\sin^2 \alpha$, where α is the angle with the dipole axis.

Now, in the absorption case, consider a wave coming from angle α. If the electric vector is in the appropriate direction, the effective length of the dipole is then $\ell \sin \alpha$, so that the induced e.m.f. is $e_o = E_o \ell \sin \alpha$



Hence the power absorbed in a matched receiver is

$$P_{abs} = \frac{\frac{1}{2}e_o^2}{4R_r}$$

$$= \frac{E_o^2 \lambda^2 \sin^2\alpha}{32\pi^2 \cdot 20}$$

Further, the flux density, according to page 8 , is

$$S\,\Delta\nu = \frac{E_o^2}{4\pi \cdot 60} .$$

Hence the absorbing cross-section, A, is given by

$$A = \frac{P_{abs}}{S\,\Delta\nu} = \frac{3\lambda^2 \sin^2\alpha}{8\pi}$$

Averaging over all directions,

$$\overline{\sin^2\alpha} = 2/3$$

so that the mean value is

$$\bar{A} = \frac{2}{3} \cdot A_o = \frac{\lambda^2}{4\pi},$$

demonstrating this universal relation in one particular case.

The maximum gain is

$$G_o = 3/2.$$

Practical antennas.* ← ref 34 It is desirable that the antenna should accept appreciable energy only over a small solid angle whose direction can be altered at will by the mounting mechanism. Outside the cone of acceptance it should have the smallest possible response so as to reduce external interference, either man-made or natural. This condition is more easily realized at wavelengths below one meter.

In the frequency range 50 to 200 Mc/s it is common practice to use Yagi antennas. A Yagi antenna is an end-fire array of dipoles in which there is only

one fed element, the remainder being parasitic, i.e. excited into forced oscillation. A rigorous theoretical treatment of Yagi antennas is difficult and their design is often finalized experimentally. However, once a satisfactory performance has been secured at one frequency it is possible to obtain good results at a different frequency by altering the dimensions in direct proportion to the wavelength. A disadvantage is the high selectivity, which restricts the use of such an antenna to a range of 5% or 10% in frequency.

In the decimeter and centimeter-wavelength region it is common practice to employ a parabolic mirror with a dipole feed at the focus. The mirror may be a metallic sheet, wire mesh, or made of parallel metal slats. A metallic sheet is seldom used at the higher wavelengths because of the excessive wind loading. When the mirror dimensions are large compared with the wavelength, the radiation pattern is analogous to a Fraunhofer diffraction pattern in optics. A characteristic main lobe directed along the main axis is produced, with a series of minor lobes of low intensity.

At wavelengths below a few centimeters, good directivity is obtained with relatively small mirror dimensions. In this case a mirror spun from sheet metal is usually employed because irregularities in the parabolic surface must be kept below a small fraction of a wavelength. Also the dipole is replaced by a horn feed.*

(In line with the usage of talking about an antenna as a transmitting antenna, the collector at the focus is called the feed system, since it feeds energy into the mirror.)

---

* The first four paragraphs of this section are taken from "The measurement of solar radio-frequency radiation," by J. L. Pawsey and L. L. McCready, Report No. RPR 74, Radiophysics Laboratory, Sydney, where photographs of several typical antennas can be found.

The feed system is designed to give the optimum concentration of energy into the mirror. If the mirror could be filled completely and homogeneously, the gain would be (following from diffraction calculations),

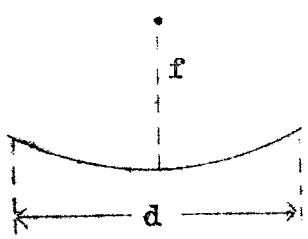
$$G_o = \frac{4\pi A'}{\lambda^2}$$

where A' is the real area of the mirror. Actually the mirror is always less effective by a factor k, which is about 0.7 in typical cases, making

$$G_o = \frac{4\pi}{\lambda^2} kA'$$

and so $$A_o = kA'.$$

The f/d ratio is usually chosen to be about 1/4 to 1/3. If the ratio is too small, the outer parts of the mirror cannot be made very effective, especially in the regions in the direction of the dipole axis. If the ratio is made too large, it becomes difficult to design a focal feed that does not spill over the mirror, and also the length of the radio-telescope may become prohibitively large.



A useful feature of parabolic antennas is their flexibility with regard to frequency. It is possible to use one mirror for the simultaneous reception of widely spaced frequencies through the use of a multiple feed system.

The following tabulation lists the various types of antenna which have been or might be used in radio astronomy:--

A. Hardly directive.

1. $\lambda/2$ dipole. (Gain = 1.65, radiation resistance = 73 ohm, no reactive component)
2. $\lambda/2$ dipole, with "one-sided" pattern. ($G \approx 4$)
   (a) with reflecting sheet $\lambda/4$ behind it.
   (b) with parasitic reflector behind it.

3. Yagi ($G \approx 12$)

B. Very directive.

4. Broadside array. (Combinations of any of earlier systems, arranged to make interference favorable in forward direction.)

5. Horn.

6. Paraboloid reflector.

7. Rhombic antenna.

8. Lenses, metal or dielectric.

C. Interferometers. (described in the section on point sources, p. 147)

9. Michelson interferometer.

10. Lloyd's mirror.

D. Antennas for circular polarization.

11. Two antennas (plane-polarized) in perpendicular planes, connected with a $\lambda/4$ phase-difference. May be simultaneously used as an interferometer, if separated.

12. Paraboloid, with a $\lambda/4$ plate in front of it, at 45° to the dipole.

13. Helical antenna.

Chapter 3. TEMPERATURE AND THERMAL NOISE (Henry J. Smith)

3.1 Thermal Fluctuations and Noise

In order to give meaning to the scale deflections measured on the recording meter attached to our radio telescope we must introduce the thermodynamical concept of temperature, and the closely associated notion of noise. When a radio telescope is receiving or transmitting a signal, it is to a certain degree in a statistical steady state. That is to say, the currents and voltages across any one or more components of the circuit are constant when measured with an instrument of all but the highest degree of precision. But a thermodynamical specification of a steady state is a process of averaging over large numbers of the systems which constitute an assembly. Individual systems (in electronic circuits, electrons) will have properties fluctuating about a mean value. These fluctuations (in velocity and density) are described by the term "thermal motion." The kinetic temperature of an assembly is a numerical measure of the magnitude of thermal motion. In communications engineering thermal fluctuations of electrons about the steady state are often referred to as noise.

The Principle of Detailed Balancing: This can be stated as follows:

> In any closed system in thermal equilibrium at a certain temperature, any detailed process is statistically counterbalanced by its precise reciprocal process.

No formal justification of this principle has yet been given. For a discussion of its plausibility and interpretation, the reader may refer to Eddington's "Internal constitution of the Stars," page 45.

The usefulness of the principle lies in the information it can yield concerning the rates at which processes occur. In general we can write the rates in forms of this type:

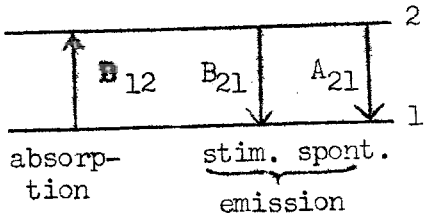
$$(\quad) = (\quad)\times(\quad)\times(\quad) \times (\quad)$$

Rate — Each factor is the number of particles of a certain kind participating in the process. — "Cross-section" and similar constants of the system for the process.

A similar equation applies to the reciprocal process. Along with "cross-sections" one has atomic constants—transition probabilities, absorption coefficients, etc. In the equilibrium state, the numbers of particles in the initial and final states will be determined by thermodynamical laws (Boltzmann, Maxwell, Saha, etc.). By the principle of detailed balancing the rates of each process and its reciprocal will be equal, and therefore we can specify the ratio of the cross-sections of the direct and reverse processes.

The use of this principle in astrophysics is as follows. If the cross-section of a process is known, then, by devising a fictitious experiment in a thermostat and postulating detailed balancing, we find the cross-section of the inverse process. Afterwards these cross-sections can be applied to situations in which cyclical and unbalanced processes occur. Examples are: the relation between cross-sections for inelastic and superelastic electron collisions, between rates of evaporation from and condensation on a surface, etcetera.

Absorption and Emission of Radiation: As an example of the principle of detailed balancing we cite an astrophysically important illustration.



The absorption of light by a quantum-mechanical oscillator is stimulated by an incident light quantum, but emission may be spontaneous or stimulated. We have in detailed balancing:

$$\underbrace{\text{Pure absorption} - \text{Stimulated emission}}_{\text{Effective absorption}} = \text{Spontaneous emission.}$$

Our principle states there is a detailed balancing between these processes in a properly isolated assembly in thermal equilibrium. This gives us an equation between the three transition probabilities (i.e. the Einstein coefficients B_{12}, B_{21}, and A_{21}) and the thermal (black-body) radiation intensity. This equation has to be identically satisfied for all values of the temperature. The branch point in the development of modern physical science, where quantum theory departed from classical, occurred when Einstein in 1917 was obliged to introduce stimulated emissions in order to reproduce Planck's law of black-body radiation. For reference we reproduce here the three important laws of black-body radiation. $B(\nu)\,d\nu$ is the radiant energy emitted per unit time in the frequency range ν to $\nu + d\nu$ by a black-body in thermal equilibrium at temperature T:

Rayleigh-Jeans ($h\nu \ll kT$)	Planck (all ν, T)	Wien ($h\nu \gg kT$)
$B(\nu)d\nu = \frac{2\nu^2 kT}{c^2} d\nu$	$\frac{2h\nu^3}{c^2} \frac{d\nu}{e^{h\nu/kT}-1}$	$\frac{2h\nu^3}{c^2} e^{-h\nu/kT}$

Balance between the three processes is obtained at all T and ν, if Planck's formula is used. In the case of $h\nu \gg kT$ the rate of stimulated emission becomes very low. We can then neglect it fully and restore exact balance between the absorption and spontaneous emission only, by using Wien's law for the thermal radiation. This is the non-classical? case.

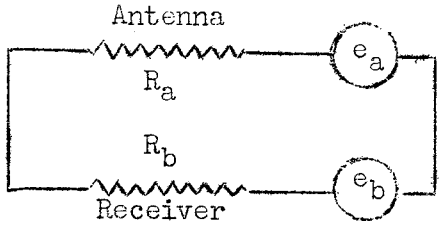
In Radio Astronomy invariably $h\nu \ll kT$. (e.g. at $T = 10^{4^\circ}$ K. and $\lambda = 1m$, or $\nu = 3 \cdot 10^8$ sec^{-1}, $\frac{h\nu}{kT} = 1.4 \times 10^{-6}$.) The rates of absorption and stimulated emission, which differ only by the Boltzmann factor $e^{h\nu/kT}$, then are nearly equal. The rate of effective absorption is exactly $(e^{h\nu/kT} - 1)$ times, i.e. in good approximation $\frac{h\nu}{kT}$ times, the rate of pure absorption. When this approximation is made the exact balance can be restored by replacing Planck's law by the classical law of Rayleigh-Jeans. The equation then is

Absorption = Emission.

This is the classical expression of the principle of detailed balancing in radiative processes. It was first stated by Kirchhoff. This is all we need, but whenever quantum processes are involved it should be remembered that the first member in this formula denotes $\frac{h\nu}{kT}$ times the pure absorption, where T is the temperature that governs the ratio of the number of atoms in upper and lower level.

Applications to Radio Astronomy:

(1) Noise emf of a resistor. A radio telescope may be considered as two resistive elements connected in a closed circuit, as follows:



We know that when a current flows through any resistor, energy is dissipated as heat. The inverse process is that a hot resistor can produce a current. Since currents of any frequency are dissipated, the produced current contains all frequencies, so it has the character of a noise. By a theorem of Nyquist, any resistor R can be regarded as a noise generator of average power

$$\frac{(\text{emf})^2}{R} = 4kT\,\Delta\nu \qquad \therefore\ (\text{emf})^2 = 4kT\,\Delta\nu \cdot R.$$

(Reference: Burgess, Proc. Phys. Soc., 53, 293, 722, 1941; Fürth and McDonald, Proc. Phys. Soc. 59, 388, 1947.)

(2) Two Resistors in a Circuit. Consider the circuit shown above. The total emf is $e_a + e_b$, so the total current is

$$i = \frac{e_a + e_b}{R_a + R_b}$$

and the total power dissipated in the circuit is

$$P = \frac{\overline{e_a^2} + \overline{e_b^2}}{(R_a + R_b)^2} \cdot (R_a + R_b),$$

for the time average of $2e_ae_b = 0$. This power consists of four terms. For instance, one term contains $\overline{e_a^2}$ in the numerator and R_b in the second factor; this term signifies the power generated in resistor a and dissipated in resistor b.

Thus by Nyquist's formula,

$$P_{a\to b} = 4kT\,\Delta\nu \cdot \frac{R_a \cdot R_b}{(R_a + R_b)^2} = P_{b\to a}$$

This equality follows also from the principle of detailed balancing. Similarly

$$P_{a\to a} = 4kT\,\Delta\nu\,\frac{R_a^2}{(R_a + R_b)^2} \cdot$$

and

$$P_{b\to b} = 4kT\,\Delta\nu\,\frac{R_b^2}{(R_a + R_b)^2}$$

For a matched set ($R_a = R_b$) all four parts are $kT\,\Delta\nu$.

(3) Radiation resistance of an Antenna. It remains to interpret these formal results. The power generated in the receiver and dissipated in the antenna is clear enough, but what meaning can be given to $P_{b\to a}$? This problem can be solved by imagining the antenna-receiver system to be enclosed in an insulating box (a "thermostat"). In the case of an atomic process where we know the rate and cross-section of a process,
we applied the principle of detailed balancing to find the cross-section of the reciprocal process. In an entirely analogous manner we can solve the present problem. The counterpart of emission is of course the absorption of radiation coming from the wall of the thermostat. Thus we have for matched ideal receiver and antenna:

Emission of energy produced in antenna and radiated out to space is $P_{em.} = 4kT\,\Delta\nu\,\frac{R^2}{(2R)^2} = kT\,\Delta\nu$

Absorption of energy emitted (or reflected) from wall of thermostat is $P_{abs.} = \frac{1}{2}B(\nu)\Delta\nu\cdot 4\pi\bar{A} = 4\pi\bar{A}\,\frac{kT\Delta\nu}{\lambda^2}$

Equating them according to the principle of detailed balancing, we obtain the two important results:
by the Rayleigh-Jeans law.

$$P_{abs} = kT\,\Delta\nu$$

$$\bar{A} = \frac{\lambda^2}{4\pi}$$

The second formula has already been quoted in our discussion of antenna properties. (p.27)

3.2 Equivalent Temperatures

If the actual systems in astronomy were in perfect thermodynamical equilibrium, there would not be much use for astrophysics. Just in virtue of the fact that the actual situation differs from the fictitious experiment in a thermostat it becomes necessary to study the rates of the individual processes. Yet the problems often become more translucent by comparing the rate of individual processes with what would happen in thermal equilibrium. This is done by introducing equivalent temperatures. Four of them in use in astronomy are shown in the following table. They may be used as equivalent to the quantities in the first column that have been defined earlier. (Some people prefer to use the original quantities and some prefer the equivalent temperatures.)

EQUIVALENT TEMPERATURES

Law of Rayleigh-Jeans:

Blackbody radiation is unpolarized, isotropic, with brightness:

$$B(\nu) = \frac{2}{\lambda^2} kT$$

Quantity	Definition of equivalent T	Name	Reasoning
N (noise figure)	$T_i = (N-1)290^\circ$	equivalent temperature of input resistance of receiver	if noise figure were 2, then $T_i = 290^\circ$ = room temperature. $k \cdot 290^\circ = 4.00 \times 10^{-21}$ watt-sec.
P_{abs} (absorbed power)	$T_A = \frac{P_{abs}}{k \Delta\nu}$ $(P_{abs} = kT_A \cdot \Delta\nu)$	antenna temperature (= result of a measurement)	if whole sky radiated as black body at temp. T_A, then this power would be absorbed.
$B(\nu)$ (brightness)	$T_b = \frac{B(\nu) \cdot \lambda^2}{2k}$ $(B(\nu) = \frac{2}{\lambda^2} kT_b)$	brightness temperature of a celestial body	a black body has to have the temp. T_b to match the brightness of the actual body at that frequency.
$S(\nu)$ (flux density)	$T_a = \frac{S(\nu) \cdot \lambda^2}{2k\Omega_s}$ $(S(\nu) = \frac{2kT_a\Omega_s}{\lambda^2})$	apparent temperature (usually referring to the sun)	a black body of the size of the solar disc Ω_S has to have the temp. T_a to match the flux density from the sun at that frequency. (Ω_s = 0.22 sq. deg. = .000068 sterad.)

$$P_{ext}/P_{int} = T_A/T_i$$

general: $P_{abs} = \frac{1}{2}\int AB\Delta\nu \, d\Omega \longrightarrow T_A = \frac{1}{\lambda^2}\int AT_b d\Omega$

homogeneously filled: $P_{abs} = \frac{\lambda^2}{2} B \Delta\nu \longrightarrow T_A = T_b$

sun in center of beam: $P_{abs} = \frac{\lambda^2}{8\pi} G_o \cdot S \Delta\nu \longrightarrow T_A = \frac{G_o}{185000} T_a$

Problems on observational methods.

1- In making a measurement of solar noise intensity at 10.7 cm. (2800 Mc/s), Covington measures the receiver output under the following three conditions:
(i) antenna pointed to empty sky, whose equivalent temperature at this frequency is about 50° K.
(ii) dipole enclosed in a small box at earth temperature (290° K).
(iii) antenna pointed to sun.

If the three readings are 8, 52, and 60 units respectively, find
(a) the antenna temperature when pointed to the sun,
(b) the apparent temperature of the sun (antenna gain = 700)
(c) the flux density from the sun.

2- At the longer wavelengths, solar noise can only be seen above the background of galactic noise if an antenna of sufficient directivity is used.
What antenna gain is necessary at a wavelength of 5 m. so that the meter reading with the antenna pointed towards the sun will be double that due to the sky alone, in a region of the sky with a brightness temperature of 10,000°K, given that the apparent temperature of the quiet sun is 1.7×10^{6}°K.
What size of paraboloid is required to realize this gain?
Show also that if the receiver noise figure is 3, the internal noise is negligible.

3- Ryle and Vonberg's receiver had the following characteristics:

noise figure	11 decibels,
overall bandwidth	4 Mc/s,
time constant	0.25 sec.

What is the fluctuation in output reading, due to internal noise, expressed in degrees?
The measured value was found to be 6.5 ± 2°K.

4- Bolton and Westfold, in their 100 Mc/s galactic noise survey, used an array of nine Yagis, in three groups of three, one wavelength apart. The directional pattern of this antenna is to be obtained, in the planes of the electric and magnetic vectors.
The pattern for an individual Yagi is as tabulated below. This is to be multiplied by an array factor, which expresses the fact that the final pattern is the resultant of three vectors of equal amplitudes, and different phases. These phase differences are obtained from the differing path lengths traveled in the direction of propagation by radiation received by the three elements of the array.

Note that, in either plane, only the appropriate cross-section of the array need be considered.

Obtain the general shape of the directional pattern in each plane, the half-widths, approximate sizes and positions of side lobes, and estimate the antenna gain.

angle from direction of maximum sensitivity (degrees)	power sensitivity	
	E-plane	H-plane
0	1.0	1.0
20	.85	.93
40	.4	.65
60	0	.3
80	.1	0
100	.1	.1
120	0	.1
140	.05	0
160	.07	.06
180	.1	.1

References

A. E. Covington, Canadian Nat. Res. Counc. Report ERB-242, April 1950.

M. Ryle and D. D. Vonberg, Proc. Roy. Soc. A 193, 117 (1948).

J. G. Bolton and K. C. Westfold, Aust. J. Sci. Res. A, 3, 19 (1950).

Solutions:

1 (a) 334° K, of which 50° is due t o the background.
(b) $75{,}000^\circ$ K.
(c) 123×10^{-22} watt m^{-2} $(c/s)^{-1}$

2. 3100 sq. meters.

3. 3.4° K.

4. Half-width in each plane $\approx 17^\circ$.
Side lobes at 90°, 140°, 180° in E-plane
and 60°, 100°, 180° in H-plane
Gain = 71.

3.3 Thermal emission from gaseous and solid bodies.

A radio telescope measures the brightness $B(\nu)$ or the flux $S(\nu)$ of a radiant source. We have seen that the nature of our detecting and measuring equipment renders it expedient to express these data as equivalent temperatures T_b or T_a. We must now consider two alternatives:

(a) The noise production is non-thermal in origin, e.g. produced by electrical discharge in gases, or by plasma oscillations. Then the T_a or T_b is a parameter without physical meaning as a thermodynamical temperature. The noise produced by solar flares is an example of this type.

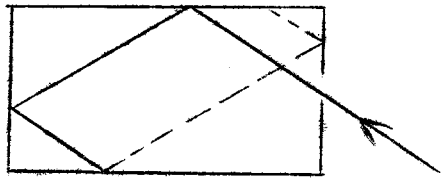
(b) The noise is thermal in origin. If the source is a black-body radiator at a homogeneous temperature then T_b is its actual temperature. But most often the source is not black or not at a uniform temperature. Our present task is to see how the measured temperature T_b is related to the true physical conditions in the astronomical source.

A Simple Black-Body--the Oven. If radiation of intensity B falls on a solid surface of reflectivity r, then a fraction $B \cdot r$ is reflected and a fraction $B \cdot p$ absorbed, where $p = 1 - r$. By definition the surface is called black if $r = 0$, so $p = 1$. Let the temperature of the surface be T. By supposing the surface enclosed in a thermostat at temperature T and applying the principle of detailed balancing we find

$$B_{\text{emitted}} = p \cdot B(T)$$

This is Kirchhoff's law and the universal function B(T) is Planck's radiation intensity. The formula shows that a black body ($p = 1$) would emit exactly this radiation intensity. That is why it is often called black-body radiation.

It is impossible to find a perfectly black surface. We can, however, construct a close approximation to a black body by making an oven (as in the figure) of imperfectly absorbing walls. Successive reflections from the walls serve to produce blackness and making the aperture narrow prevents escape of



diffusely reflected radiation. So radiation entering the aperture will be almost completely absorbed; inversely, the radiation issuing from the aperture if all walls are at the temperature T will have the intensity B(T).

Let us now calculate the brightness temperature for an oven of which different walls have different temperatures. Let wall n absorb the fraction p_n of the energy in any ray that we send into the aperture. The oven as a whole is still black, so

$$p_1 + p_2 + p_3 + \ldots = 1 \quad .$$

By virtue of its own temperature T_n each wall now emits (partially via reflection from the other walls) the intensity $p_n B(T_n)$ through the aperture in the specified direction. The total emergent radiation will be

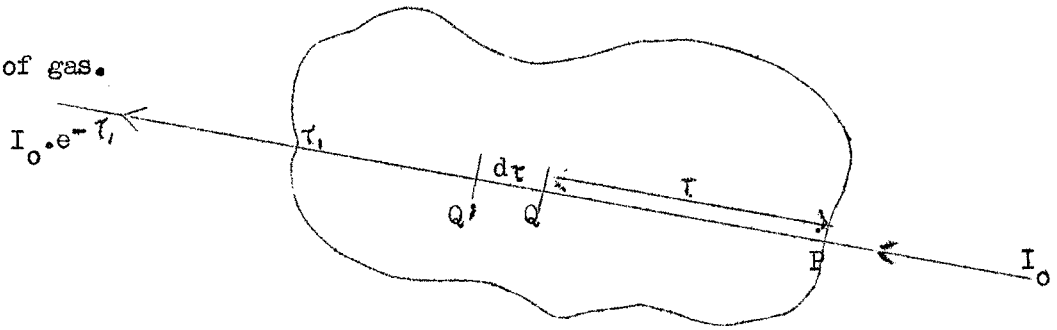
$$B_{emitted} = p_1 B(T_1) + p_2 B(T_2) + p_3 B(T_3) + \ldots$$

or, if the conditions for the Rayleigh-Jeans law, ($B \sim T$), are fulfilled:

$$T_b = p_1 T_1 + p_2 T_2 + p_3 T_3 + \ldots$$

The stratified medium--a Simple Case of Non-Uniformity. Let us next consider a mass of gas in which the temperature varies with depth. Analogously to the oven, an effective black body may be built up from many imperfectly absorbing

layers of gas.



Define the optical depth τ at the point Q as that depth at which an incident beam of intensity I_o would be reduced to the amount $I_o e^{-\tau}$ while travelling the distance PQ through the medium. If QQ' corresponds to a difference $d\tau$, then the fraction of I_o absorbed in QQ' is

$$d(e^{-\tau}) = e^{-\tau} d\tau = p.$$

Suppose now the gas is radiating, and let the beam travel in the reverse path. The brightness B at the point P is just

$$B = \int_o^{\tau_1} B(T\tau) d(e^{-\tau}) = \int_o^{\tau_1} B(T\tau) e^{-\tau} d\tau.$$

(cf. our discussion of the radiating oven!) When, as at radio frequencies the Rayleigh-Jeans law applies, we can write for the brightness temperature of the gas

$$T_b = \int_o^{\tau_1} T(\tau) e^{-\tau} d\tau.$$

This equation applies to a gaseous medium which is neither black nor at constant temperature. The following special cases are important:

(a) Black-body at uniform temperature T throughout: $T_b = T$.

(b) Non-black body at uniform temperature T throughout: $T_b = T(1 - e^{-\tau})$

(c) Body black, i.e. of infinite optical depth, but not of homogeneous temperature. In this case T_b is a weighted mean of the temperatures in the successive layers of the gas. If the temperature gradient is everywhere small, then 80% of the radiation is emitted between

$$e^{-\tau_{upper}} = 0.1 \quad \text{i.e.} \quad \tau = 2.3$$

and

$$e^{-\tau_{lower}} = 0.9 \quad \text{i.e.} \quad \tau = 0.1.$$

The "effective optical depth to which we can see" in a radiating gas is $\tau = 1$. When the layer $0.1 < \tau < 2.3$ corresponds to a narrow zone of linear depth we can frequently assume that it has a constant temperature T corresponding to the actual temperature at this effective depth. Then $T_b = T$.

The Equation of radiative transfer. We have derived the intensity of the emitted radiation from general thermodynamical concepts. No need exists to study in greater detail the emission and absorption processes that lead to this net result, or to study the intensity at any place inside the gas. However, most authors prefer to give this more complete treatment by means of the equation of radiative transfer.

Consider a small cylindrical element of cross-section $d\sigma$ and height ds in the medium. The difference in intensity of radiation in the frequency limits ν to $\nu + d\nu$ crossing the two faces normally in time dt and confined to an element of solid angle $d\Omega$, is

$$\frac{dI_\nu}{ds} ds d\nu d\sigma d\Omega dt.$$

This quantity must be equal to the difference between the energy absorbed by the medium from the incident beam of intensity I_ν:

$$k_\nu ds \cdot I_\nu d\nu d\sigma d\Omega dt.$$

(where $k_\nu \equiv$ the absorption coefficient per unit length)

and the energy emitted by the matter within the cylinder:

$$j_\nu d\nu ds d\sigma d\Omega dt.$$

(where j_ν is the emissivity per unit volume)

Hence the conservation of radiant energy through the cylinder requires

$$\frac{dI_\nu}{ds} = -\kappa_\nu I_\nu + j_\nu$$

or

$$\frac{dI_\nu}{d\tau} = -I_\nu + J_\nu \qquad \text{...(EQUATION OF RADIATIVE TRANSFER)}$$

where $d\tau = \kappa_\nu ds$ and $J_\nu \equiv \frac{j_\nu}{\kappa_\nu}$ is known as the source function (Ergiebigkeit). For the cylindrical element we have by Kirchhoff's law

$$j_\nu = \kappa_\nu \cdot B_\nu(T), \qquad \text{so} \qquad J_\nu = B_\nu(T)$$

$B_\nu(T)$ is of course the Planck Function, and can in radio astronomy applications be replaced by the Rayleigh-Jeans expression. Assuming that T is a function of τ, we shall write $B_\nu(T)$ as $B_\nu(\tau)$. Further we omit the index ν, still with the understanding that the equation refers to a particular frequency. The equation of transfer along a ray through the mass of gas now has the form:

$$\frac{dI(\tau)}{d\tau} = -I(\tau) + B(\tau).$$

It admits the integral

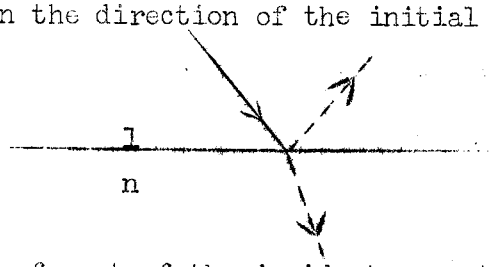
$$I(\tau) = \int_0^\tau B(t)e^{t-\tau}dt + I(o)e^{-\tau}$$

The boundary condition is obvious. In this formulation τ increases in the direction in which the radiation is going. This is the reverse convention from the previous section. Making the change, we find the formula for I(o) that was derived earlier.

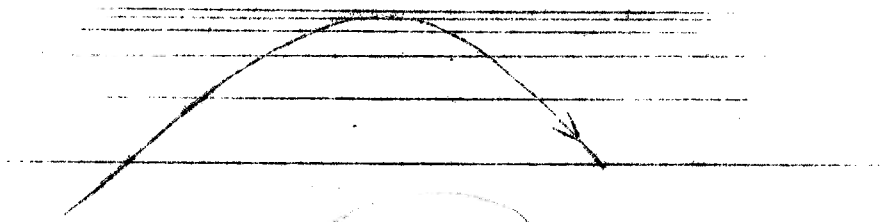
Reference: Milne, Handbuch der Astrophysik, III, 1.

Influence of Variable Index of Refraction. An index of refraction which differs from unity (empty space), or which is not constant, hardly ever occurs in astronomical problems, but at the wave-lengths used in radio astronomy refraction must be considered. There are two important effects on the optical path of a pencil of light, resulting from a variation of the index of refraction n of a medium:

(a) A change in the direction of the initial ray, following Snell's law:

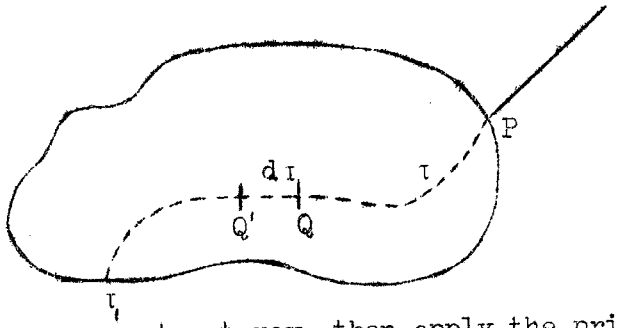


(b) Reflection of part of the incident ray at the interface. The immediate consequency of an index of refraction n differing from unity, is to change the phase velocity c into c/n. Reflection at a boundary is reduced if there is a continuous change in the refractive index. (This is the principle behind anti-reflection coating of lenses and prisms.) When the change in refractive index is gradual, then the deviation of a ray is likewise continuous.



Examples of this are mirages and the "reflection" of radio waves by the earth's ionosphere. The refraction of radio frequency electromagnetic waves in the solar corona is very important, as we shall see.

Let us now see how refraction of the ray will affect the emergent radiation from a mass of gas.



As before, we shall trace an entrant ray, then apply the principle of detailed balancing. The definition of optical depth remains unchanged. From a beam following the ray path PQQ', a fraction $d(e^{-\tau}) = e^{-\tau}d\tau$ is subtracted in the layer QQ' of optical thickness $d\tau$. Then if the mass of gas is enclosed in a thermostat, B(T) is the intensity of the pencil entering at P through unit cross-section at normal incidence, within unit solid angle and the energy

$$e^{-\tau}d\tau \cdot B(T)$$

is absorbed within the layer $d\tau$. It is immaterial what is the solid angle or cross section of the pencil as it reaches Q. Inversely, the energy emitted from QQ' and emerging at P is

$$B(T)e^{-\tau}d\tau.$$

While this derivation holds only for an isothermal gas mass, the obvious generalization is to replace B(T) by $B\{T(\tau)\}$. The result is identical with our previous one, and so the formula for T_b remains

$$T_b = \int_0^{\tau'} T(\tau)e^{-\tau}d\tau.$$

Again, a more complete derivation may be made through the equation of radiative transfer. It can be shown (see Milne) that the equation of transfer takes on the form, in a medium of refractive index n,

$$\frac{d(I/n^2)}{d\tau} = -\left(\frac{I}{n^2}\right) + B.$$

The intensity of thermal radiation inside such a medium is $n^2 \cdot B(T)$.

Emission from a solid body. A body is black if it absorbs all incident radiation: it is both non-reflecting and non-transparent. The second condition is fulfilled for the astronomical solid bodies. So, using at once the law of Rayleigh-Jeans,

$$T_b = (1 - r)\, T.$$

The reflectivity may be derived from Maxwell's theory (Fresnel's formulae) for a smooth surface and from experiment for a rough surface. The value of T refers to the layers where any incident radiation would be absorbed. Since a body like the moon has a temperature gradient in its surface layers it is important to make some estimate of how deep below the surface those layers occur.

This may be done from Maxwell's theory. A wave traveling through a medium with non-zero and non-infinite conductivity sets up currents that produce Joule heat, which entails absorption of the wave. The index of refraction for a wave of circular frequency ω, is

$$m = \sqrt{\varepsilon - \frac{4\pi i \sigma}{\omega}} = \sqrt{\varepsilon - i\,\eta}$$

Here ε is the inductive capacity of the medium and σ its conductivity. Now consider a plane wave of amplitude E_o incident normally on the surface of a medium; the amplitude at the depth z is

$$E = E_o e^{-ikmz}. \quad \left(k = \frac{2\pi}{\lambda} = \frac{\omega}{c}\right)$$

Since the intensity is proportional to the square of the wave amplitude,

$$I \sim e^{-Re(2imkz)}$$

But by the definition of the absorption coefficient, $\varkappa$,

$$I \sim e^{-\varkappa z}.$$

Therefore

$$\varkappa = Re(2imk).$$

The inverse value, $1/\varkappa$, is the length over which the intensity drops by a factor 1/e, while $2/\varkappa$ is the length over which the amplitude drops by a factor 1/e. The latter value is known as the skin depth, because it signifies the depth to which waves penetrate a conducting body. But the use of this term is conventionally restricted to the case $\eta \gg 1$, where the skin depth,

$$\frac{2}{\varkappa} = \frac{\lambda}{2\pi} \cdot \frac{1}{\mathrm{Re(im)}} = \frac{\lambda}{2\pi} \cdot \left(\frac{2}{\eta}\right)^{\frac{1}{2}} = \frac{\lambda}{\pi\sqrt{2\eta}},$$

is much smaller than the wave length. For poor conductors, on the other hand, that have a small value of η, we have to go back to the original formula for $\varkappa$. The waves can penetrate much farther than one wave length. The following table gives some examples of $1/\varkappa$ and thus gives an impression of the depth from which the thermal radiation from various bodies is emitted.

Tungsten lamp filament, visible light	0.5 μ	= λ
Hot water boiler, Infra red (2μ) radiation	0.2 μ	= 10^{-1} λ
Copper plate, 1cm. radiation	0.3 μ	= 10^{-5} λ
Copper plate, 10m. radiation	10 μ	= 10^{-7} λ
Sea water, 1cm. radiation	0.3 cm.	= .3 λ
Wet earth, 1cm. radiation	3 cm.	= 3 λ
Moon, 1cm. radiation	40 cm.	= 40 λ

Chapter 4. RADIOEMISSION BY THE MOON (Henry J. Smith)

The moon, though a cool body, is not completely cold. It is self-luminous, radiating measurably in the infra-red and centimeter spectral regions. The moon's heat of course is the result of insolation. Any particular portion of the moon's face is thus subject to periodic heating and cooling. Due to poor heat conduction, the temperature varies differently in different layers below the surface. And waves at different frequencies come effectively from different layers. In this way a study of the lunar radiation at different frequencies can disclose some of the conditions and properties of the lunar surface.

Infra-red observations:--Prior to 1945 the most thorough investigation of the moon's radiation was carried out by Pettit and Nicholson. Using the large Mt. Wilson reflectors they observed the moon with a thermocouple through the 8-14 micron window in the absorption spectrum of the earth's atmosphere. They found that the total radiation varies in intensity with lunar phase, and in phase with it. (Pettit, Ap J 81,17,(1935). The measured temperature of the subsolar point on the center of the disk was 407° K. This is 10% larger than the theoretical value 374° K, based on their deduced reflectivity of the moon and the observed solar constant. The discrepancy is probably due to inaccurate assumptions about the transmission properties of the earth's atmosphere. The temperature of the subsolar point on the dark side of the moon (i.e. the center of the new moon) they found to be 120° K. (Ap J 71,102,(1930). The accuracy of this latter measurement was low, owing to much decreased sensitivity of the thermocouple at temperatures around 100° K. Pettit and Nicholson

measured the variation of intensity across the full moon and concluded that limb-darkening follows the law

$$E = E_s \cos^{2/3}\rho$$

(where E_s is the sub-solar point intensity, and ρ the angle between the normal to the area considered and the direction to the sun). The directional emissivity of an ideal smooth, slowly rotating sphere without an atmosphere, exposed to sunlight, is given by the formula

$$E = E_s \cos\phi$$ Lambert's Law

The departure from this simple law was interpreted as due to the roughness of the moon's surface.

Micro-wave observations:--The first radio frequency measures of lunar radiation were made by Dicke and Beringer (Ap J 103 375, 1946) at a wavelength of 1.25 cm. They measured the emission from nearly full moon (phase $+18^\circ$), and found an effective black-body temperature T_b of 294° K.

Piddington and Minnett (Austr. J. Sci. Res. A 2, 63, 1949) have recently carried out an exhaustive study of lunar microwave radiation. Their equipment had the following characteristics:

Antenna: Parabolic, 44-inch diameter, equatorially mounted; half-power beam width of $0^\circ.75$; power gain, 44.8 decibels ($3 \cdot 10^4$);

Receiver: 15 Mc/s overall bandwidth at 1.25 cm. wavelength; response time 1 or 6 seconds;

Calibration: Comparison with a black body at room temperature, rotated 25 times per second into waveguide system;

Sensitivity: 1.2° K. in antenna temperature (corresponding to lunar temperature changes of about 8° K; see below).

Observing technique consisted in allowing the moon to trail across the antenna pattern near the meridian. The maximum scale reading thus obtained was proportional to the antenna temperature of lunar radiation averaged over the disk, together with telluric absorption and emission. The factor relating antenna temperature to apparent temperature, $185{,}000/G_o$ (where G_o is the antenna gain), is 6.2. The antenna directional pattern and the law of limb darkening were obtained independently; with these known Piddington and Minnett calculated a factor 0.94 to account for the fact that the edge portions of the moon are not observed with full antenna gain. Because the moon does not radiate like a perfect black body, the observed temperatures are reduced by a factor (1 - R), where R is the reflectivity:

$$R = \left(\frac{\varepsilon^{\frac{1}{2}} - 1}{\varepsilon^{\frac{1}{2}} + 1}\right)^2 .$$

A probable value of the dielectric constant ε is 5, so that (1 - R) = 0.9. By holding the antenna fixed in direction during a measure, the received atmospheric radiation was constant. Therefore the receiver output variation was due to the reception of lunar radiation as the moon passed through the beam. But as well as radiating itself, the atmosphere absorbs part of the lunar radiation. Piddington and Minnett allowed for this absorption by a method developed by Dicke et al. (PR. 70, 340, 1946). At 1.25 cm. cosmic radio noise is negligible. In this manner these investigators secured partial coverage of three lunations in April-June 1948.

The result of these observations indicates that the equatorial temperature of the moon is of the following form:

$$T = 249^\circ + 52.0^\circ \cos(\Omega t + 45^\circ).$$

Piddington and Minnett claim an accuracy in these temperatures of ± 5%.

Interpretation:--The moon's temperature at 1.25 cm. varies in a range of 197° to 301°, and lags 45° behind the lunar phase. The infra-red temperature on the other hand varied from 120° to 391°, in phase with lunar phase. The apparent discrepancy between microwave and infra-red observations must be due to the transparency of the lunar surface to the 1.25 cm. waves. Piddington and Minnett applied the classical theory of heat conduction in a semi-infinite medium subject to periodic heating of its exposed face. However the assumption of constant ratio of electromagnetic to thermal absorption coefficients failed to represent the observed phase lag or the ratio of the emplitudes of surface and internal temperature variation. Following a suggestion of Jaeger and Harper, Piddington and Minnett next considered a solid uniform model covered with a thin layer of poorly conducting dust. In this way the observed phase lag was accurately accounted for by a millimeter thick layer of pumice-like material lying on a lava base. Meteoric accretion could easily account for the origin of such a film. Moreover Jaeger and Harper (Nature, No. 4233, 1026, 16 Dec. 1950) have shown that Pettit's infra-red observations can be interpreted by such a model surface of the moon. However Lettau (unpublished) has obtained a better fit with a model in which the composition varies continuously with depth.

Chapter 5. THE SUN STUDIED BY OPTICAL (NON-RADIO) MEANS.

5.1 General information (H.C.van de Hulst.)

The sun will be the subject of many of the following pages. All observed radiation (radio and optical radiation) comes from its outer layers, photosphere, chromosphere and corona. Before discussing the properties of these layers in detail we might describe the sun as a whole as follows.

The sun is a mass of gas that is to a high approximation radially symmetric and stationary. Density, temperature, etc., are functions of r that do not change with direction or time. There are small motions (convection) in the core and below the photosphere; these belong to the stationary picture like the granulation and the spicules in the chromosphere that may be secondary effects of this convection. The density decreases all the way out. The temperature decreases outward throughout the inner body and photosphere; it may be approximately constant in the lower chromosphere and rises sharply further out.

Two agents cause deviations from this spherical stationary picture. Connected with the sun's rotation many phenomena on the sun show a dependence on latitude; large scale circulation currents are certainly behind this dependence. Some not-understood cause produces an approximate 22-year cycle in many solar phenomena; the half-cycle of 11 years is more conspicuous. Theoretically it seems probable that both 22-year cycle and solar rotation have to do with large-scale magnetic effects. A fact is that the latitude-dependence of any phenomenon changes during the cycle so that in practice we have to study both dependences together. Of course, we can study only the superficial phenomena like the wandering of the sun spot zones or the changing form of the corona.

All phenomena of solar activity obey statistical laws and all go through maximum activity during the same years (1937, 1947, 1958?). Individually they seem unpredictable.

A first survey.--The following table has not very critically been compiled from various sources. It neglects all changes with time and latitude and serves to show the typical differences between various layers, not the exact numbers,

	h(km)	$r/R_\odot$	log N_e	T	hydrogen is
outer corona	1400 000	3	5.3	1000 000	ionized
medium corona	700 000	2	6.2	1000 000	
inner corona	140 000	1.2	7.6	1000 000	
	21 000	1.03	8.2	700 000	
upper chromosphere	14 000	1.02	8.7	50 000	ionized
	7 000	1.01	9.0	25 000	
lower chromosphere	5 000	1.007	9.2	5 000	neutral
	700	1.001	11.2	4 500	
sun's limb, τ = 0.002	0	1.000	12.0	4 500	neutral
.01	-70	.9999	12.4	4 500	
.1	-200	.9997	12.9	5 200	
photosphere 1	-390	.9995	14.0	6 500	
2	-430	.9994	14.5	7 300	ionized
3	-450	.9993	15.0	8 000	

Conversion: $R_\odot$ = 700 000 km, subtending 16' from earth
0.001 $R_\odot$ corresponds closely to 1 second of arc.

A layer of the photosphere is usually indicated by its optical depth (either some defined average over the visual region or the τ at a given wave length). The continuous light from the sun comes from the photosphere. Also the Fraunhofer lines are formed there, most of them in its upper part, which corresponds to the old concept of reversing layer. A layer in the chromosphere is usually indicated by its height in kilometers above a certain zero point. As zero point the sharp limb of the photosphere is chosen; this is the

layer for which the optical depth of a tangential ray is just 1; for a radial ray it is 0.002. A layer in the corona is usually indicated by the value of $r/R_\odot$, further denoted simply by r.

Scale height.--The conspicuous difference between photosphere, chromosphere and corona is in their vertical extension. A quantitative expression for the density gradient is the scale-height H. It is defined by setting the local density proportional to $e^{-h/H}$, so H is the height we have to go in order to find the density changed by a factor e. In formula:

$$\frac{1}{H} = -\frac{1}{N}\frac{dN}{dh}$$

where
H = scale height
N = number of atoms per unit volume

If the gas is in thermal equilibrium the scale height follows the barometric formula:

$$N \sim e^{-mgh/KT} = e^{-\mu gh/RT}$$

$$\text{so } H = \frac{RT}{\mu g}$$

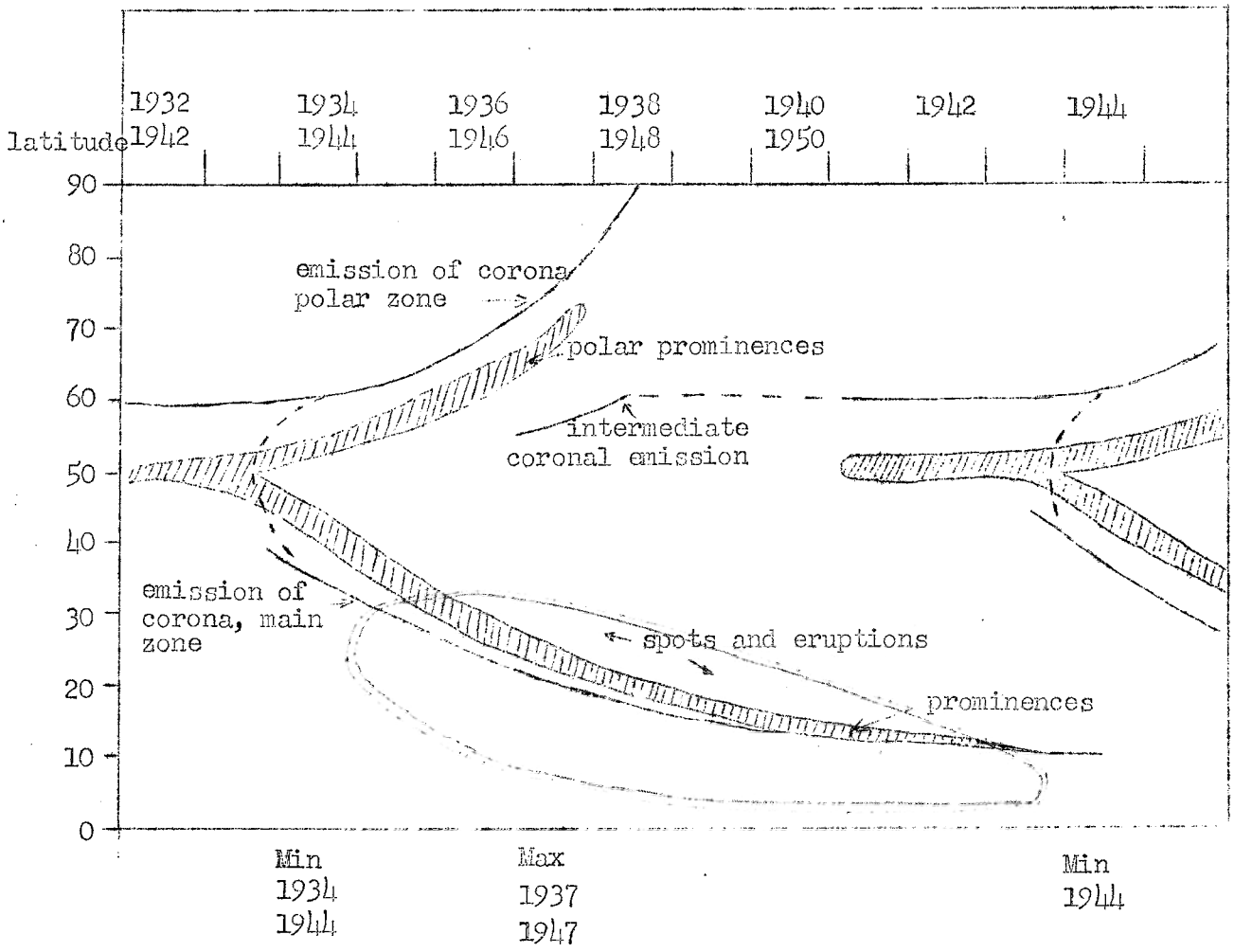
where
h = height of an arbitrary level
T = temperature ^oK
m = mean mass of the molecule
K = Boltzman's constant
g = gravitational acceleration
μ = mean molecular weight
R = gas constant = 8.31×10^7

It is thus seen that the enormous differences in H are to a large extent due to the differences in T. For the corona H = 70 000 km in the lower layers but increases outward due to the changes in g. With the gravity at the sun's surface the equilibrium relation is:

$$H_{km} \cdot \mu = 0.030\ T$$

So the corona scale height with a probable value of $\mu = 0.65$ gives $T = 1.5 \times 10^6$. This is an upper limit (see later). The computation for the photosphere is the inverse. For its upper part we find H = 100 km from the known $T = 4500^\circ$ and an assumed $\mu = 1.3$. This value is in agreement with the sharpness of the sun's limb but a quantitative confirmation is not yet obtained (compare Lindblad's eclipse cinematography).

Activity zones.--The following figure gives a sketch of the latitudes at which various phenomena were concentrated during the last solar cycles. Again, this figure serves to convey the general pattern and should not be relied upon for accurate values. A critical study still reveals many doubts as to the exact interrelation of the activity zones found for various phenomena. Roughly, however, one may say that there is one main zone, the sunspot zone, drifting from 30° to 10°. In this zone (or at slightly higher latitudes) also the chromospheric phenomena, prominences and flares, are abundant. Prominences also occur in a zone where no spots are seen; this is sometimes called the polar zone but is by no means close to the pole. Coronal emission also exhibits an "equatorial" zone coinciding with the spot zone and a "polar" zone, coincident with or at slightly higher latitude than the "polar" prominence zone. There is a strong suggestion that both zones arise from a general type of traveling disturbance that reaches the surface of the sun a year before minimum phase and then spreads out in both directions. Also there may be a real polar zone of high electron density in the corona during minimum phase; this zone is confined to 10° from the pole. When such a diagram is plotted with the time down and the latitude right and left the actual spread in latitude of the spots in the main zone gives the appearance of a butterfly (Maunder).



Where does the radio-emission originate? For all we can tell radio waves just like light waves originate anywhere in the solar body. But they can reach us only from the layers that are partially transparent. The partially transparent layers for radio waves are the chromosphere and corona. Centimeter waves come from chromospheric layers and meter waves from the corona. That this is true is demonstrated most directly by the observations, for instance:

(a) during a total solar eclipse a large fraction (up to 50 per cent for meter waves) is left uncovered.

(b) Active spots have been located outside the visible disk by interferometer.

(c) Interferometer observation of the quiet sun on 60 cm (Steinberg) has given a distribution of brightness extending far outside the sun's disk.

Quite as stringent is the theoretical argument derived from the opacity of the chromosphere and corona. This opacity will be computed on the basis of Kramers' law in Chapter 6. Since the absorption coefficient contains the product $N_e^2 T^{-3/2}$, both N_e (the number of electrons per cm^3) and T (the temperature) have to be known well for each layer. The following sections will review the information available from optical observations on these points. As a preview here are the computed heights at which the optical depth along a radial ray is 1.

at 1 cm	h = 4000 km	at 1 m	h = 25000 km (r = 1.03)
at 10 cm	h = 8000 km	at 5 m	h =200000 km (r = 1.3)

Again, these values serve to give a rough idea; the exact values are changeable and subject to a critical discussion.

5.2 Optical investigations of the Chromosphere. (R.B. Dunn.)

The theory of the chromosphere (Menzel, Wildt, Giovanelli, Thomas, Miyamoto, Woolley and Allen, and others) is still in a floating state; therefore this section has to be mainly descriptive.

Observations. The methods of observing the sun in the visible and photographic light are fairly well known and need not be discussed in any great detail here. They fall into three broad classes depending on the feature to be observed. A. Eclipse techniques, B. Instruments whose final images are essentially integrated light and C. Instruments whose images are essentially monochromatic or spectral.

Class B includes sunspot cameras, refracting and reflecting telescopes using solar eyepieces, wedges, neutral filters and projection devices. The original coronagraph (13) less filter is in this class. Class C includes spectrographs, spectroheliographs (10), spectrohelioscopes, and interference filters (5). Integrated light studies enable observations of the sunspots, limb darkening, faculae, and granulation (all phenomena of the photosphere). Monochromatic studies include chromosphere, prominences, flocculi, flares, and filaments. The eclipse allows observations of the chromosphere and the corona. The chromosphere with spicules and prominences is seen red during an eclipse, because of the strength of H_α; the corona is seen white.

Information on the chromosphere thus is obtained mainly in two ways: (a) from flash spectra (with some competition from spectra taken outside eclipse). (b) by spectroheliographs (with growing competition from monochromatic filters). This information we shall now discuss.

Flash spectrum.--At the beginning and end of totality during a solar eclipse the spectrum of the limb is seen to suffer a "reversal" and to become an emission spectrum. Since the moon covers and uncovers the sun at the rate of 300 km/sec this happens in a flash. The flash spectrum is partially the same as the absorption spectrum of the disk. Some lines appear in the chromosphere and do not show on the disk due to their high ionization potential.

The most extensive list of lines in the flash spectrum is probably Mitchell's 1947 catalog of about 3500 spectral lines (1). He has taken ten eclipse spectra from which five have been studied. The table includes the wavelength, the intensity in the flash, intensity on the sun, element, height in km., and the electron potential. Other observers have given similar tables but this is

probably the most extensive and convenient. Its theoretical consequences have been studied by Wildt (2).

Hale (3) and others have managed to observe the flash spectrum outside eclipse by placing the slit tangentially to the limb under excellent seeing conditions.

Spectroheliograms and filter observations.--The absorption in the center of strong Fraunhofer lines is so great that the chromosphere is not entirely transparent. By taking a photograph of the sun in light of such a particular wave length, layers of the (lower) chromosphere are made visible.

Normally the spectroheliograms are taken in the H_α and H and K lines. These lines are the strongest in the chromosphere. D'Azambuja (4) has obtained spectroheliograms that include H, K_1, K_2, K_3, Ca 8542, D_3, He 10,830, the Balmer series of hydrogen, and others. The photographs are quite different in the different lines. The large active areas, "plages faculaires," usually can be identified between the various spectroheliograms. Bright and dark markings appear and disappear depending on the line. Motion pictures have been obtained of the motion of Ca flocculi by the McMath-Hulburt group. These films give the impression of a "crawly" appearance and indicate that the whole surface of the sun is in violent agitation. Fairly recently photographs have been taken with a narrow bandfilter centered on H-alpha. Such filters are still too wide (1.5 A) to compete with a spectroheliograph, for a good deal of photospheric light comes through. However, for observations at the limb this is an advantage. The (photospheric) limb of the sun is seen sharp and smooth but the chromosphere above it as an irregular fringe, about 10" high with a ragged edge. Some reproductions of photos from Lyot's movies are found in his extensive article on monochromatic filters (5).

Roberts of Climax (6) 1945 using a quartz interference filter attached to a coronagraph drew new attention to the "spicules" in the chromosphere. They are also seen on good eclipse plates. The whole chromosphere seems to consist of such spicules that stick out like grass blades (early observers compared the chromosphere with a burning prairie). Movies made by Roberts and by Lyot indicate that the spicules pop up and recede in a matter of minutes but it is uncertain whether this is real gas motion. They are 1" to 2" across and occur all around the edge except where more violent phenomena like prominences make them invisible. This fact and the rough agreement in size and duration suggests a physical relation with the granulae in the photosphere. Both phenomena are always present and remarkably constant in appearance; they are features of the undisturbed sun.

Chromospheric models and theory. The problem of the chromosphere is particularly interesting and many models have been introduced to explain it. Some use equilibrium concepts and others advocate dynamical support. The enormous change in temperature that must occur somewhere in the chromosphere presents a problem that exists neither in the photosphere nor in the corona; observational clues are hard to get and often ambiguous. Woolley and Allen (7) 1950 in a recent paper describe the observational evidence that a chromospheric model should fit. These conditions are quoted without change from their paper.

a. The far ultra-violet emission should equal the amount required to produce the minimum (zero sunspot) ionosphere.

b. This emission should equal the amount of energy conducted inwards from the corona.

c. The emission of thermal radio noise should fit quiet day observations....

d. The density scale height of the low chromosphere should fit eclipse observations of decrement of the intensity of low excitation lines with height.

e. The density scale height should fit eclipse observations of decrement of strong Balmer lines extending to considerable heights.

f. Electron densities should agree with eclipse values (1) from the intensity of the Balmer continuum and (2) from the resolution of the Balmer lines.

g. Electron temperatures should agree with eclipse observations of the Balmer continuum.

h. Kinetic temperature of atoms should agree with eclipse observations of line width.

i. Eclipse observations show that both high and low excitation lines (e.g. He and Fe) appear to originate at the same level. The model should explain this.

j. The density scale height should be not less than the gravitational scale height for the kinetic temperature of the level. If it is greater, there should be some reason for it.

k. The pressure should not increase outwards.

This list gives some idea of the complexities of the problem of determining a model chromosphere. Woolley and Allen (7) then proceed to construct a model that satisfies most of these requirements. The model has spherical symmetry and a single value of temperature and density at every height so it does not describe the spicules. A prominent feature of their model is the sharp division between the upper and lower chromosphere, which was also incorporated in the survey table of the preceding section (p.61). In the lower chromosphere hydrogen is neutral and the temperature is governed mainly by radiative equilibrium, so this part is practically isothermal at the T of the sun's surface

(= 4500°). This would mean a constant scale-height H of the order of 100 km. The fact that the observed H increases by 200 km for every 1000 km height is evidence for "mechanical support," which does not seem unlikely in view of the spicule motion. In the upper chromosphere hydrogen is ionized and the temperature is governed by heat conduction inwards from the corona. This part is virtually at constant pressure (it is not supported, so to say).

Temperatures of the chromosphere.--If this model of the undisturbed chromosphere were final, we could use the N_e and T to compute the emission by the quiet sun in the centimeter range. It is still far from that, for opinions are still confused, in particular about the temperatures. In fact, the radio-emission conversely gives the most reliable value of T now available. The following short table, inspired by a table presented by Woolley and Allen, lists the reasons for believing in a high (about 30000°) or a low (about 5000°) temperature. Item (h) has been added as well as the approximate height to which the arguments refer and my personal opinions on the value of the arguments.

Reasons for high temperature:

(a) Redman's widths of H and He lines; h = 1500 km (obs. fine; interpretation has possibly to be revised),

(b) Measured scale height; h = 500 to 8000 km (no value because of the likelihood of non-thermal support),

(c) Emission of He lines; h = up to 8000 km (more likely excited by uv radiation from corona; Miyamoto)

Reasons for low temperature:

(d) Radio in cm range; h = 6000 km (obs. fine; interpretation seems foolproof).

(e) Measured excitation temperature; 500 - 1000 km (probably all right)

(f) Solar ultra-violet emission as indicated by ionosphere; lower part (the theory of the ionosphere is not clear enough to give this argument so much weight as W & A do)

(g) Int. distribution in Balmer continuum; h = 500 km (observations unreliable)

(h) Intensity of visual continuum; h = 500 km (Zanstra; observations still weak) (8)

(i) Non-appearance of forbidden lines; lower part (probably all right; Wurm)

This cannot be called a critical survey but it shows that the radio observations are supported in giving a low T by several other arguments. None of these estimates refer to the upper chromosphere; this is virtually terra incognita.

Solar activity. The phenomena in disturbed regions, that are present throughout photosphere, chromosphere and corona, are collectively called solar activity. They affect many phenomena on the earth and intensive study of these regions may also be helpful in better understanding the undisturbed sun. They seem to be more violent in chromosphere and corona than in the photosphere; in line with this, the radioemission of the sun is much more violently affected by activity than (its) light. We go briefly through the multitude of terms applied to the active photosphere and chromosphere, leaving the corona for later.

Spots.--Deepseated cool regions in the photosphere. Intriguing structural details are seen on good photographs. They have strong magnetic fields (up to 4000 gauss) at the center and weak fields extending far beyond the visible spot (Kiepenheuer, unpublished). They often occur in groups that are the center and probably the cause of a large disturbed region.

Faculae.--Bright areas seen in white light most easily near limb. They seem to give the more precise outline of the active areas. In the monochromatic light of a strong Fraunhofer line they are more conspicuous and different in details. D'Azambuja calls them "plages faculaires" (the Americans have abbreviated it to plages and the colloquial term now seems to be plage areas!).

Bright and dark flocculi.--Detailed markings seen all over a spectroheliogram. Much coarser and much more susceptible to disturbance than the granulation in the photosphere. Some of these structures would be classed as prominences when seen at the limb. But the identification of types of disk markings with types of prominences is one of the greatest problems in studies of solar activity.

Filaments.--Long dark threads seen on spectroheliograms and sometimes extending over more than 60° of solar longitude. At the limb they are seen as prominences of various appearances. D'Azambuja has studied the filaments in great detail in his extensive paper (4). They are huge curtain-like structures that hang high above the chromosphere and are remarkably stationary until after many solar rotations they suddenly disappear. They also have been described as quiescent prominences but this name is misleading because the gas in these curtains may not be quiescent at all.

Prominences.--This is the collective name for everything sticking out of the ragged limb of the chromosphere on a spectroheliogram. Their emission-line spectrum is much brighter than the continuous spectrum of the corona so they also are seen on eclipse photographs. Movies give the strong suggestion that matter streams down from the corona through the prominences, sometimes to fixed centers of attraction. There is a great variety of forms. Many classifications have been proposed and likewise many interesting theories.

Eruptive prominences.--A fairly well defined type with accelerated motion outwards. Pettit (9) has catalogued 62 eruptive prominences and has concluded that few if any shoot up fast enough to escape from the sun.

Surges.--A type of small prominences shooting up and falling down along the same path; here suddenly gravitation seems to act the normal way. They are characteristic for the vicinity of a solar flare.

Solar flares. These phenomena have a special place because they are the most violent outcome of solar activity and very important both in radio astronomy and in solar-terrestrial relationships. Flares is a new name for Hale's (10) 1930 earlier "chromospheric eruptions". They are usually observed in monochromatic light but have been seen in integrated light on occasions. To show against the photosphere they must be extremely intense. The spectroheliograph and scope are used for flare study. The spectrohelioscope is probably better because continuous observation of line widths can be made during the flare. Recently the monochromatic filters have been applied to flare observation at the Harvard station at Climax. It is possible to observe fainter flares with this instrument than with the spectrohelioscope. Various observatories are operating flare patrols and sending their results to clearing stations like the Bureau of Standards at Washington.

A paper by Ellison (11) summarizes the data on flares. He used a spectrohelioscope in his observations and used effective line widths plotted against time to describe the flare. The essential characteristics of flares as outlined in his paper follow:

a. their areas lie within the range 0-5000 millionths of the sun's hemisphere...

b. with rare exceptions they are located within 10^5 km. of a visible spot.

c. ...they are normally confined to chromospheric levels....

d. they develop out of pre-existent "normal" bright hydrogen flocculi.

e. their rise to peak intensity may be extremely rapid ... followed by a slow decay.

f. the duration of the peak intensity is very short (less than 5 min.) though the total life of an intense flare may be hours.

g. their spectra exhibit bright reversals in the Balmer absorption lines of hydrogen, in neutral helium and in the lines of ionized calcium, iron, silicon, and a few other metallic atoms..., but there is much variation in detail.

h. in a few instances an increase in brightness of the continuous spectra has been recorded....

i. they are essentially stationary in height as well as in the horizontal plane....

j. a majority of the brighter flare patches eject prominences of the surge type....

k. the probability of a flare occuring varies with the type of associated sunspot....

l. small flares are most frequent during stages of rapid growth of a spot....

m. secondary geophysical effects arise from the emission of both wave and corpuscular radiation....

The description in note m is of great importance. It has long been recognized that the "magnetic storms" are strongest during solar activity. About 23 hours after the solar storm frequently aurora (northern lights)and records of magnetic storms are seen indicating strong correlations. Indications are that the flares are much stronger in the ultraviolet than previously expected. This ultraviolet radiation seems to excite areas of the corona and send a beam of corpuscular material out. Computed velocities correspond to the 23 hour interval between the solar activity and the terrestrial disturbances.

The possibility of investigating spectroscopically the corpuscular emission during a flare has been studied by Kahn (12) in a recent paper. Indications are that such study would be difficult.

References

1. Mitchell, Ap J 105, 1, 1947.
2. Wildt, R., Ap J 105, 36, 1947.
3. Hale, Ap J 30, 222, 1909.
4. D'Azambuja, Annales de L'Observatoire de Paris, 6, 1948.
5. Lyot, B., Annales d'Astrophysique, 7, 31, 1944.
6. Roberts, Ap J 101, 1945.
7. Woolley and Allen, MN of R.A.S. 110, 358, 1950.
8. Zanstra, Proc. Acad. Amsterdam and Observatory, Fall, 1950.
9. Pettit, As. Soc. of the Pac. 52, 172, 1940.
10. Hale, Ap J, 71, 75, 1930.
11. Ellison, MN of R.A.S., 109, 1, 1949.
12. Kahn, MN of R.A.S. 110, 477, 1950.
13. Lyot, B., MN of R.A.S. 99, 586, 1939.

Selected Illustrations

Prominences: see Lyot Ref. 13.

Filaments: see D'Azambuja Ref. 4.

Spots, granulation, butterfly diagrams: Menzel, Our Sun, Blakiston, 1949; Handbuch der Astrophysik, 4, Solar Syste

Flares: Dodson, Ap J 110, p. 382, 1949.

Chromosphere: see Lyot Ref. 5. and Roberts, Ref. 6.

Flash Spectra: Menzel, Lick Pub. #17 also Davidson & Stratton Mem. RAS 64.

5.3 Optical investigations of the corona. (H.L. Zirin)

I. Determination of the Electron Density.

The first calculations of the polarization and brightness distribution which would be produced by scattering by free electrons in the corona were made by Schuster (1). Later authors corrected these results, in particular, Minnaert (2). Baumbach compiled the observational data from eclipses and fitted it to the theory (3). The latest analysis has been carried out by van de Hulst (4). In the corona, it is easy to compute the electron density N_e because the coronal light is mainly a continuous spectrum, due to Thompson scattering. We must, however, distinguish between several superimposed sources of light, which are shown in Fig. 1 below.

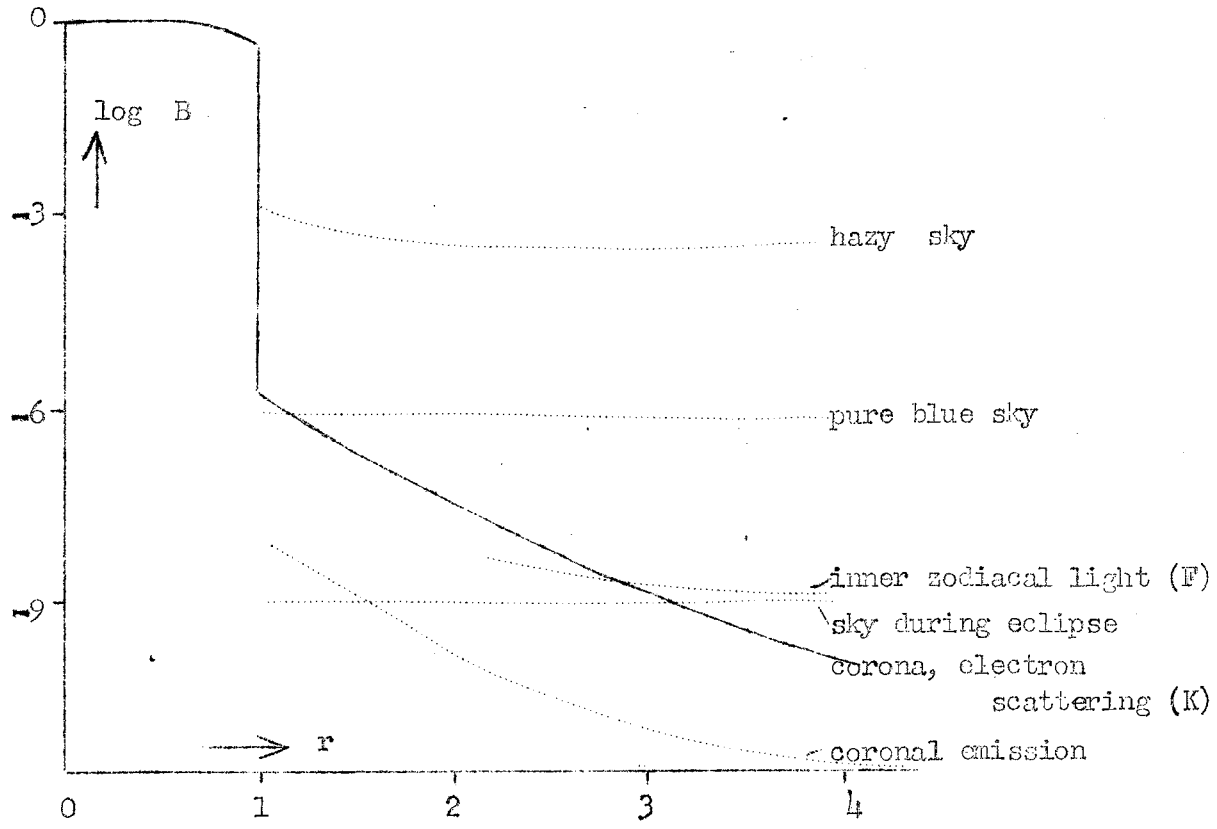


Fig. 1 .

We are interested in the range $1 < r < 2$. In this region the sky light and coronal emission may be neglected (we define the corona itself as the region above r = 1.03). We need therefore only account for the Fraunhofer component due to the inner zodiacal light.

Following Grotrian, the contribution of the real corona is designated the K-component, and the spurious corona, due to the inner zodiacal light, is called the F-corona. The former is continuous, shows coronal structure, and is strongly polarized, whereas the latter reproduces the Fraunhofer spectrum of the sun, is circularly symmetric and unpolarized. The observed brightness will therefore have to be multiplied by the factor

$$f = \frac{K}{F + K}$$

to obtain the true brightness of the K-component, and from this the electron density will be found.

The factor f is best determined by the fact that the Fraunhofer lines of the F-component are made shallow by the superposition of K. If r is the residual central intensity of a line in terms of the continuum, then we can derive the relation

$$1 - f = \frac{1 - r_{corona}}{1 - r_{disk}}$$

We can also use the other characteristics of the K and F components to separate them.

Photographs of varying exposure duration (0.1 to 30 sec.) give us the relative brightness distribution in the corona, which varies through a factor of 1000. Then the solution of two integral equations, corresponding to the two degrees of polarization of the scattered light, gives us $N_e(r)$ and the correct

polarization. These results hold only for the eclipse at which the data were obtained and in the equatorial plane.

Photoelectric photometry of the entire corona, or a ring-shaped part of it, fixes the absolute values of brightness with respect to the solar disk. The electron density can then be found, using the known Thompson scattering coefficient. There appears to be a change in the total intensity by a factor of 1.8 between sunspot maximum and minimum.

Simple inspection of the form of the corona at different phases of the solar cycle will convince anyone of the enormous changes that take place.

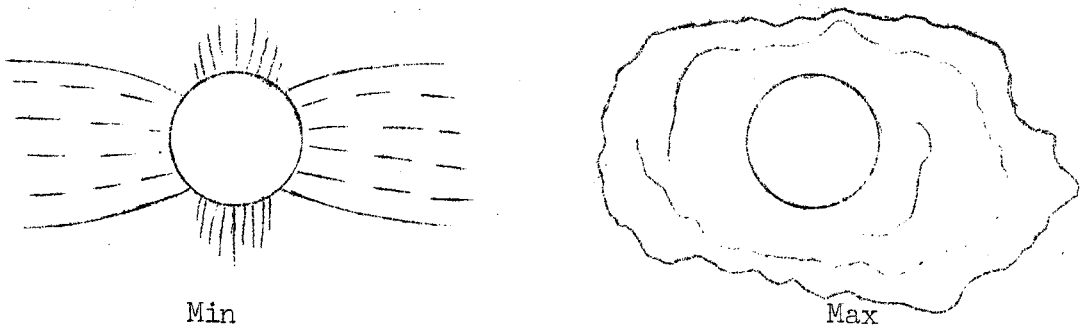


Fig. 2

Ludendorff (5), (6), and later Bergstrand attempted to put the changes on a semi-quantitative basis, by examining the ellipticity of the isophotes. An instance of the isophotes at maximum and minimum is given in Fig. 3. For this isophote the model used by van de Hulst (4) gives:

equator:	r = 1.5	K = 6.2	F = 2.7	F + K = 8.9
pole:	1.3	4.2	5.0	9.2

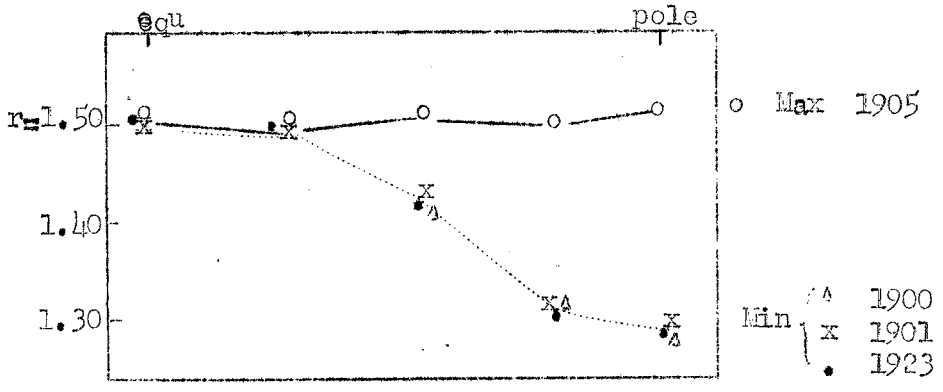


Fig. 3

The actual isophote has an even greater ellipticity. This is very apparent for r greater than 2, where the contrast between pole and equator seems to drop again.

After elimination of F the value of N_e may be derived from the variation of brightness along lines perpendicular to the solar axis. Analysis of data from the 1900 eclipse by van de Hulst indicates that at minimum phase the density does not monotonically decrease from pole to equator. It has a distinct minimum near latitude 70° and rises again to half the equatorial density at the pole. The polarization measurements are used as a general check on the analysis of the electron densities.

II. Determination of the Electron Temperature in the Corona.

The two great riddles of the solar corona have always been its great geometrical extension and the unidentified coronal lines. Both of these problems find their explanation, at least qualitatively, in a coronal temperature of a million degrees. There are various precise methods of obtaining the coronal temperature, several of which will be given below.

II. 1. Coronal temperature from scale height.

We may obtain an upper limit for T if we use the scale-height (vid. section 5.1). By plotting N_e against $1/r$, we find, using two assumptions for μ :

	$\mu = 0.69$	$\mu = 0.61$
equator	$T = 1.62 \times 10^6$	1.43×10^6
pole	1.15×10^6	1.02×10^6

It is remarkable that Van de Hulst (4) finds a density gradient inside the polar plumes that is nearly the same as the density gradient for the polar

region in general. This points to some equilibrium and shows that the polar rays are not "streamers" of rapidly outflowing matter.

II. 2. Coronal temperature from broadening of spectral lines.

The high coronal temperature gives strong Doppler broadening of spectral lines, both emission and absorption. Since the K component is produced by fast-moving electrons, the broadening of its lines is very great. The root mean square velocity of electrons at temperature T is:

$$v = T^{\frac{1}{2}}\, 10^{5.83} \text{ cm/sec}$$

so that we obtain

T = 10,000	v = 675 km/sec	v/c = 0.00225	$\Delta\lambda$ at 4000 A = 9 A
1,000,000	6750 km/sec	0.0225	90 A

Both of the above values for $\Delta\lambda$ are large enough to wash out most lines. The exact theory and profiles have been worked out by C. J. van Houten (7). The strongest lines, theoretically, should only have a central depression of 7% of the continuum.

Observations by W. Grotrian (8) show that the H and K lines together become a shallow dip visible from 4100 A, and corresponding to $\Delta\lambda$ about 100A, or a temperature of about a million degrees. D. H. Menzel (unpublished) did not find the dip at all.

Early eclipse work had already shown that the coronal emission lines were wide. The first exact work was photometry carried out by M. Waldmeier 1941 (9) of the line 5303 at a distance of 51" (r = 1.05) from the limb, using a slit spectrograph. The instrumental width was determined from the breadth of the Fraunhofer lines in scattered light. The real half-width obtained was 0.8 A. (with a dispersion of 9 A/mm). After corrections, the total width of

the green line at half central intensity is 1.10 A, which means that the intensity of the line is equal to exp - $(\Delta\lambda/0.65)^2$. With the theoretical formula and $\mu \cong 56$ (iron), this gives the temperature 3.96 x 10^6 degrees. This temperature is an upper limit because there may be other effects broadening the line.

Waldmeier also points out (10) that the lines are less wide at r = 1.07 than at 1.03, and that at r = 1.40, the half width goes down to 0.5 A. (11). 1946 Early work by Lyot (12) gave a half width of 0.80 A., which corresponds to a temperature of 2.1 x 10^6 degrees. In his George Darwin Lecture (13), 1939 Lyot gives coronagraph results from 200 spectra. The average widths of 11 lines are 0.8 A in the green, 0.9 to 1.0 A in the red and greater in the infrared. New work by Lyot and Dollfuss (unpublished) gives widths to an accuracy of 5%, which will give temperatures to within 10%. The results are not yet available. The green line is shown to have a satellite.

By the same methods line shifts giving radial velocities may be measured. The rotation of the corona can be determined this way (Waldmeier, (10)). Other radial velocities are usually not visible, but Waldmeier (10 a) notes that occasional velocities of 5 km/sec, and even (very seldom) up to 10-100 km/sec are occasionally observed as asymmetry.

II. 3. Coronal temperature determined from degree of ionization.

A third method of determining the coronal temperature is from the degree of ionization in the corona. We are not certain that the ionization is purely thermal, but we can at least estimate the influence of T. As is well known, the coronal emission lines were first identified by Edlen (14) and his work is summarized by Swings (15), (16). The list below gives these identifi-

cations, together with the observational data of Lyot (13), published before Edlen's work. Note: we write Fe 13 instead of the somewhat clumsy official notation Fe XIII, i.e. iron missing 12 electrons.

Line	Intensity	Identification element	ion. pot.	Behavior
3388	44	Fe 13	325 volts	These lines show
5116	3	Ni 13	350	fine arches, but
5303 (grn ln)	120	Fe 14	355	somewhat fuzzy
6702	5	Ni 15	422	and filled up.
7057	1			Not enhanced in
8024	1	Ni 15	422	flares.
10747	240	Fe 13	325	
10798	150	Fe 13	325	
6374 (red ln)	20	Fe 10	233	(These two first identified by Grotrian)
7892	30	Fe 11	261	Fairly sharp detail, not filled. Enhanced in flares.
5694 (yellow ln)	1	Ca 15	814	Few details; seen only exceptionally and only if continuous spectrum is very strong. Around active spot regions. (vid. Waldmeier (17))
5536	1-0	A 10		Observed at eclipses. Never since.
3328 and 11 other lines in blue & UV	3	Ca 12 Fe 11-13 A 14 Ni 13-16		

Table 1

The identifications for 5694 and 5536 have recently been questioned by Naqvi (unpublished). The best general reproductions of the coronal spectrum are given by Lyot (13, plate 15). Waldmeier (18) gives a fine reproduction of the line 5303. He has also noted reduced intensity of the red line near stationary prominences (filaments).

Edlen's identifications indicate, obviously, a high coronal temperature. The detailed ionization theory of the corona has been worked out in recent years by several authors (Woolley and Allen (19), Biermann (20), S. Miyamoto (21), Schklovsky 1948 (22)). The ordinary Saha ionization equation cannot be applied here, because thermodynamic equilibrium does not prevail. We may, however, consider the detailed processes active in the corona.

Suppose we consider the equilibrium between a twelve times ionized ion and a thirteen times ionized ion. The densities will be denoted by N_{13} and N_{14}, respectively. Then the processes governing the equilibrium will be the following:

	ionization	recombination
by radiation	$N_{13}\, Q$	$N_{14}\, N_e\, \alpha$
by collision	$N_{13}\, N_e\, S_1$	$N_{14}\, N_e^2\, S_2$

The expressions in the table give the probability for the occurrence of each process; Q depends on the radiation field, while the recombination coefficient α and the collision coefficients S_1 and S_2 depend on the temperature.

The authors cited agree that ionization by radiation and recombination by collision are unimportant in the equilibrium state, the former because no evidence for the presence of the soft X-rays necessary to produce this ionization has been observed, and the latter because of the unlikelihood of three

body encounters. The remaining two processes can then be balanced, to give

$$\frac{N_{14}}{N_{13}} = \frac{S_1}{\alpha}$$

The degree of ionization is independent of the electron pressure! S_{12} and α are estimated by Woolley and Allen on theoretical grounds. They are:

$S_1 = 2 \times 10^{-5} \frac{1}{\chi^2} \exp(-\chi/86)$, with χ = ion. potential in volts;

$\alpha = 1.5 \times 10^{-12}$ for Fe 14 at 10^6 degrees.

Curves giving the ratios for different ions and various assumed temperatures are given in Fig. 4, below.

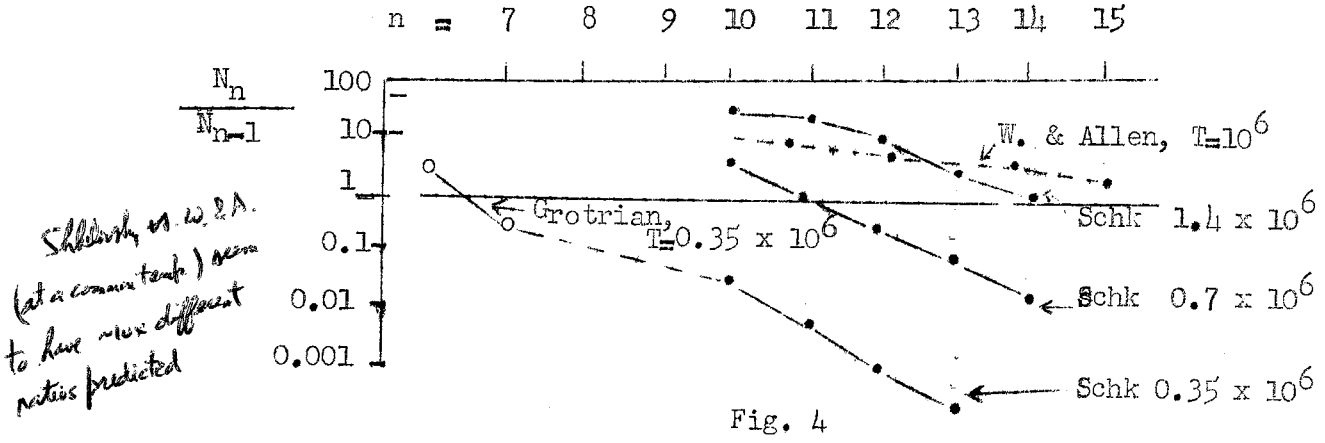


Fig. 4

The points on the ordinate n = 10, for example, give the ratio N(Fe 10)/N(Fe 9). The absolute values of the constants here are not too reliable, but the change in relative abundance of ions with temperature is extremely rapid, for we multiply by several factors. For example:

$$\frac{N(\text{Fe 14}) \text{ (green line)}}{N(\text{Fe 10}) \text{ (red line)}} = \begin{cases} T = 1.4 \times 10^6 & \{ 40 \times 25 \times 5 \times 2 = 10{,}000 \\ T = 0.7 \times 10^6 & \{ 2.5 \times 0.8 \times 0.25 \times 0.08 = 0.04 \end{cases}$$

The change produced by doubling the temperature is a factor of 2×10^5. In fact, a change of only 13% in the temperature will change the relative abundances

of these two ions by a factor of 10. Any theory of radio emission requiring large local changes in the temperature of the corona must be wrong.

The question of the mechanism of excitation is not completely settled. Excitation by electron collision seems most likely.

Waldmeier (23) discusses the invisibility of Hα in the coronal emission spectrum. The obvious reason for its absence, viz. that the temperature is so high that there are few recombinations, is confirmed by his computations.

III. Structure of the Corona. Motion of Coronal Gases.

III. 1. Zonal structure.

The solar rotation, alone and in interaction with the sun's magnetic field, gives rise to primary and secondary effects having a zonal structure (see Unsöld, (24)). The rotation has been measured from prominences by D'Azambuja (25) and from the corona by Waldmeier, both for the equator (10) and the pole (26).

The general solar magnetic field is certainly much smaller than was thought by Hale. Yet it seems to be present and clearly marked by the polar plumes. Its magnitude and change with the solar cycle are still entirely uncertain. The drifting of the sunspot zone and the second prominence zone has been mentioned in section 5.1; the changing forms of the white corona were discussed in the beginning of this section.

III. 2. Coronal emission regions.

When Waldmeier constructed his coronagraph in 1939, he started to estimate the brightness of 5303 around the sun's limb (27). This material now covers a full solar cycle. The first real isophotes were obtained (18) in 1940

(by estimating the limit of visibility) and an ellipticity estimate was made in the same way as Ludendorff had (5). Since 1942, Waldmeier has presented this material in a more systematic manner. He defines as "Koronastrahlen" each maximum along the sun's limb. The material is collected in: (a) A catalog (28); (b) Heliographic maps (29), (30), (31, fig. 2-4); (c) Polar maps (31), (32), for latitudes greater than 50 degrees. The statistical discussion of the dependence on the phase of the cycle starts with a butterfly diagram (33), based on the catalog (positions only). In a later paper (34), (35) a more refined treatment is given, in which the intensities are used and averaged statistically. Lyot's statistics only date back to 1944. In an early paper Waldmeier (36) shows that, on the average, the form of the corona, in the sense of dependence on latitude, is the same in 5303 and 6374.

III. 3. Fine structure in the corona.

Quite a bit of effort has been devoted to observations of the fine details of the corona. The best known of these are the spectacular prominences. Because of their spectrum, the prominences are generally considered as part of the chromosphere. Yet, extending up to 100" above the chromosphere (or up to r = 1.10), geometrically they belong to the corona.

Another small-scale phenomenon in the corona is the presence of persistent regions of high coronal emission. These are variously termed "jets" or "rays," or filaments." Waldmeier uses the term "filaments" for regions of high emission (in 5303 or 6734) that persist for several days. He maps them and puts them together in a "heliographic map" of coronal emission. Some of these, for the region of the great sunspot, March, 1949, have been reproduced (37). The general impression conveyed is that there is a great deal of washing out

in longitude, due to the fact that one cannot distinguish whether the light is emitted in the plane of projection or before or behind it.

During the period 1939-1945, Waldmeier (31) observed the filaments to remain concentrated around latitude 60 degrees. Then they began moving higher, reaching the poles just in time for the sunspot maximum (1947). They show the same rotational velocity as prominences (vid. III. 1) and often last for several months. They are invariably elongated in longitude.

Lyot has also remarked the persistence of these emission regions. He calls the "koronastrahlen" "jets coronaux," and notes that the projection effect is important. This is shown clearly by the fact that a jet that is visible for a week reaches minimum latitude in the middle, producing an apparent path that is an arc bent towards the equator. A polar jet has been visible for several months, and oscillating neatly about the polar axis.

Waldmeier (37) terms all features with lifetimes of the order of several hours or half a day coronal "condensations" or "arches" (the latter, of course, only when they are arch-shaped). These are "Detailed structures in the vicinity of spot groups, in which the intensity changes rapidly, and can temporarily increase by a factor of 5."

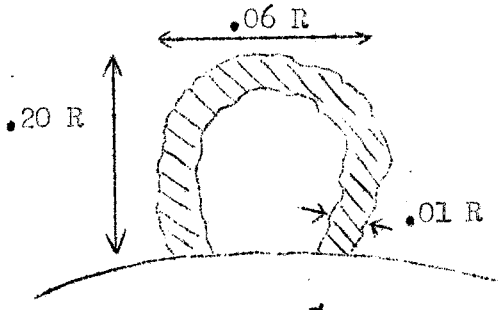


Fig. 5

The best known of the features are the arches, diagrammed in Fig. 5 above. The best available photographs of these are given by Lyot (38, plates

III and IV). According to Lyot the details remain stationary during six hours (total duration of visibility) while the relative intensities in the structure remain constant only about one hour. The thickness of the arch is about 10", or .01 R (in the red line). This fuzziness is certainly not due to the instrumental broadening, for the prominences are perfectly sharp. Lyot remarks that the green line gives less detail than the red. This has not been noted by Waldmeier.

Plate IV of Lyot's paper (38) shows clearly that there is "a nearly complete independence between corona and prominences." This certainly is true of the relative intensities (sometimes a strong coronal arch with hardly any prominence visible in the vicinity, sometimes vice versa). Yet even in the few pictures reproduced we find a coincidence in position, if not intensity. This is confirmed in one exceptional case studied by Waldmeier (39). A rough sketch is given in Fig. 6 . The yellow line 5694 shows just one region of con-

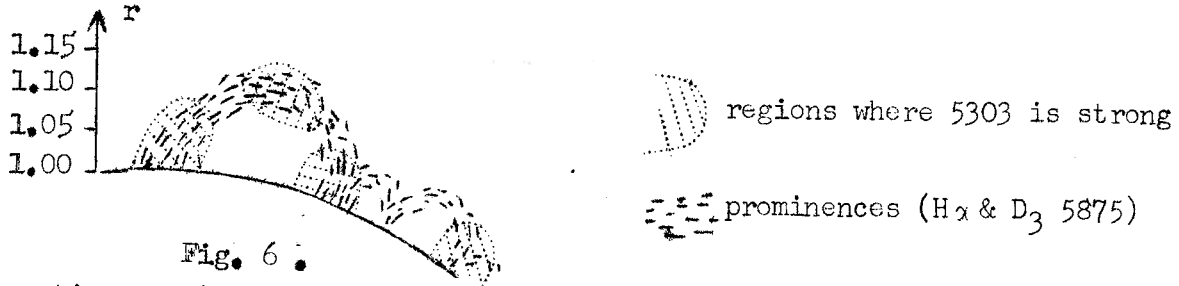


Fig. 6 .

densation, again coinciding with the prominence heads. There are, again, changes in relative intensity of duration about 1 hour.

The coronal condensations also turn up in slitless grating spectra of the eclipsed sun, e.g., the eclipse of 1930 (S. A. Mitchell (40) and W. Grotrian (41)). There are also reports on earlier visual observations. Eclipse observations might be used to ascertain the lower boundary of coronal emission, which is impossible with the coronagraph.

Lyot has remarked (oral communication) that it is quite general for a prominence to be located in a dark "shadow region" of the corona. This region appears relatively dark, both in the light of the emission lines and in the continuous coronal light (Fig. 7 (a)).

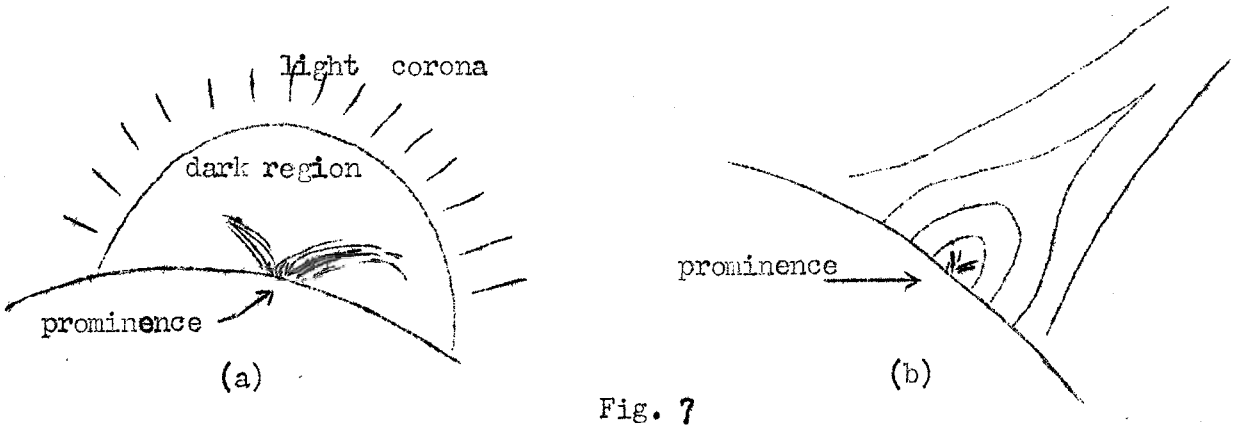


Fig. 7

"Domes" around prominences are well known in eclipse photographs. They often have an alternate light-dark structure (Fig. 7(b)) and gradually change into the typical form of a coronal "streamer." The inner dark area of such a dome is probably identical with the shadow region noted by Lyot. Examples may be found in the corona at the following eclipses: 1901 (SE quadrant, 3-fold), 1905 (SW, a double one; NE two large ones), 1919 (SW double; E, single) and others.

The size of the domes is of the same order as that of the arches seen in the light of the emission lines. They may be just very striking examples of these.

see Fig. 2 on p. 78

The polar rays, or plumes, are short and brush-like, visible in the polar regions of the corona at eclipses near minimum phase. Large scale photographs are necessary to show good detail. Van de Hulst (4) has analyzed them photometrically and shows that the brightness exceeds the background by 20%. But

if the F corona is eliminated, the resulting N_e is five times larger in the rays than between them. The data rejects the idea of ejected streams, because the density falls off too sharply. The suggestion is made that the rays are formed by matter lining up along the magnetic lines of force.

III. 4. Further observational questions.

There are still many unanswered questions on the corona. Motions in the corona are still poorly known. Some motions have been discovered in the form of radial velocities, but they are seldom over 5 km/sec. Yet some motions of ten or one hundred km/sec noted by Waldmeier (see p.81) seem to be real. The case is not so good for proper motions, although numerous reports of these have been made. The best discussion of this question is given by Lyot (38, p. 42), who discusses the various reports of eclipse observations of velocities of the order of 5-10 km/sec. Lyot observed the corona with photographs and tri-color filters. He reports relative movements of the arches of less than one km/sec, noting that these could be easily accounted for by perspective and solar rotation. Waldmeier's answer to this (42) shows that some motions are real, but the main evidence rests on radial velocities.

How large are the density and intensity variations in the corona (locally and with the cycle)? There are very little data available on actual intensity variations. Waldmeier estimates that the condensation represents (in certain cases) a density increase of a factor 20. Very rough estimates of the change with phase made from Lyot's daily records give for the main spot zone in arbitrary units:

	emission lines	continuous light
Min (1944)	50	30 (25-35)
1945	100	50
Max (1947)	100-150	55 (50-60)

The intensity of the continuous light is measured from the polarization. The data for emission lines in 1947 omits all emission peaks; were these included, the intensity would be about 400. These estimates, that are not final, seem to confirm the change in electron density by a factor 1.8. The factor for the emission lines may be 3 or 4, without peaks, and 10 if they are included. Waldmeier (43) finds a factor of five between green line intensity at minimum and maximum. The intensity change with phase and with latitude is given in (44).

A few more theoretical questions that may be asked about the corona are: Is the electron-proton gas a good plasma, in the sense that we can ascribe to it a fixed Maxwellian velocity distribution? and, how long does it take to establish ionization equilibrium in the corona?

Bibliography on the Corona

1. A. Schuster, M.N. 40, 35, 1879.
2. M. Minnaert, Zs. f. Ap. 1, 209, 1930.
3. S. Baumbach, A. N. 263, 121, 1937.
4. H. C. van de Hulst, B.A.N. #410, 1950.
5. Ludendorff, H., Sitzber. d. Preuss. Ak. d. Wiss. 10, 185, 1928;
6. " " " 16, 200, 1934.
7. C. J. van Houten, B.A.N. #410, 1950.
8. W. Grotrian, Zs. f. Ap. 3, 220, 1931.
9. M. Waldmeier, Zs. f. Ap. 20, 323, 1941.
10. M. Waldmeier, Astr. Mitt. d. Eidg. St. Zürich, # 147, 1946.
10a. M. Waldmeier, Astr. Mitt. d. Eidg. St. Zürich, # 151, 1947.
11. M. Waldmeier, Physica 12, 733, 1946.
12. B. Lyot, Comptes Rendus 202, 1259, 1936.
13. B. Lyot, M.N. 99, 580, 1939.
14. B. Edlen, Zs. f. Ap. 22, 30, 1942.
15. P. Swings, PASP 117, 1945.
16. P. Swings, Ap J 98, 116, 1943.
17. M. Waldmeier, Astr. Mitt. 146, 1946.
18. M. Waldmeier, Zs. f. Ap., 22, 30, 1942.
19. R.v.d.R. Woolley and C. W. Allen, M. N. 108, 292, 1948.
20. L. Biermann, Naturwiss. 34, 87, 1948.
21. S. Miyamoto, Pub. Astr. Soc. Japan 1, 10, 1949.
22. I. S. Schklovsky, Russ. Astr. J. 25, 1948.
23. M. Waldmeier, Experientia I, 1, 1945.

24. A. Unsöld, Physik der Sternatmosphären, Berlin, 1938.

25. L. D'Azambuja, Ann. Meudon 6, fasc. VII, 1948.

26. M. Waldmeier, Astr. Mitt. # 165, 1950.

27. M. Waldmeier, Zs. f. Ap. 19, 21, 1939.

28. M. Waldmeier, Astr. Mitt. # 170, 1950.

29. M. Waldmeier, Zs. f. Ap. 21, 109, 1942.

30. M. Waldmeier, Naturwiss. 32, 51, 1944.

31. M. Waldmeier, Astr. Mitt. # 165, 1950.

32. M. Waldmeier, Zs. f. Ap. 27, 24, 1950 (same as # 31.).

33. M. Waldmeier, Astr. Mitt. 157, 1949.

34. M. Waldmeier, Ibid. # 171, 1950.

35. M. Waldmeier, Zs. f. Ap. 27, 237, 1950 (same as #34.).

36. M. Waldmeier, Ibid. 20, 172, 1940.

37. M. Waldmeier, Astr. Mitt. # 169, 1950.

38. B. Lyot, Ann. d'Ap 7, 31, 1944.

39. M. Waldmeier, Astr. Mitt. # 146, p. 14, 1945.

40. S. A. Mitchell, Smiths. Inst. Publ. # 3453; Smiths. Rept. 1937, pp. 145-167; Ap. J. 75, 1, 1932.

41. W. Grotrian, Zs. f. Ap. 8, 124, 1934.

42. M. Waldmeier, Astr. Mitt. # 151, 1947.

43. M. Waldmeier, Astr. Mitt. # 164, 1949: Zs. f. Ap. 27, 236, 1950.

44. M. Waldmeier, Zs. f. Ap. 27, 236, 1950.

5/4/87 -6r ch M & HA I
BTS #41
1923

Chapter 6 RADIOEMISSION FROM THE QUIET SUN (A. A. Wyller)

6.1 Kramers' theory for the emission and absorption by free-free transitions.

We shall derive the following formula, governing the absorption of radio waves in a gas of electrons and ions in thermal equilibrium:

$$\kappa(\nu) = \frac{1}{n} N_i N_e \frac{4\sqrt{2}}{3\sqrt{\pi}} \frac{1}{(mkT)^{3/2}} \frac{Z^2 e^6}{c\nu^2} \ln(p_a/p_b)$$

Here n = refractive index (see later) $\quad -e$ = electron charge

ν = frequency $\quad Ze$ = ionic charge

N_e = number of electrons per cm^3 $\quad m$ = electron mass

N_i = number of ions per cm^3 $\quad k$ = Boltzmann constant

c = velocity of light $\quad p_a$ and p_b are limiting impact

T = temperature $\quad$ parameters (see later).

The absorption coefficient $\kappa(\nu)$ is a macroscopic one, not an atomic one. It occurs in the equation of absorption $\frac{dI}{dx} = \kappa_\nu I_\nu$. The dimension of the coefficient is cm^{-1}.

Fourier analysis. We will here introduce some theorems from Fourier analysis that will prove useful in the derivation. Let $f(t)$ be real. We define the "Fourier transforms"

$$C(\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty} f(t)\cos\omega t\,dt \quad , \quad S(\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty} f(t)\sin\omega t\,dt$$

Now Fourier's theorem states that

$$f(t) = \int_0^{\infty}\left[C(\omega)\cos\omega t + S(\omega)\sin\omega t\right] d\omega$$

Parseval's theorem is also important:

$$P = \int_{-\infty}^{\infty} f^2(t)\,dt = \pi\int_0^{\infty} C^2(\omega) + S^2(\omega)\;d\omega$$

The physical implications of these theorems are important: $f(t)$ will represent the amplitude (more precisely a component of the acceleration) in a wave train

and ω the circular frequency $2\pi\nu$. The Fourier transforms express how strongly various frequencies are present in the wave train. P represents the total energy in the train. Parseval's theorem shows that we can think P to be composed either of the contributions of each small element of time (but these not specified with respect to frequency) or of the contributions of each frequency band width (but these not specified with respect to time). This situation is typical for Fourier transforms and is similar to that behind Heisenberg's uncertainty principle.

Emission by one encounter.--We can use classical electron theory. Radiation is emitted by a charged particle only when it is accelerated; this occurs during an encounter. We start from the classical formula that the radiation per unit time by an accelerated electron is (all frequencies and all directions together):

$$Q(t) = \frac{2}{3}\frac{e^2}{c^3}\ddot{x}^2.$$

$\ddot{x}$ denotes the acceleration; the factor 2/3 arises from integration over the typical dipole radiation pattern, which has zero intensity in the direction of $\ddot{x}$. The energy emitted during the entire encounter is $Q = \int_{-\infty}^{\infty} Q(t)dt$. Our idea is to divide this energy up with respect to frequency rather than with respect to time. Thus $Q = \int_0^{\infty} Q(\nu)\, d\nu$. In order to do so we replace the f(t) of the Fourier theorems by $\ddot{x}$. Then apparently

$$Q = \frac{2e^2}{3c^3}P = \frac{2e^2}{3c^3}\,\pi \int_0^{\infty} \left[C^2(\omega) + S^2(\omega)\right] 2\,\pi\, d\nu,$$

so that

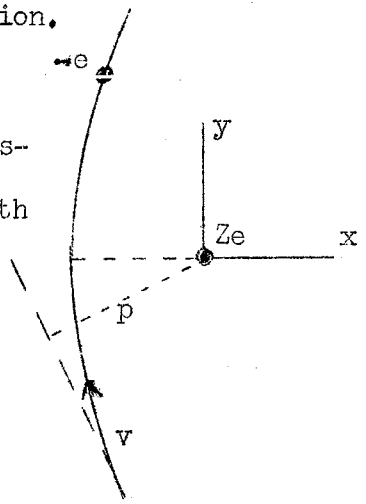
$$Q(\nu) = \frac{4\pi^2 e^2}{3c^3}\left[C^2(\omega) + S^2(\omega)\right].$$

A similar contribution from the y and z component of the acceleration should be added. But by proper choice of coordinate system (see figure) we make $\ddot{x}$

even, $\ddot{y}$ odd and $\ddot{z}$ zero. The formula for $Q(\nu)$ is then complete by substituting

$$C(\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty} \ddot{x}(t)\cos\omega t\,dt, \qquad S(\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty} \ddot{y}(t)\sin\omega t\,dt.$$

Specification of the encounter.--Three kinds of encounters occur in the gas: ion-ion, electron-electron, and ion-electron encounters. In the first two kinds the particles have equal charge and equal but opposite accelerations. The dipole radiation of one particle will then exactly cancel that of the other and only a small quadrupole radiation remains; this we neglect. In the ion-electron encounters the dipole radiation from the ion reinforces that from the electron; the latter is by far the bigger because of the small mass. So we deal only with the emission by an electron during its encounter with an ion.



The orbit is a hyperbola. It is specified by the initial velocity v and impact parameter p which is the distance from the (fixed) ion to the undisturbed straight path of the electron. We define

$$\alpha = \frac{\text{potential energy at distance p}}{2 \times \text{initial kinetic energy}} = \frac{Ze^2}{pmv^2};$$

There are two limiting cases, both worked out by Kramers in his original paper. If α is small compared to 1 the deflection will be small and the orbit is nearly a straight line. If α is large compared to 1 the electron falls toward the ion with not much initial energy and the orbit is nearly a parabola. By good luck we have to deal in radio astronomy with the nearly straight encounters only. Put t = 0 at closest approach; with good approximation: $r^2 = p^2 + v^2t^2$. The acceleration Ze^2/mr^2 thus has the components:

$$\ddot{x} = \frac{Ze^2}{mr^2}\cdot\frac{p}{r}, \qquad \ddot{y} = \frac{Ze^2}{mr^2}\cdot\frac{vt}{r}$$

We can go on making approximations. The duration of the encounter is actually infinitely long but the effective duration will be of the order of p/v. In radioastronomy the frequencies are small. We may thus assume that $p\omega/v \ll 1$, which implies that $\omega t \ll 1$ during the entire encounter. So, approximately, $\cos \omega t = 1$ and $\sin \omega t = \omega t$. This shows at once that the S-integral is an order of magnitude smaller than the C-integral. The latter can be found by direct integration:

$$C(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} \mathfrak{k} dt = \frac{Ze^2 p}{m\pi} \int_{-\infty}^{\infty} \frac{dt}{(p^2 + v^2 t^2)^{3/2}} = \frac{2Ze^2}{\pi pmv} .$$

The energy emitted per collision per c/s (in all directions together) will thus be:

$$Q(\nu) = \frac{16\, e^6\, Z^2}{3c^3 p^2 m^2 v^2} .$$

In this approximation the spectral intensity is independent of frequency. It arises only from the rapid change in momentum given to the electron during the encounter.

Integration over many encounters.--First we assume one electron velocity v. The number of encounters, or as they are usually called collisions, per cm^3 with impact parameters in a small range dp is $N_i N_e v . 2\pi p dp$. Multiplying this by $Q(\nu)$ and then integrating over p we find the energy $E(\nu)$ emitted per c/s per second by one cm^3 of the electron-ion gas in all directions. It is

$$E(\nu) = N_i N_e \frac{32\pi e^6 Z^2}{3c^3 m^2 v^2} \int_0^{\infty} \frac{dp}{p} .$$

The integral diverges at both ends; this is due to the approximations we have made. As a formal remedy we replace the integration limits by the values p_b and p_a; these signify the lowest and highest impact parameter for which our

approximations were admissable. So the integral is replaced by a logarithmic factor:

$$\ln (p_a/p_b).$$

We still must integrate over the Maxwell velocity distribution. The average value of $1/v$ is $(2m/\pi kT)^{\frac{1}{2}}$. We neglect the influence of the Maxwell distribution on the values of p_a and p_b because rough estimates only are needed under the logarithm. The result is

$$E(\nu) = nN_iN_e \frac{32}{3}\left(\frac{2\pi m}{kT}\right)^{\frac{1}{2}} \frac{Z^2e^6}{m^2c^3} \ln(p_a/p_b)$$

In this and the following formulae we already include the factors due to a refractive index n that may differ from 1 (see below).

This result for emission might be used for computing the brightness of a transparent cloud of gas at radio frequencies. We find it more useful to work with absorption coefficients. So we apply Kirchhoff's law. The black body radiation intensity in a medium with refractive index n is $B(\nu) = \frac{2\nu^2kT}{c^2}\cdot n^2$. By Kirchhoff's law the brightness of 1 cm^3 of the gas in a given direction (i.e. the energy emitted by it per second per unit solid angle) may be written in two equivalent ways as

$$\frac{E(\nu)}{4\pi} = \varkappa(\nu) \cdot B(\nu).$$

Solving for $\varkappa(\nu)$ we find

$$\varkappa(\nu) = \frac{1}{n} N_iN_e \frac{4\sqrt{2}}{3\sqrt{\pi}} \frac{1}{(mkT)^{3/2}} \frac{Z^2e^6}{c\,\nu^2} \ln (p_a/p_b),$$

which is the formula quoted at the beginning. Our further task will be to specify what is meant by p_a, p_b, and n.

Lower and upper impact parameter.--The values of p for which our approximation breaks down should automatically appear when once more we review the approximations made during the derivation.

(1) v is small compared to the velocity of light. This approximation was made in applying Newton mechanics and in neglecting retardation. It is well fulfilled.

(2) The orbit is nearly a straight line. This requires $\alpha \ll 1$, which poses a lower limit to p, which we denote by $p_{b_1} = \frac{Ze^2}{mv^2}$. For smaller p the encounter becomes more nearly parabolic and the total impulse suffered by the electron is much smaller. Incidentally, when approximation (5) holds a direct integration of the Fourier coefficients can be made without making assumption (2); this is of no use to us.

(3) the orbit is not influenced by the loss of energy by radiation. Follows from (1) and (2). The electron remains free. It is a free-free transition.

(4) The electrons were thought as localized points. If p becomes as low as the De Broglie wave length of the electrons this holds not true. We put the effective cut-off at half this wave length, i.e. $p_{b_2} = \frac{h}{4\pi mv}$.

(5) We have assumed that ω was smaller than the inverse duration of the encounter: $\frac{\omega p}{v} \ll 1$. This gives an upper limit to p: $p_{a_1} = \frac{v}{\omega}$. For any p larger than this value the encounter is so slow that frequencies as high as ω are not emitted by it.

(6) We have assumed that the collision is complete and undisturbed, i.e. that the electron moves in the field of only one ion. This leads us to assume that the interionic distance is an effective upper limit to p. So $p_{a_2} = N_i^{-1/3}$. This argument is not as stringent as it would seem. It is conceivable that electrons still feel the influence of ions even behind several other ions and electrons. Cohen, Spitzer and Routley, in discussing the related problem of the

conductivity of an ionized gas, advance arguments why not the interionic distance but the Debye wave length of the plasma is the effective shielding length. Then $p_{a_3} = (\frac{kT}{4\pi N_e e^2})^{\frac{1}{2}}$. In all published computations and also in the following sections this possibility has been neglected; it does not seem quite certain.

Summarizing, we have 2 lower limits and 2 upper limits to p. Evidently, one reason for the approximation to break down is sufficient. So we have to take the narrowest range, i.e. the higher value of the p_b's and the lower value of the p_a's. All four combinations have been mentioned in the literature. Advisable in cases of doubt is a numerical estimate of all four values. Typical values for corona and chromosphere are mentioned in section 6.3. Fortunately the ratio p_a/p_b is at least 1000 so that an error of a factor 2 in these rough estimates affects the logarithm only slightly.

The refractive index. The classical expression for the refractive index of an ionized gas from the theory of H. A. Lorentz is

$$n^2 = 1 - \frac{N_e e^2}{\pi \nu^2 m} = 1 - \frac{\omega_0^2}{\omega^2}, \qquad \text{where } \omega_0^2 = \frac{4\pi N_e e^2}{m}$$

The expression had been derived on the assumption that electromagnetic waves travel through the gas and that the electrons are free to swing back and forth on the electric fields of these waves. This value of n is the one we have to use in the formulae of emission, absorption and black body radiation just given. That the emission contains n, the black-body intensity n^2 and the absorption coefficient n^{-1} may be derived in several ways. Most simple is to assume formally a different value of the dielectric constant of the vacuum or medium through which the electrons and ions move and then go through the derivation from scratch.

Lorentz indicated a different way. The electrons are not free to move; at odd times their oscillation is disturbed by collisions (i.e. encounters with

ions). Let γ be the collision frequency = number of collisions per second. Lorentz then finds the modified formula for n:

$$n = 1 - \frac{\omega_o^2}{\omega(\omega - i\gamma)}$$

The refractive index now has become complex; this denotes in general an absorption of the electromagnetic waves along with the refraction. This absorption we shall now derive. We decompose n in the real and imaginary part: $n = n_o - in_1$. The amplitude in a wave traveling in this medium in the z-direction is proportional to

$$e^{-i\omega nZ/c} = e^{-i\omega n_o Z/c} \cdot e^{-\omega n_1 Z/c} .$$

The intensity thus contains a factor $e^{-2\omega n_1 Z/c}$, which means that the linear absorption coefficient is $\kappa(\nu) = \frac{2\omega n_1}{c}$. The rest is algebra; we assume that γ is small compared to ω and find by straightforward expansion (left to the reader) that $n_1 = \frac{\gamma\omega_o^2}{2n\omega^3}$. So that finally

$$\kappa(\nu) = \frac{\gamma\omega_o^2}{nc\omega^2}$$

This formula confirms by a quite different way that the absorption coefficient is proportional with n^{-1} and with ν^{-2}. By comparing the formula just found with the earlier derived expression of $\kappa(\nu)$, we can find an expression for the collision frequency γ, which corresponds to the one derived from conductivity studies.

References

H. A. Kramers, Phil. Mag. 46, 836, 1923.

S. F. Smerd and K. C. Westfold, Phil. Mag. 41, 831, 1949.

K. C. Westfold, Phil. Mag. 41, 509, 1949.

R. S. Cohen, L. Spitzer, P. McR. Routley, Phys. Rev. 80, 230, 1950.

6.2 Observations of the quiet sun

The classification of the components of solar radio emission to which we adhere in the following chapters is shown in this table. All components other than the quiet sun are collectively called the active sun and treated in chapter 7.

Phenomenon	Approx. const. during:	Localization	Polariz.	T_a	T_e
Quiet Sun	1 year	whole disk and outside	absent	7000-10^6	same
Enhanced radiation at short waves, e.g. 10 cm.	1 week	spotgroups sunspots	15%	10^5	10^6
enhanced radiation at meter waves = noise-storms	12^h-3^d	spotgroups	100%	to 10^9	to 10^{12}
This contains Stormbursts	2 sec.	spotgroups	100%	to 10^9	to 10^{12}
Isolated bursts	2 sec.	?	0	to 10^8	to 10^{11}
Outbursts	1-10 min.	flares	0, except in tail	to 10^{13}	to 10^{16}

How to isolate the quiet sun from the observations. The radioemission measured when the instrument is directed towards the sun is made up of three parts. First the galactic noise from the region of the Milky Way around the sun. This has to be eliminated by measuring the intensity of that part of the sky when the sun has moved further along the ecliptic. But even now there will be variations in the solar radio emission itself. Careful investigations have shown

these variations to be associated with the occurrence of certain optically active regions on the sun like spotgroups and flares. These contributions come from the active "radiosun." When they have been removed there still remains a finite emission in the radiofrequencies, which we call the emission from the quiet sun. The difficulty in observing the quiet sun lies in allowing for the additional noise on the records of observation. The bursts can well be eliminated as they have a short lifetime and high intensities. They will show up as sharp peaks on the records. But the enhanced radiation has longer durations so its effects often are indistinguishable on the records from the quiet sun's emission. Two ways are open, in principle, to make the distinction.

1. Analyze the time sequence of events.--The record of any day may seem fairly constant, or somewhat fluctuating with a lot of stormbursts. But if the level is measured and well calibrated each day the variations soon show up. Depending on circumstances one may then proceed as follows.

(a) Observe many days and assume that the sun was perfectly quiet on the day with the lowest noise level. This is a crude way but it works in the cm range.

(b) Make a histogram, i.e. a graph of the frequency distribution of the recorded daily levels. This is useful for waves of 1 meter and larger, for Pawsey has shown that the histogram has a fairly sharp cut-off at some low level. The obvious interpretation is that this base level below which the sun's emission never sinks refers to the quiet sun.

(c) In the 3-50 cm range the histograms look more like a normal curve so are not very useful. Denisse and others have shown that for these wave lengths a fairly sharp correlation between daily level and sunspot area exists. By extrapolating on the correlation graph to zero sunspot area we find the quiet sun's level.

2. Use the localization of the active phenomena.--All emissions from the active sun are localized on or near the sun's disk in the vicinity of spots or in other active regions. By proving that a phenomenon is localized we prove that it does not belong to the quiet sun. The practical forms of using this principle are:

(d) Use of a partial or total eclipse. The record of the sun's total radiation as a function of time shows a smooth fall and rise during the hours that the eclipse lasts. In addition, sudden drops occur when an active region is covered by the moon and sudden rises when it is uncovered. In principle these sudden changes may be eliminated from the results and only the quiet sun's radiation retained.

(e) Also an interferometer (sea-interferometer or meridian interferometer) will reveal the presence of localized active regions. For instance, when the sun drifts through the pattern of a meridian interferometer with sharp fringes, the sun's disk (quiet sun) will give a smooth curve and a spot (active sun) will give fringes. So the first part can be isolated by simply drawing the lower envelope to the observed fringes.

(f) The ideal way, of course is to have a pencil beam that is narrow enough to scan the sun's disk and locate any spots. This can now be done with the biggest telescopes in the 1 and 3 cm range.

Observed values of the total radiation of the quiet sun. This radiation may be expressed either as the flux density S or as the apparent temperature T_a. We choose the latter presentation. The figure (page 107) shows observed values

as dots. Each dot is found from a series of observations by isolating the quiet sun by one of the methods (a) to (f). Details are shown in the table on the next page.

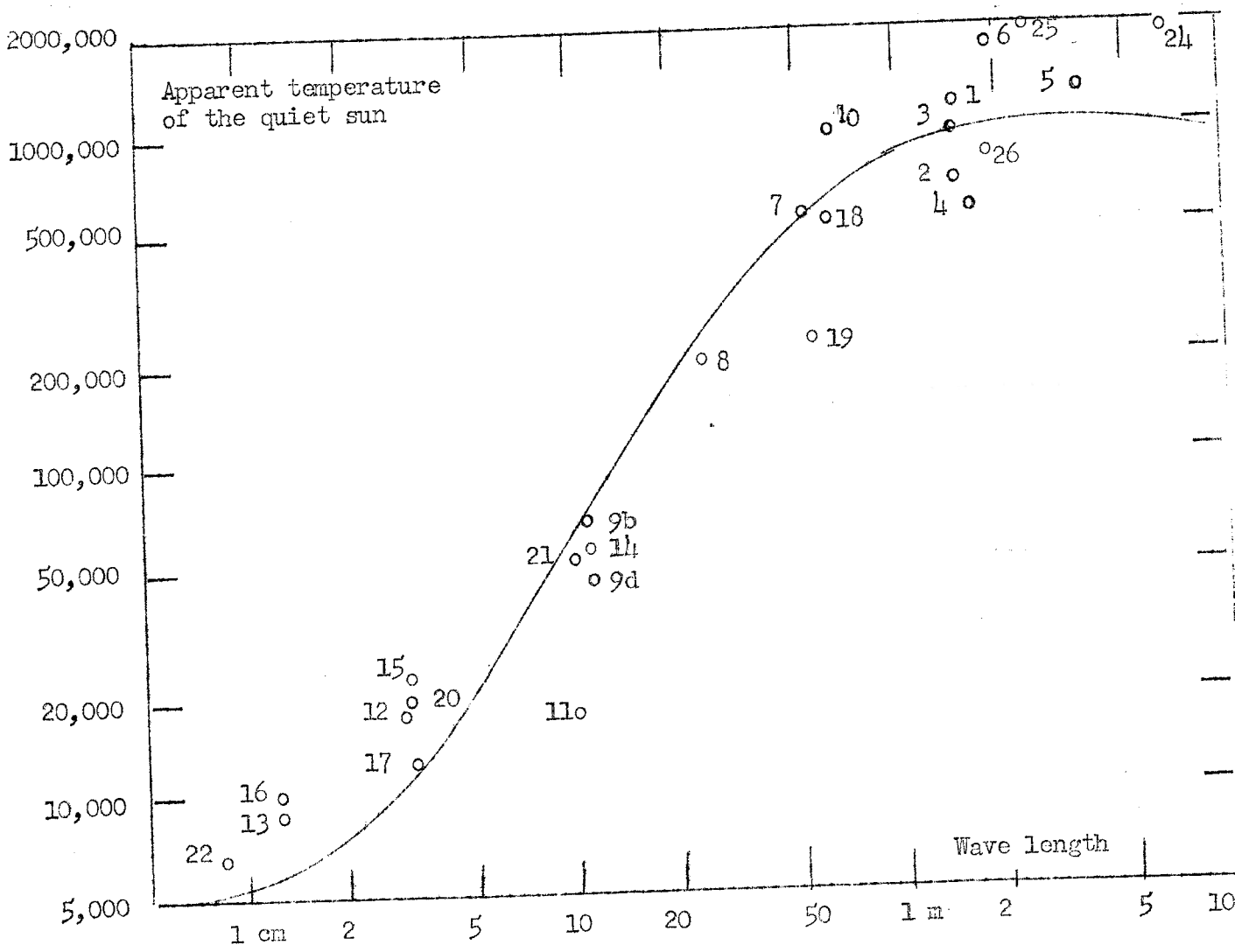
Observations 1 to 16 are copied from the paper by Pawsey and Yabsley (points 7 and 8 later corrected for better calibration). The other observations have been added from independent sources and reduced in a similar vein. The accuracy is 0.5 db or 15-20 per cent in the cm region and 1-2 db or 25-50 per cent in the meter region. Observations 11, 12, and 19 are isolated observations that may have larger errors than the others. For wave lengths larger than 5 meter the quiet sun is too weak with respect to galactic (and extragalactic?) noise to be measured at all with existing antennas. The whole temperature range corresponds to a factor 200 or 23 db so that errors of 1 db may be considered relatively small.

List of observed values T_a for the quiet sun (as of March, 1951):

		λ	year	period in months	$T_a \times 10^{-6}$	method
1.	McCready, Pawsey, Payne-Scott	150 cm	1945-46	6	1.2	histograms
2.	Pawsey, Yabsley	150	1947	3	0.7	histograms
3.	Lehany, Yabsley	150	1947	4	1.0	histograms
4.	Ryle, Vonberg	171	1947	5	0.6	histograms
5.	Ryle, Vonberg	375	1947	5	1.3	histograms
6.	Reber	187	1944	1	1.8	histograms
7.	Lehany, Yabsley	50	1947	4	0.6	correlation
8.	Lehany, Yabsley	25	1947	4	0.19	correlation
9.	a) Covington	10.7	1947	9	0.078	lowest poi
	b)		1948	12	0.065	correlatio
	c)		1949	12	0.055	lowest poi
	d)		1950	12	0.044	lowest poi

10.	Reber	62.5	1946	3	1.0	mean
11.	Southworth	10	1943	1	0.018	mean
12.	Southworth	3	1943	1	0.018	mean
13.	Piddington, Minnett	1.25	1948	5	0.009	correlation
14.	Covington	10.7	Nov.1946	e	0.056	eclipse
15.	Sander	3.2	July1945	e	0.022	eclipse
16.	Dicke, Beringer	1.25	July1945	e	0.010	eclipse
17.	Naval Research Laboratory	3.15	?	?	0.012	?
18.	Stanier	60	1949	1	0.54	interferomet
19.	Laffineur, Houtgast	55	1948	1	0.24	mean
20.	Minnett, Labrum	3.18	1949	3	0.019	correlation
21.	Piddington, Hindmann	10	1948-49	6	0.054	correlation
22.	Hagen	0.85	1948-49	5	0.0067	pencil beam
23.	Christiansen, Yabsley, Mills	50	Nov.1948	e	No independent calibration	
24.	Machin	670	1950	?	2.0	lowest point
25.	Machin	245	1951	?	2.0	quiet days
26.	Blum, Denisse	187	1950	1	0.9	quiet days

References are on page 118.



The observations collected in this figure are in very good agreement with theory. At meter waves we see the corona with a temperature of a million degrees while in the cm range we look through to the lower chromosphere, which has a temperature close to that of the sun's surface. Precise computations are discussed in the next section. The curve in the figure is based on the computations by Unsöld, corrected by Burkhardt and Schlüter (See p. 112)

Changes with solar cycle.--This table and figure should not leave the impression that the quiet sun is entirely constant. It is even likely that its radiation

should change with the solar cycle. The result obtained from optical observations is that the temperature of the corona is fairly constant but that its density varies (perhaps by nearly a factor 2) with the solar cycle. This variation would be most noticeable at radio wavelengths where a change in optical depth of the corona yields a big change in brightness. This occurs at the steep part of the curve, i.e. in the range where T_a is well between 5000° and 1000000°. Observations so far show precisely what is expected. The lowest antenna temperatures measured by Covington at 10.7 cm were 295° in 1947; 235° in 1948; 210° in 1949 and 165° in 1950. When reduced to zero sunspot area the first and last of these values may become something like 260° and 160°. Multiplication by 265 (compare problems, section 3.2) then gives a drop from T_a = 69000° in 1947 to 42000° in 1940. The Australian observers have found a similar effect in the 10cm region but none in the meter range. Southworth's observation of 1943 now does not seem so far out of range as it seemed before.

Intensity distribution across the solar disk. The methods for localizing bright spots on the sun can in principle also give the intensity distribution over the entire disk.

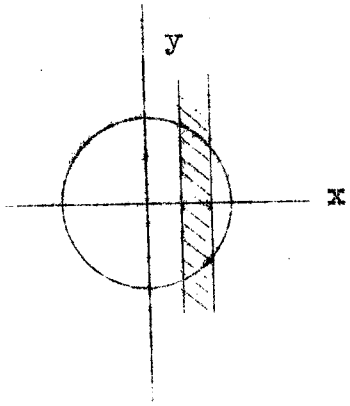
(a) Eclipses have been used in some instances but the interpretation of the records often is uncertain. This is inherent in the problem. The moon is just big enough as a screen for the visual sun but too small to give convenient measures of the brightness distribution in the radio sun. As mentioned before, active spots show up when they are covered and uncovered. But the evidence for limb brightening from eclipse observations is so far inconclusive. One convenient parameter is the fraction of the total radiation that is left when the visual disk is totally eclipsed. This fraction is less critically dependent

on the ratio of the sun's and moon's disk than optical measures of the corona brightness, because there is no sharp change at the edge. Observed values of this fraction are:

	λ	fraction of radiation left at deepest point
Eclipse May 1947, Bahia. (Hagen; Haikin and Chikhachev)	3.2 cm	0.04
	150	0.40
Eclipse Sept. 1950, Alaska (Hagen, Haddock, Reber)	3.2	0.08
	10.7	0.17
	65	0.28

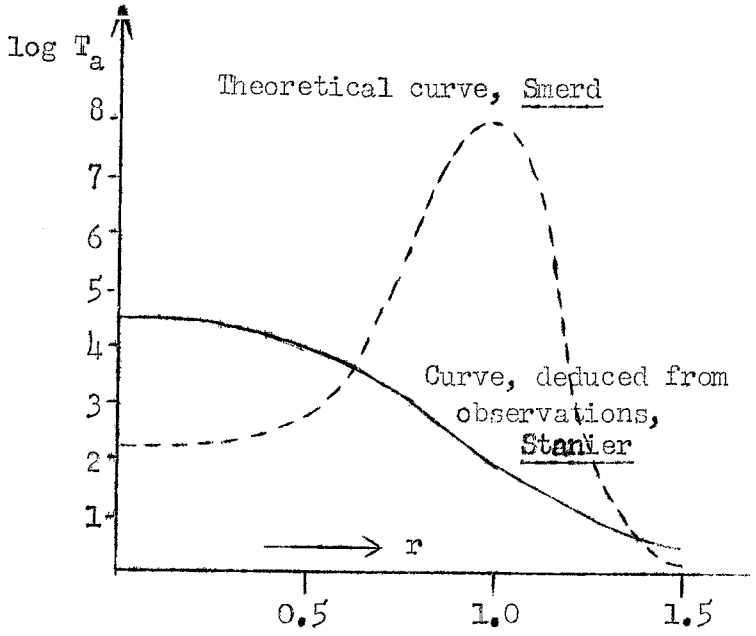
Such observations, when corrected to give fractions of the total quiet sun's radiation will eventually be helpful in settling the matter of limb brightening. Observations of partial eclipses in which less than 80 per cent of the disk is covered are less useful for this purpose (Sander; Dicke and Beringer; Covington; Minnett and Labrum; Piddington and Hindman; Christiansen, Yabsley and Mills; Laffineur, Michard, Steinberg and Zisler).

(b) Interferometer measurements have been applied to the problem of finding the brightness distribution across the disk only by Stanier at 60 cm.* He measured the brightness contrast in the interference fringes for a number of different separations of the interferometer elements. This contrast as a function of the antenna separation is essentially the Fourier transform of the intensity distribution I(x), where (see figure) I(x) dx is the integrated brightness in a strip across the sun parallel to the y-axis or meridian. Then, assuming circular symmetry he finds the brightness distribution with r from I(x) by solving an integral equation. The result is highly important, for it flatly



* Now also by Machin at 2.45 meter, see reference on p.118

contradicts the theoretically predicted limb brightening. The observed curve is shown below, together with the theoretical one. Among the theoretical attempts to explain this discrepancy the assumptions of a latitude dependence would not help; only a strong brightness near the pole could explain the observations but that contradicts what we know of the optical form of the corona. Also a different temperature or density distribution with r does not help, for on any model the brightness of the partially transparent corona will be greatest for the rays that travel farthest through it, i.e. the rays passing just outside r = 1.03. Van de Hulst has tried to solve the discrepancy by invoking the coronal streamers, i.e. irregularities in the density of the outer corona. The limb brightening may be made invisible by the irregular refraction by the coronal streamers in front of the regions where the radiation is emitted. Thereby the distribution with a bright limb will be smeared out as if we looked through opal glass. Qualitatively the effect may be explained this way, but quantitative estimates are difficult since the densities and frequency of occurrence of the coronal streamers are uncertain.



6.3 Theory of the quiet sun

The absorption coefficient and refractive index.--All theoretical computations are based on Kramers' law for absorption by free-free transitions, complemented

in the meter waves with Lorentz' formula for the refractive index. The equation derived in section 6.1 is, with the atomic constants replaced by their values in cgs units:

$$\varkappa(\nu) = 0.0097 \sum \frac{N_i N_e Z_i^2}{T^{3/2} \nu^2} \ln \frac{p_a}{p_b} .$$

A summation sign has been inserted to indicate that we have to sum over different kinds of ions. We eliminate it by writing $\sum Z_i^2 N_i = \beta N_e$. Most authors use $\beta = 1$ but the most likely value (for 1 He ion to 10 H ions and few heavier elements) is $\beta = 1.2$. Also we have to estimate the impact parameters p_b and p_a. Again, different authors have used different estimates, partially because in the beginning the situation with the 4 possible combinations (see 6.1) was not correctly understood. If we take T and N_e for representative layers contributing to the radiation at each wave length, we obtain the table:

wave length	T	N_e	p_b(cm)	p_a(cm)	$\ln(p_a/p_b)$
1 cm	10^4	10^{10}	1.5×10^{-7}	2×10^{-4}	7.2
10 cm	10^5	10^9	1.5×10^{-8}	10^{-3}	11.1
1 m	10^6	10^8	1.5×10^{-9}	2×10^{-3}	14.2
10 m	10^6	10^7	1.5×10^{-9}	5×10^{-3}	15.0

Thus, we may write Kramers' law in the form:

$$\varkappa(\nu) = \alpha N_e^2 T^{-3/2} \nu^{-2} n^{-1},$$

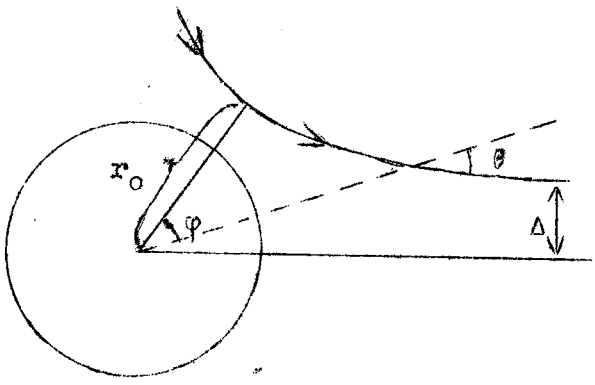
where $\alpha = 0.0097 \times 1.2 \times \ln(p_a/p_b)$ is of the order of 0.17 for meter waves (corona) and half that value for cm waves (lower chromosphere).

Lorentz' formula in cgs units is

$$n = 1 - \frac{8.1 \times 10^7 N_e}{\nu^2} .$$

Computations based on a "model" of the quiet sun. A "model" of the sun may be any assumed distribution of N_e and T_e in chromosphere and corona. So far, only spherically symmetric models have been used for computations of the radio-emission. Here N_e and T are given as functions of r. The whole computation then involves the following steps.

(1) Ray tracing: In a spherically symmetric atmosphere with changing refractive index any ray is given by $nr \sin\theta = \Delta$ (= constant). This is a formula from the "old books" on refraction in the earth's atmosphere; θ is the angle with the normal. The meaning of Δ is seen in the figure. Because n is smaller than 1 in the corona, the ray trajectory turns outwards. Its shortest distance to the center, r_o, is found as the solution of the equation $r \cdot n(r) = \Delta$; this equation can be solved graphically.



As known functions of r we now find first $\sin\theta$, then $ds = dr \sec\theta$ and $d\varphi = dr \cot\theta/r$ (by simple substitutions) and finally φ itself (by integration). The ray trajectory is then known in polar coordinates (r, φ) and can be drawn if we want to.

(2) Integration along a ray, giving optical depth: we have $d\tau = \varkappa ds$ (a factor n^{-1} is contained in $\varkappa$). By integration over r we find the optical depth for any point. In particular for the turning point $\tau_o = \int_{r_o}^{\infty} \varkappa ds = \int_{r_o}^{\infty} \frac{\varkappa d\tau}{\cos\theta}$. Along the entire trajectory in and out the optical depth is $2\tau_o$. Quite as important as optical depth is the travel time, which plays a role in the interpretation of time lags of bursts. The group velocity is $v_g = nc$ (c is the

velocity of light), so $dt = \frac{ds}{nc}$, which may be integrated to give the total travel time from any point along the trajectory. These times become appreciable only if n comes near 0, which happens only if Δ is close to 0, i.e. for points seen near the center of the disk.

(3.) Second integration along the path, giving T_b: This is the integration that we have discussed in section 3.3 (chapter 3). The formula is

$$T_b = \int_0^{2\tau_0} T(\tau) e^{-\tau} d\tau.$$

Most authors simplify the computation by assuming both the corona and the chromosphere to be isothermal. The formula then becomes quite simple. For instance, a ray passing only through the corona (such are all rays for meter waves and the rays with $\Delta > 1.03$ for all wave lengths) gives

$$T_b = T_{cor}(1 - e^{-2\tau_0}).$$

Rays passing through the corona over an optical depth τ_{cor} and finding the chromosphere virtually opaque (any point on the disk for waves below one meter) give

$$T_b = T_{cor}(1 - e^{-\tau cor}) + T_{chr} e^{-\tau cor}$$

Even further simplifications are possible if τ_0 or τ_{cor} are either very small or very large compared with 1. The resulting simple expressions for T_b all have clear physical meanings in terms of opacity or transparency of the layers. The discussion is left to the reader; it is instructive to study closely the figures given in the papers by Smerd and others with these simple interpretations in mind.

(4.) Integration over the sun's disk, giving T_a: The formula is

$$T_a = \frac{1}{\Omega_s} \int T_b d\Omega \quad = \int_0^{\infty} T_e d(\Delta^2),$$

where Δ is as before the projected distance of the measured point from the center of the sun's disk. In this formula (and in all papers on the subject) Δ is expressed in radius of the visual disk as a unit.

Illustrative examples.--Many authors (Unsöld; Burkhardt and Schlüter; Waldmeier and Müller; Hagen; Denisse; Smerd; Nicolet) have made more or less complete computations on a model of the quiet sun. They vary in the assumed values of α and in details about the temperature and density distribution. None of these models is final and in fact a final model does not exist because the sun itself is changeable. We list here some places where illustrative diagrams or tables may be found (not complete!).

Ray trajectories (also showing "reflection" of central rays)	Burkhardt and Schlüter, Fig. 1 Jaeger and Westfold, Fig. 3 & 4.
Values of τ for the central ray, various λ.	Smerd, Fig. 2. Jaeger and Westfold, Fig. 7 & 8.
Travel times for central ray, various λ.	Jaeger and Westfold, Fig. 5 & 6.
Distribution of T_b across disk.	Unsöld, Fig. 5 (without n-effect) Hagen, Fig. 40 (only 3.14 cm) Smerd, Fig. 4
Idem, but plotted for several fixed Δ as a function of λ (this is more useful for interpolation)	Smerd, Fig. 5
Total radiation of sun, expressed as T_a (these curves have to be compared with the observed points, preceding section)	Smerd, Fig. 8, Fig. 9 Unsöld Denisse Hagen, Fig. 7
Total radiation of sun, expressed as flux density	Smerd, Fig. 7

The most complete work is Smerd's. It may be used as a standard reference except for waves shorter than 10 cm, because the bulk of his computations are for a chromosphere at $T = 30000^{\circ}$, which seems too high. Hagen gives very complete tables and figures for the cm region using a cooler (non-isothermal) chromosphere.

Direct inferences about T and N_e from radio data. The purpose of model calculations is to vary the assumptions in the model until a satisfactory agreement with the observations is obtained. This is not only very laborious but one always wonders whether the solution is unique and how critically the agreement depends on the various assumptions. It would be much preferable to find a more direct solution giving inferences about T and N_e from the observations. This is not possible in general (whenever integral equations are involved such solutions become unreliable or impossible). But one important example in which a solution may be made we shall discuss in some detail.

Level of emission of 1 cm radiation.--The observed value of T_a at 1 cm is 8000°. Some limb brightening and radiation outside the limb is expected so T_b at the center must be lower, say 7000°. Since the chromosphere is opaque (the turning point is in the photosphere, i.e. at very large optical depth) this brightness temperature must correspond to the actual T at some "effective" layer in the chromosphere. We may assume that this layer is near $\tau = 1$ and wish to find where this layer is located. Let the electron density at this layer be N_o and let the density gradient be represented by the scale height H so that $N_e = N_o e^{-h/H}$, where h is the height above the level sought. Further let T be virtually constant. Then, with $\varkappa = \alpha N_e^2 T^{-3/2} \nu^{-2}$ $(n = 1)$ we find by direct integration:

$$\tau = 1 = \int_o^\infty \varkappa \, dh = \alpha T^{-3/2} \nu^{-2} N_o^2 \, H/2,$$

which gives $N_o^2 = \frac{1}{\alpha} T^{3/2} \nu^2 \cdot \frac{2}{H}$.

With the numerical values, T = 7000°, $\alpha = 0.08$, $\nu = 3 \cdot 10^{10}$ c/s and H = 1000 km $= 10^8$ cm, this gives log $N_o = 10.1$ so $N_o = 1.3 \times 10^{10}$ cm^{-3}. Reference to the table in section 5.1 shows that this density is reached in the lower chromo-

sphere, somewhere near h = 2000 or 3000 km. (here h again has its usual meaning of height above the surface).

The advantage of such a more direct computation is that one can more easily judge the influence of various uncertainties. First we make sure that n is indeed close to 1 (within 0.1 per cent) and that the contribution of the corona is negligible (10^{-5}); this computation is left to the reader. Then we wonder about the scale height. This may be far different and also there may be a filamentary structure. But the absolute minimum is H = 100 km as in the upper photosphere. This changes N_e by a factor 3 only, bringing log N_e to 10.6; this is still at, or above, h = 1000 km. The greatest uncertainty thus is not in finding the density but in estimating where this density is reached. In Woolley and Allen's model N_e has a local maximum at the layer h = 6000 km, where the ionization sets in. Here the value log N_e = 10.4 is reached, so probably this layer is opaque for cm radiation. In conclusion we may say that the 1 cm waves come from layers that are at the same height or higher than those to which Redman's observations of Balmer line widths refer (see 5.2).

Temperature gradient between chromosphere and corona.--To determine this directly from observations, as Hagen and Denisse have tried, is already a good deal more difficult, certainly if we consider our ignorance of the exact density distribution. The correct method is to make analyses of the kind shown above also for wave lengths of 2, 3 and 4 cm. At 5 cm the corona contributes already half and at 10 cm most of the received radiation. This is as far as we can judge from present computations and from the observations that refer to the years near maximum activity, 1948-49. It is important to note that the transition from low to high T in any case must be fairly sharp. The theoretical

curve based on the assumption of an abrupt change from 5000° to 1000000° (Unsöld-Burkhardt-Schlüter), that is drawn in the graph of the preceding section, fits the observations very well except for the 1-5 cm range. This range is exactly where a deviation is expected, due to the gradual change in the upper chromosphere. Hagen, Denisse and Piddington have shown that a model that fits the observations exactly can easily be found.

The sun's general magnetic field. As an extension to the computations on a model quiet sun one may assume a weak general magnetic field of the sun. Magneto-ionic theory then gives a certain percentage of circular polarization of the emitted radiation. The polarization has opposite signs on the northern and southern hemisphere so is invisible for the sun as a whole, but might be detected during a partial eclipse. Observations so far have given negative results.

Results as given by Smerd:	Polarization	H_{pole}
50 cm eclipse	$< 3\%$	< 8 gauss
10 cm eclipse	$< 1.5\%$	< 25 gauss

This is in agreement with the findings of Thiessen, Von Klüber, Kiepenheuer and others that H_{pole} is smaller than 1 gauss, if present at all.

References (far from complete)

Observations of the Quiet Sun: (numbers (1) to (26) refer to list on pages 105 and 106).

J. L. Pawsey & D. E. Yabsley: Austral. Journal. Sci. Res. (AJSR), 2, 198, 1949. (1-16)

F. J. Lehany & D. E. Yabsley: AJSR, 3, 350, 1950. (7,8)

J. H. Piddington & H. C. Minnett: AJSR, 2, 539, 1949. (13)

H. M. Stanier: Nature, 165, 354, 1950. (18)

M. Laffineur & J. Houtgast: Ann. d'Ap. 12, 1949. (19)

J. H. Piddington & J. V. Hindman: AJSR, 2, 524, 1949. (21)

W. N. Christiansen, D. E. Yabsley, B. Y. Mills: AJSR, 2, 506, 1949. (23)

M. Ryle: review in Reports on Progress in Physics 13, 184, 1950. (24)

J. P. Hagen, NRL report 3504, 1949. (17,22)

H. C. Minnett & N. R. Labrum, AJSR, 3, 60, 1949 (20)

Computations:

S. F. Smerd: AJSR, 3, 234, 1950, and 3, 265, 1950.

J. F. Denisse, Ann. d'Ap 13, 185, 1950.

M. Waldmeier and H. Müller, Astr. Mitt. Zürich 154 and 155, 1948.

J. C. Jaeger and K. C. Westfold, AJSR 3, 376, 1950.

J. H. Piddington, Proc. Roy. Soc. A 203, 417, 1950.

A. Unsöld, Naturwissenschaften 34, 194, 1947.

G. Burkhardt and A. Schlüter, Z. f. Astrophysik 26, 295, 1949.

Further references to observations, added later:

K.E.Machin, Nature 167, 889, 1951 (25, my integration of data from fig.2)

E.J.Blum & J.F.Denisse, Comptes Rendus 231, 1214, 1950 and 232, 387, 1951 (26)

Chapter 7. RADIO EMISSION FROM THE ACTIVE SUN (H. Zirin)

7.1 Introduction and survey

A survey table of solar radio emission components is in section 6.2. (page 102)

History.--The first indications of strong non-thermal radio emission from the sun (although not then interpreted as such) were the reports by amateurs of hissing noises heard on the 7.5 to 30 meter bands. At the same time, strong noise was reported by professionals during radio fadeouts. The first published reference to this noise was by Arakawa in 1936. In 1942 high-level noise caused severe interference with radar equipment operating on the 4 to 6 meter band. Investigation by a group under Hey revealed that the bearing of the source was always within a few degrees of the sun. In 1944 Reber, during his survey of galactic noise at 1.85 meters, noted that the solar noise was approximately as intense as that from Sagittarius (1).

In February 1946 more detailed study of solar noise started, and the British and Australian workers came up with these results (almost all these observations at $\lambda > 1$ meter): The radiation from the active sun consists of three components, viz:

(1) "enhanced radiation" with a duration of days.

(2) "bursts" with a period of seconds.

(3) "outbursts" with a duration of about ten minutes and very high intensity.

Further details of the enhanced radiation were soon found:

(1) The radiation is circularly polarized. (2), (3).

(2) The enhanced noise is strongest when there are large spot groups near the meridian. (3), (4), (5).

(3) The enhanced noise actually comes from small areas on the sun, in the vicinity of sunspots. (6).

The outbursts are connected with flares, and are the most intense radiation from the sun. Some records of outbursts still go off scale at 10^{-15} watt m^{-2} $(c/s)^{-1}$, which is already 10^6 times the normal intensity received from the sun. The most intense flares have almost invariably been connected with outbursts, although for lesser outbursts the correlation is somewhat obscure (7).

Since 1946 a great deal of quantitative and qualitative data have been amassed. Two improvements of basic interest may be mentioned:

(1) Most of the bursts occurring during the enhanced noise storms are circularly polarized. A close connection is therefore suggested. Other bursts, and, so far as we know, all outbursts are unpolarized.

(2) The development of spectral work, especially by Wild (8), beginning with simultaneous observations at neighboring frequencies and developing into regular spectroscopy.*

Programs.--A few words may be said here about programming new observations. The possibilities for new work are manifold but, depending on two rather different aims one may have in mind, the lines of research will be different. If forecasting of ionospheric disturbances, etc., is the aim, routine work with standard methods, every day and preferably at all longitudes on earth, is advisable. The results may be disseminated in some kind of code (see the Quarterly Bulletin of Solar Activity) and large-scale (brute) statistics may be made in the hope of finding significant correlations. If understanding the physical

*Wild's spectral work is now being extended to include the range from 30 Mc/s to 300 Mc/s. (Dafte)

phenomena is the first aim, statistics are also needed but representative samples over a wide frequency range are more important than complete data at one frequency. We are more willing to put odds and ends together, and instead of using standard equipment instruments for new research have to be developed. For these reasons the Sydney group has developed the spectrum analyzer and the fast interferometer, both inspired by the idea that radiation of high intensity (as are bursts and outbursts) permits observations with very small time constants so that much detail on one transient phenomenon can be collected. An effort to understand everything theoretically has to go hand in hand with the observations.

Survey of intensities.--It is more useful to express the intensity of the emission from the active sun in terms of flux density than in terms of apparent temperature. The relation between them is, in MKS units:

$$S = \frac{1.9 \times 10^{-27}\, T_a}{\lambda^2} \quad (\text{watt m}^{-2}\ (\text{c/s})^{-1})$$

The T_a is defined by comparing the radiation with that from a hypothetical black body with the size of the sun's disk. The active region emitting this radiation may be e.g. 0.01 times the disk's area. The average brightness temperature over that region is then 100 times T_a.

It would be desirable to have at various frequencies a graph showing for just how large a fraction of the total time a certain level S is exceeded. A first attempt at constructing such a (p, S) diagram from odds and ends is made here; p is defined as the fraction of time during which the flux density is larger than a given value. We choose the wave length around 4 meter, or frequency 75 Mc/sec. Pawsey's histograms give the levels belonging to

p = 1, down to 0.1 and sometimes 0.02. For instance, p = 0.1 refers to a steady level of enhanced noise in one day of ten. The histogram of Ryle's observations at 80 Mc/sec was used. Appleton and Hey (13) give 5-minute averages of S during many hours per day of a particularly disturbed period of 12 days in February 1946. The fractions of observing time during this period can be found from simple counting. We assume that such a period might occur once a year and thus scale down all values of p by a factor 30. For instance, $S = 3 \times 10^{-18}$ was exceeded during 4 x 5 minutes in 75 observing hours, which makes p = 0.004 but scaled down by 30 it gives log p = -4. Finally, reports on very big outbursts give an estimate on their p and S. The largest one on record during several years was the outburst of March 8, 1947, reaching $S = 10^{-15}$ during 1 minute. Dividing by 500 observing days of 8 hours (assumed) we obtain $p = 2 \times 10^{-5}$. Collecting the data we have:

log p	log S	log pS	
0	-22.0	-22.0	
-0.5	-21.2	-21.7	Based on Ryle and Vonberg, mainly
-1.0	-20.6	-21.6	enhanced radiation
-1.5	-19.9	-21.4	
-2.0	-19.0	-21.0	
-2.5	-18.3	-20.8	Based on Appleton and Hey, enhanced
-3.0	-17.9	-20.9	radiation and outbursts
-3.5	-17.7	-21.2	
-5.0	-15.3	-20.3	Payne-Scott, big outburst.

It is remarkable that these very rough estimates together define a good linear relationship. An investigation of this kind is needed if we want to compute the average output of the sun at this frequency. The fact that p·S gradually increases means that the big outbursts in spite of their short duration and rare occurence contribute strongly to this average. In fact, we may write

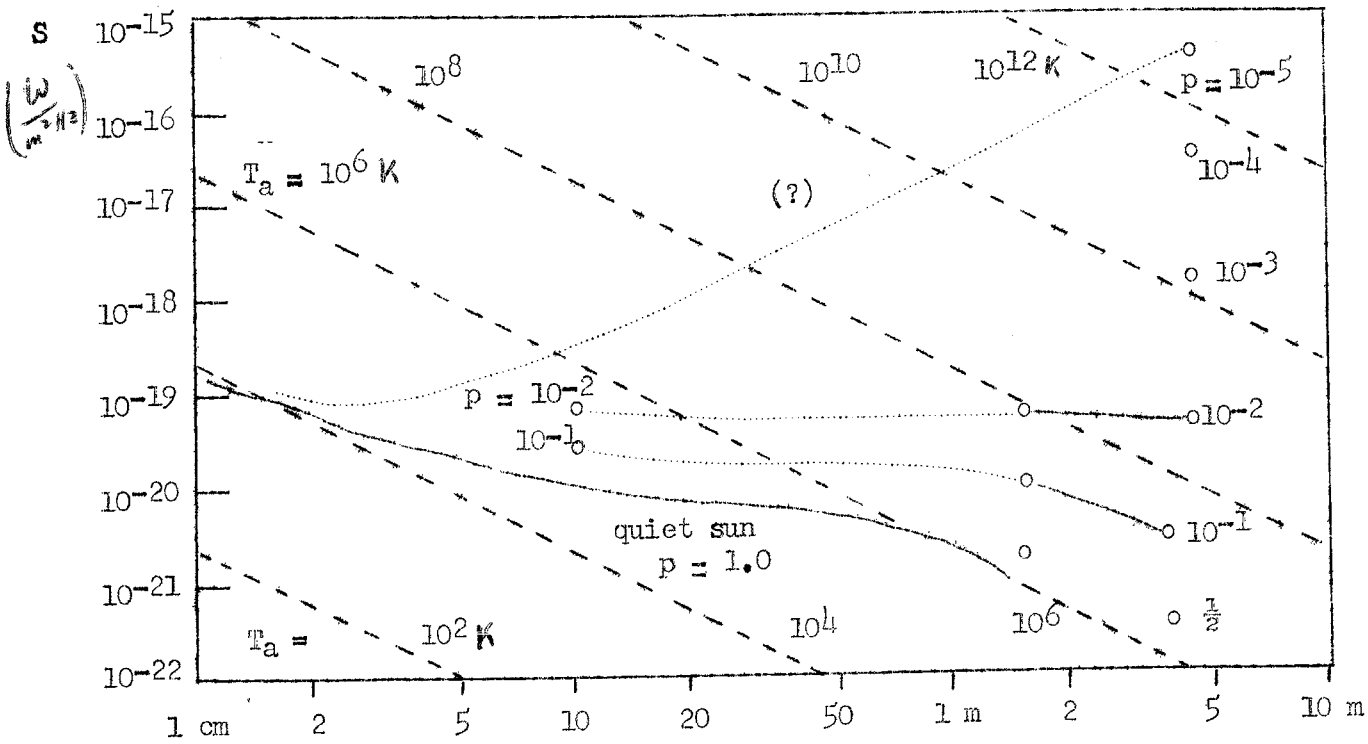
$$S_{average} = \int_0^1 S(p)dp = 2.3 \int_{-\infty}^0 pS(p)d(\log_{10} p)$$

and thus find that the figures do not yet indicate where the integrand of the second integral goes down. Assuming that it goes down soon after the last point we obtain roughly:

$$S_{average} = 2.3 \times 4 \times 10^{-21} \text{ (value of pS)} \times 3 \text{ (range of log p)} = 3 \times 10^{-20}.$$

This corresponds to $T_a = 2.5 \times 10^8$ or 250 times the (presumed) value of the quiet sun at 4 meter wave length.

Spectrum.--It is clear that the phrase "the spectrum of the active sun" has no meaning unless precisely defined. Either we must study individual phenomena by observing them simultaneously on different frequencies, or we should try to construct (S, p) diagrams at various frequencies and say that values of S corresponding to p = 0.1 define the spectrum of the sun at the p = 0.1 level, etcetera. The quiet sun then is the spectrum at the p = 1 level. Very few good estimates are yet available. The figure gives an idea of how a complete graph of the spectrum of the active sun would look.



— vdH never uses the term "slowly varying component" or cites Dennis as "discoverer"

7.2 Enhanced radiation at cm waves; correlation with sunspots

All observers in the range of wave lengths below 1 meter have noted a fairly good correlation between the daily steady level of enhanced noise and the area of the visible sunspots. Let T_a be the apparent temperature of the sun and let A be the sunspot area in the usual definition, i.e. with 10^{-6} of the area of the disk as a unit (The Australians have used 10^{-5}). Then the correlation graph defines a linear relation of the form

$$T_a = T_o + \alpha A.$$

There are still ambiguities in the measurement of area, both in the definition (true or projected) and in the practice of measurement. The extrapolated T_o, which is the apparent temperature of the quiet sun (section 6.2) is free from these ambiguities. The values of α, listed below, are affected by it. Most slopes were newly read from the published graphs.

Series	Authors	Wave length (cm)	α
13	Piddington and Minnett	1.25	0.2
20	Minnett and Labrum	3.18	0.8
21	Piddington and Hindman	10	6
9	Covington	10.7	14
8	Pawsey and Yabsley	25	75
7	Pawsey and Yabsley	50	130

(really Lehany+Yabsley, but also reported by P+Y)

At longer wave lengths the correlation is poor but if one wants to draw a straight line at all the slope indicates values of α from 100 to 150 (up to 1.5 meter).

Emission by coronal condensations.--If the size of the visible spot is taken as the actual size of the emitting region, then the average T_b in these re-

gions is α times 10^6 degree. It seems more likely that the emitting region coincides with a "coronal condensation" above a spot. A typical spot may have A = 300 and the corresponding coronal condensation may have A = 5000, which is 16 times as much. Direct estimates of the size of the emitting regions may be made from the time that it takes to cover them during an eclipse. In this way Piddington and Hindman estimated A = 2000 for one region at 10 cm and Christiansen, Yabsley and Mills at 50 cm detected eight "hotspots" estimated at A = 4000 each, which makes 32000, or nearly 40 times the then visible spot area. Such measurements are still of great importance.

Suggested explanation.--Increasing the area by a factor 20, for instance, brings the brightness temperature over this area down to $\alpha/20$ times a million degrees. So T_b is smaller than 10^6 in the areas of enhanced radiation for wave lengths of 10cm and smaller. This suggests thermal emission from a partially transparent layer of the corona. The idea is thus that the corona is made locally more opaque, simply by a higher density and that this effect causes the extra radiation. No change in T and no deviation from the quiet sun mechanism is involved. Waldmeier and Müller (37) have worked out this idea for a model coronal condensation and find good agreement at 10 cm, i.e. Covington's wave length, where the most complete observations have been made. A further point that seems to agree with the suggestion of thermal emission is that Covington (unpublished) finds no change in α with the years, like he finds in T_o. This agrees with the assumption that the corona changes its density with the solar cycle and not its temperature.

There are, however, complications. An obvious one is the magnetic field of the spot. The opacity of the condensation will be different for the ordinary

and the extraordinary ray so the emitted radiation will be circularly polarized. Covington (38) found indeed a polarization of the order of 5 per cent referred to the entire sun, i.e. of the order of 15 per cent for the extra radiation then present.

A serious difficulty arises when one tries to make a theory that fits all the cm observations simultaneously. One expects an increase of α proportionally to λ^2 as long as the condensations are optically thin and a less steep increase for the wave lengths for which they become opaque. The observed trend is in the right direction but is too steep: α increases somewhat more rapidly than λ^2 and inadmissably high values of T would be needed to explain the continuing increase beyond 10 cm. This is not discouraging. We know that in the meter waves non-thermal processes cause the enhanced radiation. At the same time the character of the radiation is changed by the appearance of stormbursts, the narrower angular emission diagram, and the complete polarization. It seems quite likely that already at 10 cm the thermal emission is slightly boosted by the unknown mechanism causing the enhanced radiation at meter waves. Piddington has suggested that the thermal and non-thermal mechanism could be co-existent and distinguishable in the range from 25 cm to 2 meter; more probably there are not two mechanisms but one whose character gradually changes.

7.3 Enhanced radiation at meter waves; noise storms and storm bursts

It is clear that there is a complex of phenomena, all circularly polarized, comprising the enhanced radiation and associated bursts. Allen calls this complex "a noise storm, with storm bursts." Payne-Scott has termed the burst

phenomena "variations in the enhanced radiation," but in view of their short duration this is really not a good term. Moreover there actually are variations lasting one to two minutes of ionospheric origin. These are called "non-selective fading" by Pawsey and just "fading" by Booker.

The variation in designation is not a confusion of terms but a difference in viewpoint. The real question is: Which of the following concepts is right?: A. Enhanced radiation is itself steady, and the bursts are an additional phenomenon superposed on it. B. Enhanced radiation is merely a great number of bursts, giving the integrated effect of a steady level. Pawsey analyzes his data on the basis of the first line of reasoning. This appears to be correct, in view of the fact that sometimes (but seldom) the level is steady and high, with few bursts, as noted by Allen (7) and Øvren (Cornell) (unpublished). The data on polarization given by Payne-Scott (15) provides strong corroboration. The second point of view has been advocated by Ryle and Vonberg. Wild's spectral investigations may further clarify this point. See reference (41).

Appearance of enhanced radiation on tracings.--Illustrations of intensity records of enhanced radiation are given by many authors, and here a picture is worth a thousand words. McCready, Pawsey, and Payne-Scott (6, Fig. 2) show records of observations of the sun both at noon and sunrise from stations located ten miles apart (at 200 Mc). No systematic change with solar altitude is found, and the graphs show the slow variation to be solar and not ionospheric. Allen (7, Fig. 1), shows records of various phenomena. Extremely important is the variable ratio of bursts to noise (background). Records are reproduced of: frequent bursts, high noise level; and no bursts, high noise level (enhanced noise). Outbursts also appear on these records (this is also at 200 Mc). Burrows (9, Fig. 3) gives a good illustration of enhanced noise and of the quiet sun with and without bursts.

Ryle (10, Fig. 11) shows the record of solar radiation at 0.60, 1.70, and 3.70 meters during transit of a spot. The record, obtained with a meridian interferometer, differs at each wavelength. Ryle and Vonberg (11) give a similar illustration at the latter two wavelengths. Pawsey (12, Fig. 12) shows observations of enhanced noise with the horizon interferometer, including the meridian passage of the great sunspot of February, 1946.

The phenomenon of non-selective fading is illustrated by Pawsey (12), but is given on a polarization tracing and therefore not so conspicuous. Records show that fading is similar on different frequencies, e.g. 60 and 100 Mc/s. They show a cyclic variation with a period of minutes. Simultaneous observations from different stations show that it differs in different places, and hence is probably an ionospheric phenomenon. As expected, fading is only observed when the level is enhanced because the radiation then comes from a local source. Observations of fading at 200 Mc/s have been carried out at Cornell by Owren and Booker.

Circular polarization.--The circular polarization of the enhanced solar radio noise was discovered independently by several workers, as already noted. The polarization is fairly complete, except when several sources are present. Martyn's original announcement that the sign of polarization changes with meridian passage of the spot has not been borne out by later observations.

Ryle (10) has attempted to find out what sign of sunspot magnetic field will produce a given sign of circular polarization in the noise. Since two possible regions may produce the spot noise, a knowledge of this would permit us to discriminate as to which zone was producing the radiation. The results are inconclusive, because the resolving power at out disposal is not great

enough. However the indication is, from unpublished work of Stanier (at 1.4 and 3.7 m.) that the so-called "extraordinary wave" is responsible (vid. 7.5, the magneto-ionic theory). This corresponds to a higher atmospheric level than the "ordinary wave." That the "storm bursts" are polarized the same way is shown conclusively by the data of Payne-Scott (15, esp. Fig. 3).

Correlation with spots; angular emission diagram.--The correlation of enhanced noise with spot numbers is not so good here as in the centimeter region. Thus Pawsey, Payne-Scott and McCready (5) initially found a close relation, but later observations showed no such simple connection.

Better established is the strong connection between the enhanced noise and the presence of spot groups near the center of the solar disk. This has been shown from direct correlation statistics, but is even more directly apparent from studies of noise intensity during the days near meridian transit of a large spot group. This was noted right from the beginning of solar radio noise observations. Studies of this kind give quantitative measurements of the angular emission diagram of the spots. Various results are available: Allen (7) found a drop to half-intensity in 1.8 days from the meridian at 200 Mc/s (Fig. 4 of his paper). The number of bursts drops in the same way as the steady enhanced flux. Hey, Parsons and Phillips (14) find, at 1.4 m., a drop to half-intensity in 1.2 days. Their Figs. 6-7 look much steeper than Allen's. Table V of this paper shows that there also is a latitude effect, favoring spots on the same heliocentric latitude as the earth. They conclude that the sunspot emits a fairly directional beam, about 40 or so degrees wide. Work by Machin (reported by Ryle (10), pp. 209-210) shows that at 3.75 meters the intensity drops to one half in $\frac{1}{2}$ day, and at 1.7 meters it does so in one day. The emission

diagram seems to get narrower with increasing wave length. In line with this trend we may assume that at 10 cm. the projected area of the spot is important; this gives a drop to half intensity in 60° of solar rotation, i.e. in about 5 days.

Observations to ascertain the precise region of emission were carried out with interferometers by McCready, Pawsey and Payne-Scott (6), as well as by Ryle and Vonberg (4). These results localize the radiation in relatively small areas in the vicinity of the spot. The size of the emission regions is less than 10' across, so that their area is less than one-tenth the solar disk. Big spotgroups may have an area of 1/200 of the disk, so that a reasonable estimate of the region of emission is 1/100 of the disk.

Storm bursts. Much of the material relating to storm bursts, which is the term used for the bursts superimposed on and associated with the enhanced radio noise, has already been discussed. The best way to study these bursts is the spectrum analysis technique used by Wild and McCready (8).* Their spectrograms (see Plate 2a of their paper) show that the width of the disturbance is only 4 Mc/s at 100 Mc/s (between ¼ power points; between ½ power points it is only 3 Mc/s) and the bursts last for 1-20 seconds (the one in question lasted 8 seconds). The frequency of maximum is closely constant in time, although the height of maximum may fluctuate during the fleeting life of the burst. They call this radiation of spectral type I (Type II = Outbursts and Type III = isolated bursts).

Reber (unpublished) has studied simultaneous tracings at 156 and 160 Mc/sec giving special attention to the sharpest single peaks, which he calls "pips" and which presumably correspond to stormbursts. Time and duration of the pips at these frequencies are coincident within 0.1 second but the amplitudes

*Also (41).

are rarely the same. Thus a spectral width of the order of 4 Mc/s is again indicated. Reber (unpublished) has also tried to estimate the duration of the pips (avoiding any more complex groups of bursts) at different wave lengths. He finds, e.g. at 62.5 cm, that the duration is quite independent of the top intensity of the pip, which may range from 0.1 to 100 times the background intensity. All these bursts have durations between 0.3 and 0.8 seconds, with an average of 0.5 seconds. Similar data are obtained at other wave lengths and are shown below. (see also figure on next page)

wave length	time (seconds)
10 cm	too fast to be determined
20 cm	too fast to be determined
62.5 cm	0.5 (0.3 - 0.8)
1.8 m	1.5 (0.8 - 2.5)
6 m	5 (3 - 10)
2.8 m	3

The last value was estimated from Wild's graph; the others are quoted from Reber's unpublished work and the values after the final reduction may be different from these. Yet already a remarkable relationship appears: the duration in seconds is about equal to the wave length in meters.* Quite probably an understanding of the stormbursts or pips will be crucial in any theory of enhanced radiation.

*Reference (44), referring to bursts during weak storms, does not support this relation. Point BD in graph.

7.4 Isolated bursts and outbursts

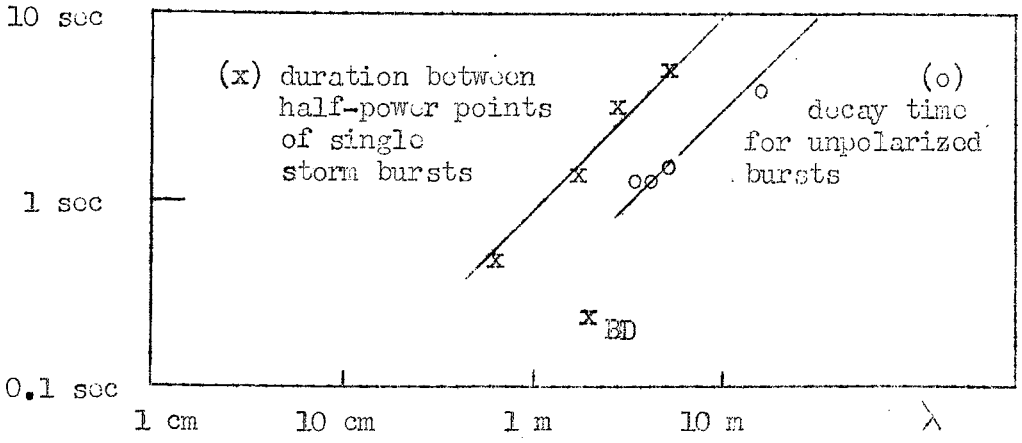
Isolated bursts. The isolated bursts are of the same order of magnitude as the storm bursts, so far as the power radiated is concerned. Close study, however, shows that they differ in several important respects:

1. They are unpolarized, as has been shown clearly by Payne-Scott (15).
2. They occur almost simultaneously at various frequencies.
3. Their spectrum differs markedly and shows various types.

Payne-Scott gives a good summary of single frequency data on numerous unpolarized bursts, both on 60 and 85 Mc/s. This material clearly shows a lag in the time of arrival on lower frequencies. The form of the burst is a rapid rise, rounded top, and exponential, but fast, decay. Approximation by an exponential of the form e^{-at} gives, for the descending curve:

a	1/a	freq
0.6 sec^{-1}	1.7 sec	19 Mc/s
0.7	1.4	60
0.25	4.0	75
0.7	1.4	85

The 75 Mc/s datum is by Williams, quoted by Payne-Scott. We again have a proportionality to wave length, but the data are meager. The figure below shows the time-wavelength graph for both kinds of bursts.



The illuminating paper of Wild and McCready (40) gives us excellent information on the isolated bursts. They observed the spectra of 32 isolated bursts, or groups of them, in the frequency fange 70-130 Mc/s. Some bursts have several maxima. They note that an average of 0.6 of the lifetime is spent on the descending branch of the intensity curve.

Among single maximum bursts, the frequency of maximum either remained approximately constant or drifted rapidly toward lower frequencies at a rate of around 20 Mc/sec^2. The latter are termed spectral type III. The complex bursts (those with more than one maximum) seem to conform to one or the other type. Bursts that occur within short periods of one another (order of one minute) show markedly similar spectra, and they seem to occur in decreasing order of amplitude. Payne-Scott points out that the single hump burst is uncommon, and the most frequent is the double-humped (21 out of 45, vs 5/45 single). In 17 of these the first peak was larger. An echo phenomenon is suggested.

Outbursts. The outbursts, of course, are the most spectacular of the radiations from the "active sun." Intensity recordings of outbursts have gone off scale at 10^{-15} w m^{-2} (c/s)$^{-1}$, corresponding to an increase by a factor of at least 10^7 in the power received (25) (26). Furthermore, correlation of outbursts with flares anywhere on the sun's disk indicates a wide beam width.

The most valuable work here is again the spectral work of Wild and McCready (8). The spectrum reveals extreme complexity, but there always is a sharp start. This start occurs later at lower frequencies at a rate of $\frac{1}{4}$ Mc/sec^2 in all the five cases observed by Wild.* The outbursts are classed as spectral type II. This work only covers 70-130 Mc/s, and it is important that it be extended to other bands, and be supplemented with the data on other wave lengths in the literature.

The outbursts have been correlated with solar flares, as well as geophysical phenomena. Ryle (10) says: "There seems little doubt that the largest outbursts are nearly always associated with intense flare activity, but that for lesser disturbances the correlation is very much worse."

*See reference (39) for full details.

Material on solar flares is available in a number of papers; general, Hale (16), Waldmeier (17); detailed description of individual flares, Ellison (18), Miss Dodson (19); statistics, Waldmeier (20); methods of observation, Waldmeier (21); theory, Giovanelli (22). Flares were known to cause sudden ionospheric disturbances (23), (SID = fadeout); ultraviolet radiation of the sun establishes an ionospheric D-layer in which the absorption of 16-25 meter waves is so strong that they cannot get through. Also galactic noise has been found to suffer a fadeout. Further geophysical effects are magnetic crochets (24) and intense noise at km waves; and a day later:--aurorae and magnetic storms, caused by corpuscular emission from the flare. Evidence has been accumulated in recent years that also cosmic rays, i.e. a very energetic corpuscular emission, are emitted by the flares.

All of these phenomena may in turn be correlated with outbursts. A good report was given by Hey, Parsons and Phillips (14); Miss Dodson has a very large collection of current data.

Time delays.--Many data have been accumulated on time delays of the arrival of outbursts on different frequencies. Wild (39) reports a delay of 170 seconds between 70 and 110 Mc linear in this entire range; Payne-Scott, Yabsley and Bolton (25) give 200-100 Mc/s, 2 min., 100-60, 4 min; Ryle (10, p. 212), gives 500-175 Mc/s 0.9 min, 175-80, 0.1 min, 80-45, 0.5 min; Covington and Medd's (27) observations on 3000 and 200 Mc are inconclusive. Continuing delay with decreasing frequency does not seem to be a strict rule. Reber (1951) (28) notes the following interesting delay estimates for one outburst:

9500-3200 Mc/s, 2 min, 3200-1400, 1.5 min, 1400-160, 1.5 min, 160-75, 2 min, 75-51, 2 min, 51-35, 4 min, and the disturbance reappeared at 51 Mc/s

three min later; at 25 Mc/s no disturbance was received. If a continuing delay may be interpreted in terms of matter shooting right off the sun, this case would indicate matter shooting up high in the corona, not quite reaching the 25 Mc/s level and then falling down. Neither interpretation is very secure; see 7.5. A constant of exponential decay can be found from Covington's curves at 10 cm: a^{-1} = 1.0 to 1.5 minutes.

Displacements of the source of the outbursts may be measured with the new fast interferometer developed by Payne-Scott and Little at Sydney. The first results for outburst observations with this instrument are reported by Bracewell (30). The source was found to be in rapid motion across the solar disk away from the flare with which it was associated. It crossed the sun's limb and in thirty minutes had reached a point in the corona more than half a solar radius beyond the limb. It seems of interest that Bracewell's figure does not show the disturbance to move radially outward from its origin. At the end the outburst begins to show some circular polarization.

Two outbursts studied very completely are reported in (42). Read also (47).

7.5 Theoretical interpretation

The most important concept discussed in explaining the source of the radio emission of the active sun* is that of plasma oscillation. A plasma is simply a neutral collection of charged particles. We recall our formula for the index of refraction n for waves of circular frequency ω in a medium of characteristic frequency ω_o:

$$n^2 = 1 - \frac{\omega_o^2}{\omega^2}, \qquad \omega_o^2 = \frac{4\pi N e^2}{m}$$

If $k = \frac{2\pi}{\lambda}$, the propagation number, then the phase and group velocities are given by

phase vel. v: $1/v = k/\omega$ group vel. u: $1/u = dk/d\omega$

*See e.g. (45), (46).

In the present case, for n between one and zero this gives v = c/n and u = nc. As n goes to zero, v becomes infinite and u becomes zero. The physical meaning of infinite phase velocity is that the phase lag between different regions is zero, so that the disturbance can be in phase throughout a large region. The fact that u is zero means that no energy can be propagated. The rigorous theory of plasma oscillations shows many reasons (for instance, already the thermal motion) why u is not strictly zero and energy propagation not strictly forbidden. Much is known, both in experiment and theory, about plasma oscillations, but the application to solar problems is still rather obscure. The narrow spectral width of stormbursts almost certainly indicates that this theory may be of importance in explaining enhanced radiation.

The magneto-ionic theory of Appleton has been applied to the propagation of radio waves through the corona by several authors (32), (33), (34), (43). The theory shows that in the presence of a magnetic field the refractive index of the region is modified, and the propagation along the magnetic field is given by

$$n^2 = 1 - \frac{\omega_o^2}{\omega(\omega + \omega_H)} \qquad \text{extraordinary ray}$$

$$n^2 = 1 - \frac{\omega_o^2}{\omega(\omega - \omega_H)} \qquad \text{ordinary ray}$$

Here ω_H is the gyrofrequency, or frequency of Larmor precession of the electron. Here again certain regions of zero n, suitable for plasma oscillations, occur but for some of these the radiation cannot get out. The circular polarization of the storm bursts and enhanced radiation clearly indicates that magneto-ionic theory will be important in the interpretation of these phenomena. It must be emphasized, however, that a magnetic field alone in connection with a gas in thermal equilibrium will not give a radiation exceeding the thermal

radiation of that temperature. The fact that T_b in enhanced radiation goes far above 10^6 degrees requires non-thermal effects for its explanation.

Delay times.--Further search for enlightenment has been in trying to account for time delays and damping factors. The damping constant might be connected to, or identical with, the frequency of collisions at the level where the disturbance occurs, as Westfold has suggested. This is in qualitative agreement with the fact that a decreases (time 1/a increases) with decreasing frequency, i.e., higher level in the corona.

There are two possible explanations for the time delays. The first attributes the delay to the movement of the disturbance. Suppose the disturbance travels outward through the corona and excites a plasma-oscillatory phenomenon wherever it goes. Then the high frequency levels will be reached first and the lower ones later. Wild applies this idea to his observations (8) and gets for outbursts a velocity of about 500 km/sec, and for isolated bursts a velocity of 20,000 km/sec. Only for outbursts does this interpretation seem likely. The outburst reported by Bracewell (30) appears to move 1.0 R in projection in thirty minutes. The velocity thus is at least 700,000/1800, or about 400 km/sec, in good agreement with Wild's calculation. For auroral particles, the velocity may be calculated from the fact that they travel one A.U. in one day, which gives 1500 km/sec. This is close enough to consider this idea a promising approach.

The second explanation is that time delays are caused by waves traveling from a fixed disturbance. This idea has been advanced by Jaeger and Westfold (35). Suppose that somewhere in the corona a disturbance emits waves of many different frequencies simultaneously. Then the waves for which the frequency

is closest to the local plasma frequency will have the lowest group velocity, and therefore will be the last to emerge. The computed time delays are of the order of a fraction of a second, i.e. of the same order as is observed for the isolated bursts. This idea also implies that the delay increases with the wavelength as observed. A special feature of this type of explanation is that it also may explain the second hump as an echo, but opinion is divided over this suggestion.

The angular emission diagram of the enhanced noise gives further food for thought. If the noise originated at or near the plasma level, where n is zero, then the radiation would emerge in a narrow beam around the normal direction. But this would also mean that during the days of meridian transit the apparent location of the spot on the disk would have to move more slowly than corresponds to solar rotation. This is not observed. The effect is not understood.

Bibliography

1. G. Reber, Ap. J. 100, 229, 1944.
2. D. F. Martyn, Nature 158, 308, 1946.
3. E. V. Appleton and J. S. Hey, Ibid., p. 339.
4. M. Ryle and D. Vonberg, Ibid., p. 339.
5. J. L. Pawsey, R. Payne-Scott, and L. L. McCready, Ibid., 157, 158, 1946.
6. L. L. McCready, J. L. Pawsey, and R. Payne-Scott, P. R. S. 190A, 357, 1947.
7. C. W. Allen, M. N. 107, 386, 1947.
8. J. P. Wild and L. L. McCready, Aust. Jour. Sci. Res. 3A, 387, 1950.
9. C. B. Burrows, Electronics, 22, no. 2, 75, 1949.
10. M. Ryle, Reports on Progress in Phys 13, 184, 1950 (Review paper).
11. M. Ryle and D. D. Vonberg, P. R. S. 193A, 98, 1948.
12. J. L. Pawsey, Proc. Inst. Elec. Eng. 97, 290, 1950 (Review paper).
13. E. V. Appleton and J. S. Hey, Phil. Mag. 37, 73, 1946.
14. J. S. Hey, S. J. Parsons, and J. W. Phillips, M. N. 108, 354, 1948.
15. R. Payne-Scott, A. J. S. R. 2, 214, 1949.
16. G. E. Hale, Ap. J. 73, 379, 1931.
17. M. Waldmeier, Zs. f. Ap. 16, 276, 1938.
18. Ellison, M. N. 109, 3, 1949.
19. H. Dodson, Ap. J. 110, 242, 382, 1949.
20. M. Waldmeier, Astr. Mitt. Eidgen. Stern. Zürich 153, 1948.
21. M. Waldmeier, Astr. Mitt. Eidgen. Stern. Zürich 162, 1949.
22. R. G. Giovanelli, M. N. 108, 163, 1948.
23. H. W. Newton and H. J. Barton, M. N. 97, 594, 1937.
24. H. W. Newton, M. N. Geophys Supp. 5, 200, 1948.

25. R. Payne-Scott, D. E. Yabsley, J. G. Bolton, Nature 160, 256, 1947.

26. A. B. Lovell, Nature 158, 517, 1946.

27. A. E. Covington, W. J. Medd, Jour. R. A. S. Can. 53, 106, 1949.

28. G. Reber, Cent Rad. Prop. Lab. Preprints 51-14, 1950. Now published in Science 113, 312, 1951.

30. R. N. Bracewell, Observatory 70, 185, 1950.

31. E. V. Appleton, J. Elec. Engrs. 71, 642, 1932.

32. M. N. Saha, B. K. Banerjea, U. C. Guha, Ind. J. Phys. 21, 199, 1947.

33. D. F. Martyn, P. R. S. 1934, 42, 1948.

34. M. Ryle, P. R. S. A 195, 82, 1948.

35. J. C. Jaeger and K. C. Westfold, AJSR 2, 322, 1949; 3, 376, 1950.

36. J. L. Pawsey and D. E. Yabsley, Ibid., 198, 1949.

37. M. Waldmeier and Müller, Zs. f. Astrophysik. 27, 55 (1950).

38. A. E. Covington, Proc. I.R.E. 37, 407, 1949.

Added later:

39. J.P.Wild, Aust.J.Sci.Res. 3A , 399, 1950

40. J.P.Wild, same journal 3A, 541, 1950

41. J.P.Wild, same journal 4A, 36, 1951

42. W.N.Christiansen, J.V.Hindman, A.G.Little, Ruby Payne-Scott, D.E.Yabsley and C.W.Allen, Aust.J.Sci.Res. 4A, 51, 1951

43. K.C.Westfold, J.of Atm., and Terr.Physics 1, 152, 1951

44. E.J.Blum and J.F.Denisse, Comptes Rendus 231, 1214, 1950 and 232, 387, 1951

45. E.J.Blum, J.F.Denisse and J.L.Steinberg, Comptes Rendus 232, 483, 1951

46. Y.Rocard, Comptes Rendus 232, 598, 1951

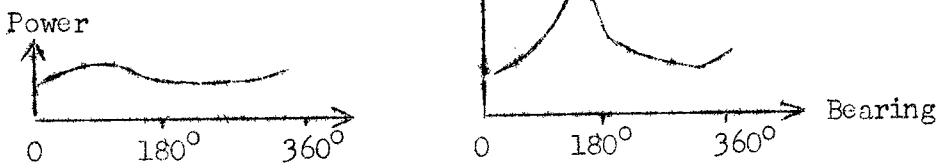
47. M.A.Ellison, Nature 167, 941, 1951

Chapter 8 GALACTIC AND EXTRA-GALACTIC RADIATION

8.1 Low Resolution Surveys (Henry J. Smith)

The Discovery and Earliest Observations. In 1931 Jansky carried out an investigation of terrestrial atmospheric noise on 14.7 meters wave-length. He found that when the atmospherics were "quiet," a thermal-like noise persisted. His antenna was fixed in altitude, and rotated through 360° in azimuth, with a period of 20 minutes. The background noise appeared to be highly directional. During each sweep his receiver recorded one maximum. At some times of the day the maximum was sharp-peaked, at others flat. When it was discovered that the bearing of the maximum moved with a sidereal period, and coincided with the direction of the galactic center, the interpretation of Jansky's observations followed readily.

well...



Reber, in 1943, used a 31 foot parabola with better resolution (12° half-power beam width). Reber's antenna had a fixed bearing (southwards) and swept through a vertical circle. His first observations were made at too short a wave-length, so that he received no signal at all. By moving to 1.85 meters (165 Mc/s) Reber was able to map the intensity of galactic radio-frequency emission. Reber's pioneer work

has been of great importance.

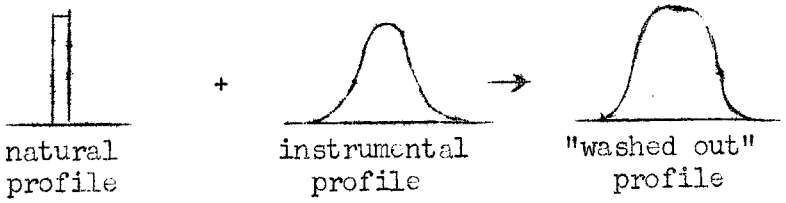
In 1946 Hey, Parsons and Phillips carried out a survey of the Milky Way at 4.7 meters (64 Mc/s). Their antenna consisted of an array of four Yagis

with a wire screen reflector. With this equipment they were able to secure an angular resolution of $14^{\circ} \times 13^{\circ}$. Like Jansky's, their antenna was fixed at 12° altitude, and swept through 360° azimuth. It was during this survey that the first radio point sources were discovered.

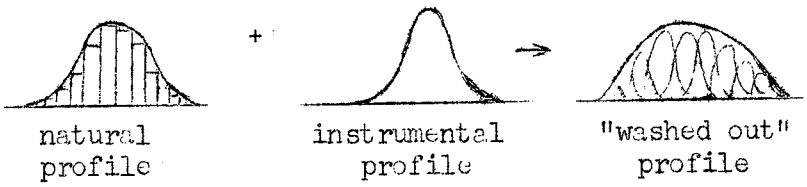
High versus Low Resolution Surveys. Up to 1946 it was thought that the galactic radio frequency emission was continuously distributed over the sky, finding its origin in the interstellar medium. There were theoretical troubles. In addition, the discovery of radio point sources initiated high resolution surveys (with interferometers or antennas of less than 3° half-power width) in the search for these objects. At the present time it is accepted as a working hypothesis that the noise emission from the Milky Way is chiefly due to the integration of distant radio stars. However, for the purpose of studying galactic structure low resolution surveys remain important.

The Observational Data and their Interpretation (Low Resolution). From the meter record of the radio telescopes we can secure in the usual way the brightness $B(\lambda, l, b)$ or temperature $T(\lambda, l, b)$ of galactic emission at a given wave length λ for each galactic longitude l and latitude b setting of the antenna. By using an antenna with resolution of $3^{\circ} - 4^{\circ}$ we can secure considerable directivity, yet average into the background all but the very brightest point sources.

However in order to reduce the measured brightness or temperature to the actual brightness or temperature of the sky at a given point, we must eliminate the effect of averaging. This is the familiar "blurred image" problem as occurs in spectroscopy. For simplicity we shall consider a one-dimensional problem. The low resolution of the receiver results in the "washing out" of a sharp spectral line:



When the natural profile is itself degraded, we may consider it as composed of successive rectangular profiles, and study the effect of washing out each separately.



If f(x) be the natural profile, g(x) the instrumental profile, then the broadened profile h(x) must be

$$h(x) = \int_{-\infty}^{\infty} f(x)\, g(x - y)\, dy.$$

The total intensity $\int_{-\infty}^{\infty} f(x)\, dx = \int_{-\infty}^{\infty} h(x)\, dx$ remains unchanged, but the washing out process will smooth out fluctuations in the true intensity distribution which are of the order of the instrumental profile or smaller. It is a relatively simple task to construct a "washed out" profile, given the instrumental and natural profiles. But to find the true from the observed profile is a task requiring discretion. Minute variations in h(x) will lead to strong fluctuations in f(x), which may or may not be real.

There are two general methods of solving for the observed profile. One is to make a Fourier transform of the integral equation above, yielding

$$F(\omega) \cdot G(\omega) = H(\omega).$$

Then $F(\omega_i) = \frac{H(\omega_i)}{G(\omega_i)}$ $(i = \pm 1, 2, 3, \ldots)$ for narrow frequency intervals. The intrinsic difficulty appears in this procedure in the form of dividing zero

by zero. The second method is iterative:

$$F = \frac{H}{G} = \frac{H}{1 + (G - 1)} \cong H\left[1 - (G - 1) + (G - 1)^2 \ldots\ldots\right]$$

First approximation: $F_1 = H$

Second approximation: $F_2 = H(2 - G)$

Third approximation: $F_3 = H(3 - 3G + G^2)$, etcetera

In other words one washes out the observed profile (first approximation), then applies the increment with the opposite sign to the first approximation to obtain the second approximation, etc.

The same procedure can be used to take into account side and back lobes of the antenna pattern. The magnitude of the changes produced by thus correcting for finite antenna resolving power can be seen from the following example from Bolton and Westfold. Their brightness temperature of the galactic center and anticenter, before and after correction, are as follows:

$$\text{Galactic Center}\begin{cases} T_{obs} = 2850^\circ \\ T_{cor} = 6200^\circ \end{cases} ; \quad \text{Anticenter}\begin{cases} T_{obs} = 710^\circ \\ T_{cor} = 650^\circ \end{cases}.$$

The changes are seen to be considerable. Since the resolving power of an antenna is less at longer wave-lengths, the apparent wave-length dependence of galactic concentration and temperature will be influenced by instrumental profile.

Results of Low Resolution Surveys. The distribution over the sky of galactic noise bears some resemblance to that of light. The belt of the Milky Way is clearly defined by a strong negative gradient towards the poles. The [illegible] width of the radio frequency Milky Way is about 30° in the direction of the center.* The galactic center is well-defined by the observations of Hey, Parsons

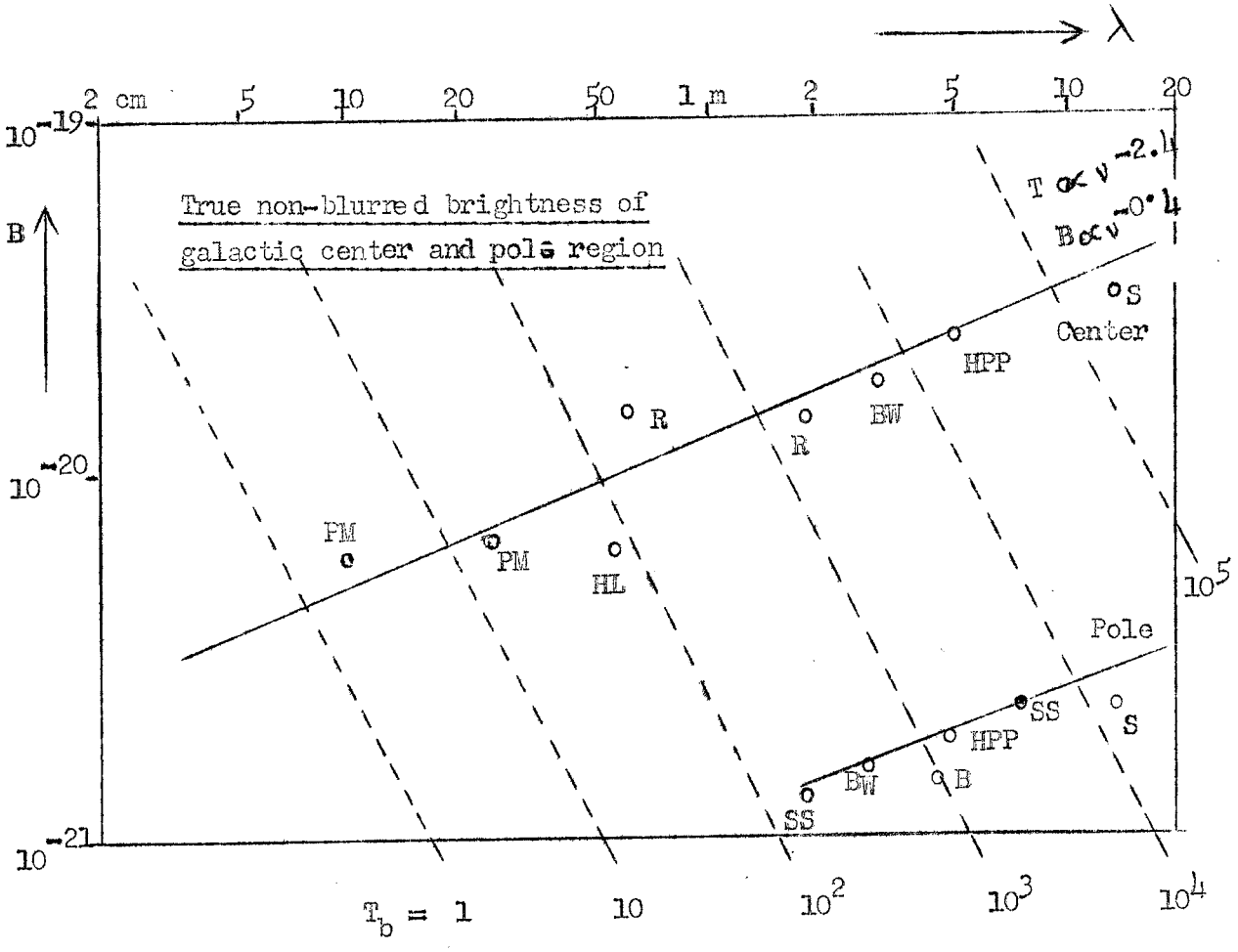
*Between half-power points at 100 Mc/sec; see further p. 165.

and Phillips, and Bolton and Westfold. Both groups of workers found a maximum intensity at $\ell = 325^\circ$, with a secondary maximum at about $\ell = 150^\circ$.

A second result from the low resolution surveys is the dependence of B and T_b on wave-length. From the definition of T_b,

$$B(\lambda) = \frac{2kT_b}{\lambda^2} .$$

In the following figure we have plotted B versus λ. Lines of constant temperature are straight lines - - - .



R	Reber	HL	Houtgast, Laffineur
HPP	Hey, Parsons, Phillips	PM	Piddington, Minnett (unpublished)
BW	Bolton, Westfold	B	Hanbury Brown (unpublished)
SS	Stanley, Slee	S	Shain (unpublished, provisional reductions)

From the mean of the best observations it is found

$$B \sim \nu^{-0.4}$$

$$T \sim \nu^{-2.4}$$

We shall postpone interpretation of these observations until we deal with problems of galactic structure.

References: Jansky, Proc. Inst. Rad. Eng. NY 20 1920, (1932)

Reber, Ibid., 36, 1215-8 (1948); Ap. J. 100 279, (1944)

Hey, Parsons and Phillips, Proc. Roy. Soc. A 192 425-45 (1948)

Bolton and Westfold, Aust. J. Sci. Res. A 3, 19-33 (1950)

Stanley and Slee, Australian J. Sci. Res. A 3, 234 (1950)

8.2 Observations of the Point Sources (R. B. Dunn)

Historical. During their studies of general galactic radio noise in 1946, Hey, Phillips and Parsons discovered that the noise intensity from the direction of Cygnus fluctuated some 15% in a matter of minutes. The presence of such rapid fluctuations suggested the existence of a source of small dimensions (a "point source") in that region. This was established when Bolton and Stanley in 1948 obtained a lobe pattern with their 3 meter sea-interferometer from this source, since an extended source could not give such a lobe pattern. They concluded from the observed pattern that the source has an angular diameter of less than 8'.

Bolton (1948) located several more point sources of relatively high intensity and small diameter. Later Ryle and Smith reported additional sources, including those in Cassiopeia and Ursa Major. To date 50 or more of these sources have been found, few of which have been correlated with visible objects. It has not yet been possible to observe any parallax for any of these sources; of

course, the limited resolving power available would make any but a very large parallax unobservable. Ryle (8) concludes that the sources must be at distances greater than 2×10^{16} cm., and hence presumably at stellar distances.

Ryle and Smith (1948) showed that the polarization of the radiation from the Cygnus source was random, and not circular, like the enhanced solar noise.

The fluctuations were at first thought to be intrinsic, suggesting that the sources were somewhat similar to the sun, which shows large intrinsic variations. The possibility of an ionospheric origin for the fluctuations had been recognized from the start, but this seemed unlikely when Ryle and Smith reported that the Cassiopeia source was remarkably steady, on the same days that the Cygnus source showed large variations.

Recent work with spaced receivers has shown, however, that all (or nearly all) of the fluctuations arise in the ionosphere, and the apparent disparity between the Cassiopeia and Cygnus results has been accounted for by diurnal variation.

Detection of point sources. The problem of detecting a point source is one of seeing a small, localized increment on top of the general background of galactic radiation. In general it is necessary to use interferometric techniques to provide a differentiation between point source and background. Since most of the sources are very weak, extremely sensitive and stable receiving equipment must be used. Systems have been developed which can detect sources with intensities of only 10^{-5} of the galactic background.

Two types of interferometer have been employed to study point sources. One uses the "Lloyd's mirror"technique with the sea as a reflector. This method was first used by McCready, Pawsey, and Payne-Scott (1) for solar radiation,

later by Bolton and Stanley (2) for point sources. The other method, first used by Ryle and Vonberg (3), is similar to the Michelson stellar interferometer and consists of two spaced antennas.

The advantages of the sea interferometer, as given by Stanley and Slee (4) are as follows:

i. The same sensitivity is achieved with a single antenna.

ii. The source first appears sharply in contrast to the general background at rising, whereas with the vertical system a slowly increasing amplitude is obtained.

iii. The sharp appearance at rising is useful in discriminating between a number of sources close together.

The advantage of the two-antenna type is that the observations are made at meridian transit, when the angle of incidence of the radiation on the ionosphere is a minimum. Refraction troubles are thus minimized.

An interferometer permits the detection of sources of quite low intensity, although complete coverage of all sources of a given intensity is limited by the overlapping of the lobe patterns from sources which are close to each other, making recognition of the individual patterns difficult.

Measurement of position.* The obtaining of precise positions is important in any attempt to correlate the point sources with visible objects. The methods of measuring position are in principle as follows:

1. High resolution telescope--of limited use so far, because of the large antenna needed.

2. Occultation by moon or planet--practical, but there are few occultations.

* (Added later) The numerous precautions and corrections needed in really accurate position measurements with a meridian interferometer are described in an instructive tabular form by Mills and Thomas (15).

3. Sea interferometer--several methods, discussed below.

4. Meridian interferometer--used by Ryle and Smith.

The principle is obvious in the first two cases. In both interferometer methods the declination may be obtained from the rate at which the source passes through the fringes, since daily motion gets slower towards the celestial poles. The detailed methods are as follows:

In the sea-reflection interferometer, the altitudes h_1, h_2, h_3, ... of the various lobe maxima can be computed from the height of the cliff and the wavelength. Then the altitude of a given object is related to the hour angle θ, and the declination δ, by:

$$\sin h = \cos \varphi \cos \delta \cos \theta - \sin \varphi \sin \delta$$

and $$\frac{dh}{d\theta} = \frac{\cos \varphi \cos \delta \sin \theta}{\cos h}$$

where φ is the latitude of the observing station.

Thus this type of observation admits of several ways of position finding:

1. Time of rising, or more precisely the time of passing through the first maximum, can be measured.
2. Time of setting, similarly.
3. Time to travel from h_1 to h_2, etc., gives declination, but not very accurately. The lobe separation was about 1°. Typical values of $dh/d\theta$ are:

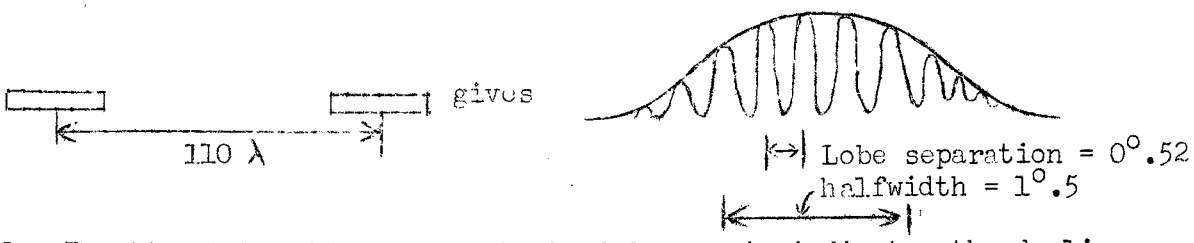
   $\varphi = -33^\circ$, $\delta = 0^\circ$, $h = 0^\circ$, $dh/d\theta = 0.84$, or 1° in 4.8 minutes.

   $\delta = 45^\circ$, $h = 0^\circ$, $dh/d\theta = 0.45$, or 1° in 8.9 minutes.
4. Direction finding in azimuth can be carried out. With the 17° beam used, this is possible to about 3°.

Two of these data are needed for a position. The combination used for the best determined positions was 1. and 2.; most other positions were determined by 1. and 3., with 4. as a check.

A good description of the methods used with the meridian interferometer is given in Ryle's review paper (8). The two methods used are:

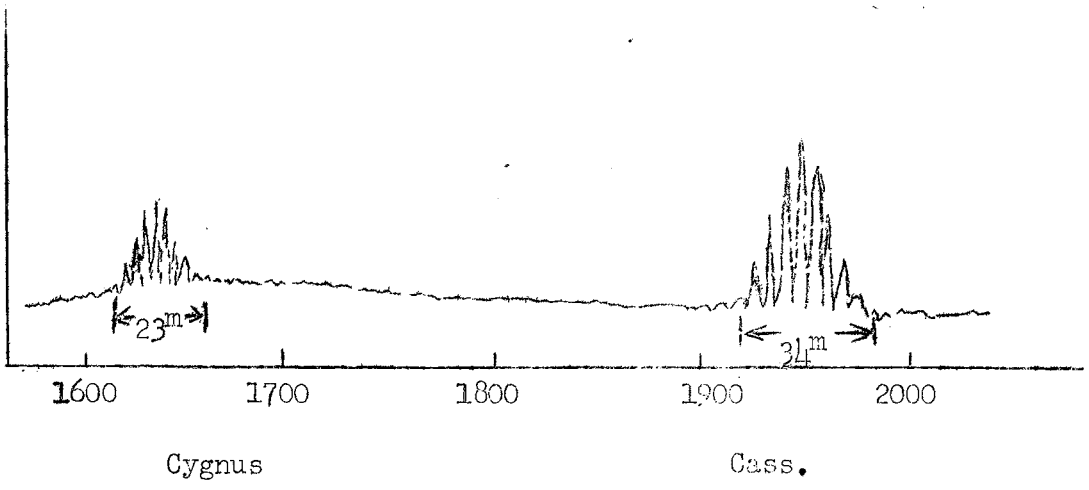
1. Time of meridian transit gives right ascension, the meridian transit being marked by the central lobe of the interference pattern. In order to know which is the central lobe, the entire pattern must be made rather narrow. This is done by making the individual antennas of the interferometric pair sufficiently large.



2. The time taken to go through the lobes again indicates the declination, for 1° is traveled in 4 minutes/cos δ .

Practising example: Find α and δ for the two sources in Fig. 14 of Ryle's review (8), which is reproduced below.

Fig. 1--Tracing of a transit of the Cygnus and Cassiopeia Sources about November 15.



Size Determination. As in the Michelson interferometer experiments a source whose angular aperture is very small compared to the angular separation of the maxima produces minima that are close to zero. By measuring the depth of the minima one can determine, or place a limit on, the size of the point source. In the case of the sea interferometer the angular diameter w of the source can be expressed by the equation

$$w = \frac{\lambda}{\pi h}\sqrt{3R},$$

where

λ = wavelength
h = height of the aerial above sea level
R = observed ratio of the received power (above the background level) at the maxima and minima of the interference pattern

At first glance it appears that an unlimited resolution can be obtained by merely increasing the height of the observing site. There are two limiting factors. One is the fact that the "effective sea reflection coefficient" is less than unity for altitudes below about 2°, arising from the curvature of the earth (4). The other is the finite bandwidth of the receiver which also makes the minima fill in. Though these factors can be computed (4), uncertainties remain, since the fringe pattern decreases in amplitude and becomes less easy to detect.

All sources were found to be small. The diameter of all is less than 30' and of the Cygnus source less than 1'.5.

Intensity and spectrum. Temperature definitions are not very relevant in this case, since the size of the sources is unknown. Instead the flux density, S, is used to denote intensity, with the unit 10^{-24} watt/ m^2 (c/s).

The table below gives a representative list of the properties of the principal point sources. The most complete published list is that of Stanley and Slee (4), but many more point sources are now known, whose positions have not yet been published. Approximate locations of a number of sources found by Ryle are however given in a small-scale chart in his review (8). *

Spectra have been obtained only for the most intense southern sources. Measurements have been made at seven frequencies from 18 to 1200 Mc/s (4,12). The accuracy of these determinations was not high, but was sufficient to show differences in the spectra of various sources. For three of the sources, the intensity was approximately inversely proportional to frequency, whereas Taurus A (the Crab nebula) showed an intensity that was constant over a large frequency range.

So far no intrinsic fluctuations of intensity have been established or suspected except for **

(a) the occasional "bursts" observed by Ryle, which may possibly have been solar or terrestrial in origin,

(b) the vanishing of the source in Centaurus after 10 good observations in 1947 (4). The possibility of equipment variations accounting for this are believed to have been ruled out. Other smaller and less certain sources are also believed to have disappeared.

The stronger point sources show a pronounced galactic concentration, whereas the weaker ones appear to have a fairly random distribution in direction. The space distribution of the point sources will be considered in section 8.3.

---

* Full list in reference (13).

** Ryle (unpublished) reports that over 1½ year he has not found any long-term variations in intensity greater than 10 per cent for most sources and 5 per cent for the strong sources.

Representative table of point sources:-- 100 Jy units

Name	Celestial coordinates α	Celestial coordinates δ	Size	Intensity at 100 Mc/s	Spectrum	Identification and remarks
Cassiopeia	$23^h\ 8^m$	+58° 0'	<6'	210(R)	$\sim\lambda$	
Cygnus A	19^h58^m	+40°36'	<2'	125(SS) 130(R)	$\sim\lambda^{1.2}$	
Ursa Major	12^h20^m	+58°	<6'	40(R)		To be discarded *
Near-center source	17^h44^m	-30°	<1°.5	30(P)	curved	
Taurus A	5^h31^m	+22° 1'	<6'	18(SS)	$\sim\lambda^0$	Crab nebula
Centaurus A	13^h22^m	-42°37'	<7'	18(SS)	$\sim\lambda^{1.2}$	Possibly NGC 5128
Virgo A	12^h28^m	+12°41'	<5'	12(SS)	$\sim\lambda^{1.5}$	Earlier called Coma A, possibly NGC 4486.
(Centaurus)	(unknown)			8		10 observations made in 1947, thereafter completely vanished.
No.10 on list of SS	-			3		perhaps no. 15 on complete list
No. 25 on list of SS No. 5 on list of R	-			1		perhaps no. 40 on complete list
No. 25 on list of R	-			0.3		perhaps no. 100 on complete list
M31, Andromeda Neb.	0^h40^m	+40°59'	~1°	1.3(H)		

SS Stanley and Slee
R Ryle
H Hanbury Brown
P Piddington (unpublished)

For comparison, it is interesting to look at various figures for the radiation from the sun and the whole galaxy at the same frequency in the same units, namely:

Quiet Sun	200
Average Sun (rough estimate during active years, see p.123)	6000
Total radiation integrated over sky, consisting of our galaxy and general background	34,000
General background alone (of presumed extra-galactic origin , see pages 162 and 174)	21,000

*This source turned out to be spurious, namely, the Cass source seen in the back lobe.

Hence our galaxy alone	13,000
Cygnus peak of Milky Way (without background)	300
Orion peak of Milky Way (without background)	1000

The total intensity of our entire galaxy is about 60 times that from the brightest "star," a ratio which is close to the corresponding value in the visual case.

Crab Nebula. The limits of the position of the radio source Taurus A, as measured by Bolton and Stanley ($5^h31^m.3 \pm 0^m.5$; $22^o2' \pm 8'$), enclose NGC 1952, the Crab Nebula, whose center is at $5^h31^m.5$, $21^o59'.5$ according to photographic observations. This nebula is the remains of the supernova of A.D. 1054 observed by the Chinese. The angular dimensions are 4' by 6', with an angular expansion of 0."23 per year. Measured Doppler shifts show an expansion of 1100 km/sec, implying a distance of 1000 parsec.

We shall now compare the observed intensity with that which might be expected on the basis of free-free transitions over the visible area of the nebula.

The measured 100 Mc/s flux density is 18×10^{-24} watt/m^2 (c/s). Dividing by $2h/\lambda^2$, i.e. 3.1×10^{-24}, we get that $T_a \cdot \Omega = 5.8$. The solid angle subtended by the amorphous mass of the nebula is 10^{-6} sterad., so that if this is responsible for the radiation the apparent temperature is 6×10^6 degrees K. This in itself would seem all right, because the absence of Balmer lines in the spectrum had already suggested that the temperature would be close to a million degrees. However, a comparison between the radio and optical flux densities shows a large disparity.

From the magnitude, the light emitted in all directions is about 0.7 x 10^{36} ergs/sec (this value might be off by a factor of 4 for $1^m.5$ interstellar absorption). This is for the interval λ 3000 to λ 5000. This interval is a frequency range of 1.0 x 10^{15} to 0.6 x 10^{15} sec^{-1}, so $\Delta\nu = 0.4 \times 10^{15}$. Consequently L_ν = 1.75 x 10^{21} erg/sec (c/s) = 1.75 x 10^{14} watts/(c/s). Now divide by the surface of a sphere with a radius of 1000 parsec, i.e. of area $4\pi \times (3 \times 10^{19} \text{ meter})^2 = 1.0 \times 10^{40}$ m^2. This gives $S(\nu) = 1.75 \times 10^{-26}$ watts/m^2(c/s) in the optical case compared to 18 x 10^{-24} in the radio case. Of the factor of 1000 difference a factor of 10 is taken up by the factor $\ln(p_a/p_b)$ which is 18 for radio and 1.8 for photographic light in this case. There still remains a discrepancy of 100. Evidently there is a radio-emission effect of unknown type that is about 100 times as effective as the free-free emission that is supposed to cause the photographed light.

Two other remains of supernovae (Tycho's and Kepler's) have been observed for radio sources with no result.

NGC 5128: This nebula was formerly believed to be extragalactic. It consists of an elliptical nebula crossed by a dark dust band. Evans (7) suggests that the nebula is not extragalactic, but is a gas cloud illuminated by a point at its center.

Ionospheric Twinkling. The first belief that the fluctuations of the point sources were intrinsic was supported by Ryle's observation that the Cassiopeia source was steady on days when the Cygnus source was showing large fluctuations. However it later appeared that the Cassiopeia source also showed fluctuations at times, and this redirected attention to the possibility that the fluctuations

might arise in the ionosphere, in a manner analogous to the production in the lower atmosphere of the "seeing" fluctuations in visual astronomy.

The crucial experiment, then, was to obtain simultaneous observations on spaced receivers. Such an experiment was carried out by the Cambridge and Manchester groups (11), at 3.7 and 6.7 meters. They found that with sufficiently great separation of the antennas the fluctuations were quite uncorrelated, which indicated that they are local in origin, and hence presumably ionospheric. Their results are summarized in the following table:

	6.7 m.		3.7 m.
spacing	< 20 km.	some correlation	100 m. complete identity
	20 km.	no correlation	4 km. some difference
	up to 160 km.		210 km. no correlation

There was however a correlation between the degree of disturbance at Manchester and Cambridge. A later paper by Ryle and Hewish (9) describes subsequent systematic observations at 3.7 m. on four sources, which have led to a clearer understanding of the fluctuations.

In their studies and in analyzing their records a "fluctuation" index was used. The monthly mean of these indices for all of the sources was plotted and the same diurnal variation was found in each case, with a maximum at about 1 a.m. This diurnal effect accounts for the earlier disparity between the Cygnus and Cassiopeia results, since in those observations the sources culminated at different times of day. Ryle and Hewish also note angular displacements as large as 3 min. of arc at 6.7 m. and 1-2 min. of arc at 3.7 m. The variations have a duration of about 30 sec., and can cause difficulties in accurately locating galactic sources. Similar fluctuations have been noted in solar noise and in moon echoes. Apparently there are "blobs" of ionized gas in the earth's upper atmosphere. Suggested origins of these "blobs"

have been meteoric impact (spread E layer echoes--studied by Appleton and Naismith (10)), impact of corpuscular radiation from sun during magnetic storms and interstellar particles that are falling into the sun but hit the earth (9).

Booker (unpublished) has recently found a correlation between these fluctuations and the occurrence of "spread F" echoes from the ionosphere. This suggests that the origin of the twinkling is in the F region, 250 or so km. above the earth.

It thus seems to be well established that the fluctuations are terrestrial. There are still some puzzling features, however, such as the report by Stanley and Slee (4), that the period t decreases with decreasing frequency. This may be a chance result, since the fluctuations in moon echoes at 20 Mc/s reported by Kerr and Shain (12) had a period about 5 times as long as those of Stanley and Slee at 100 Mc/s.

Further extensive reports about the fluctuations of the Cygnus source, that were not available when the above was written, are found in references (14) and (15). They describe observations on 205 Mc/sec and 97 Mc/sec, respectively.

BIBLIOGRAPHY

(1) McCready, Pawsey, and Payne-Scott Proc. Roy. Soc. A190: 357 1947.

(2) Bolton and Stanley Nature 161: 312 1947.

(3) Ryle and Vonberg Proc. Roy. Soc. 193: 98 1948.

(4) Stanley and Slee Aust J of Scientific Research. Series A 3 p234-50 1950.

(5) Baade ApJ, 96: 188 1942.

(6) Minkowski ApJ 96: 199 1942.

(7) Evans M.N. 109: 94 1949.

(8) Ryle Reports on the Progress of Physics 8: 238 1940.

(9) Ryle and Hewish M.N. 110: 381 1950.

(10) Appleton and Naismith Proc. of Phys. Soc. 59: 461 1947.

(11) Lovell; Smith Nature 165: 422, 1950.

(12) Kerr and Shain, Proc. I. R. E., 39, 230, 1951.

Added later:

(13) M.Ryle, F.G.Smith and B.Elsmore, to appear in M.N. 111, , 1951.

(14) Ch.L.Seeger, J. of Geophys.Research , , June 1951.

(15) B.Y.Mills and A.B.Thomas, to be published soon.

8.3 Galactic structure (F. Q. Orrall)

Origin of galactic radio emission. Several interpretations have been suggested at various times for the origin of galactic radiation. It was easily seen that stars of the most common types, with temperatures of about 10^4 degrees K. and a dilution factor of 10^{-14}, would give an average temperature over the sky, if radiating thermally, of only 10^{-10} degrees K. At the galactic center, with the stars assumed to be 100 times more numerous there, the temperature would be only 10^{-8} degree K., which is still far too low. Even if we assume that the average brightness is enhanced to 10^8 degrees K., as occurs on the sun, the integrated brightness temperature at the galactic center would still only be 10^{-4} degree K., leaving a discrepancy of 10^7, when compared with the observed value of 10^{+3} degrees K. Such considerations led investigators to look to interstellar matter for the source of the radiation.

Whipple + Greenstein An early suggestion, which was discussed seriously for the galactic center, invoked the interstellar dust. However, the real dust temperature of perhaps 100° K. is the absolute upper limit, and this is far too low!

An origin in the interstellar gas has long held the main position, because it gave about the right intensity in the meter range, and the correct kind of wavelength distribution. However self-absorption prevents the T_b from exceeding the actual temperature, and too high values of T_b have been observed at the longer wavelengths. It is now clear that the interstellar gas cannot alone account for the observed intensities. Thus this explanation is also excluded as the main cause.

Point-sources are now considered to be the principal explanation; this is only a half-explanation, since we still have to explain what the point sources

are. It is found that the numbers and brightness of the known point sources can be made to agree very well with the conception that the general galactic brightness is due to the total effect of all unresolved point sources.

We shall therefore accept this concept as a working hypothesis, and proceed to discuss whether it is possible to find from galactic-structure considerations what known objects these point sources resemble.

Plane of Symmetry of the Galaxy. A point of the galactic circle may be determined by finding at one longitude the latitude of maximum radiation. This works well, except in some directions close to the anticenter, where the gradient with latitude is weak. Northcott and Williamson applied this method to published surveys. A careful direct investigation was made by Seeger and Williamson at 205 Mc/s (Ap.J. 113, 21, 1951). Over 300 determinations of single points are combined into 16 normal points, with latitudes between + 1°.9 and - 2°.2 . The authors remark that several normal points are drawn off a smooth curve by the point sources Cyg A, Cas A, and Tau A. Possibly further deviations are due to groups of yet unknown point sources. The entire curve might thus be determined by relatively nearby objects, just like the bright stars outline "Goulds Belt".

Seeger and Williamson determine by least squares a small circle fitting these points; it has the mean latitude of - 1°.3 . The weakest point in this solution is the gap of the unobserved southern Milky Way. The computed small circle reaches latitude - 3°.8 in this gap, but according to Bolton and Westfold's 100 Mc/sec survey the median latitudes are positive in this region. The small-circle solution must therefore be regarded as spurious. Since the sun's distance above the galactic plane, as found by optical means, seems to be only 10 or 20 parsec, the expectation for a fully transparent galaxy is that deviations from a great circle must be extremely small.

Center of the Galaxy. The best determination of the longitude and latitude of the galactic center is that of Bolton and Westfold, since in their observations they had the center high in the sky. Their position, when corrected for precession, to the system based on the "international pole" of the galaxy, is $\ell = 325^{\circ}.4$, $b = -0^{\circ}.9$. The most recent optical determination, from all reliable data up to 1940, by van Tulder (B.A.N. No. 353, 1942), gives the same longitude and a latitude of $-1^{\circ}.0$. The agreement is excellent.

Intensity Distribution and Stellar Populations. Westerhout and Oort (B.A.N. No. 426,1951) have made a systematic study of the intensity to be expected on the supposition that the general intensity of the galaxy comes from star-like objects. Intensity distributions have been obtained from a model of the galactic system derived from observations of visible stars, and compared with the radio-frequency distribution observed by Bolton and Westfold.

Three types of stellar distribution were considered beginning with that of Baade's Population I, of which typical objects are O and B stars and the interstellar matter. These objects have a large galactic concentration. The average z-coordinate of O and B stars is 30 parsec, which corresponds at the distance of the center to a halfwidth in latitude of about $0^{\circ}.3$, much smaller than the observed 100 Mc/s value of 15°. Also these stars are not likely to be found in the galactic nucleus.

The second distribution used was that of the common K and G dwarf stars. The distribution of these cannot be studied directly, as they can only be seen at short distances. However they are by far the most abundant objects in the galaxy, and can reasonably be regarded as having the same distribution as the mass of the galaxy. For the space distribution of mass, a model derived by

Oort (B.A.N. No. 338) from dynamical data was used. This contains a number of spheroidal equidensity surfaces extending beyond the sun, together with a superimposed central mass almost extending to the sun, combined with a concentration of about 10% of this mass very near the center. This large central mass was assumed to have a spheroidal shape.

It was found that the radio distribution, both in longitude along the galactic circle, and in latitude at the longitude of the center, was represented satisfactorily by this model, except for a constant residual of about 600° all over the sky. The axial ratio of the big central mass cannot be estimated very well from dynamical data alone, so that this is a weakness in the comparison. Westerhout and Oort take the view that, since the agreement is good elsewhere, the radio data may be used to improve our knowledge for this central mass. The ratio between the axes as obtained in this way is about $4\frac{1}{2}$.

The third type of distribution tried was that of Population II, including globular clusters and RR Lyrae variables. This distribution shows too small a galactic concentration, and may be discarded.

Thus it may be concluded that the bulk of the radio stars, if not all of them, are distributed in about the same way as the common stars in the galaxy. None of the objects considered ban account, however, for the relatively high "residual" intensity found away from the galactic plane and in the region of the anticenter. A galactic origin for this radiation would require the existence of a further system of sources, in the form of a large spherical atmosphere around the galactic system, which is highly improbable. It appears most likely that the residual can be attributed to radiation from large numbers of unresolved external galaxies. This interpretation has been strengthened by the recent detection of radiation from a number of such galaxies individually.

Contribution of the Interstellar Gas. We shall now consider the contribution of the interstellar gas since it alone seems capable of competing with the point sources in determining the general galactic radiation.

We do not have to consider emission and absorption separately. Emission will first give measurable values. At 100 Mc/s, for which these computations are made, the amount becomes measurable if the T_b of the interstellar emission gets higher than 50° or 100°.

Only when the density becomes much higher, and the computed emission T_b approaches the actual temperature of 4000° - 10000° for the H-II-regions, absorption becomes important. The H I regions are certainly unimportant because of the lower electron density. This absorption would act on the point sources as well as on the emission of the gas itself.

We shall estimate the emission of the interstellar gas, assuming that there is no self-absorption. We shall use the free-free formula developed earlier for the corona and quiet sun. The emitted energy per cm^3 per unit solid angle per unit band width is

$$\frac{E(\nu)}{4\pi} = 3.0 \times 10^{-39} \frac{N_i N_e Z^2}{T^{\frac{1}{2}}} \ln \frac{p_a}{p_b} \cdot n.$$

For the quasi-constants we assume:

$N_i N_e Z^2 = \beta N_e^2$, with $\beta = 1.2$,

$T = 10{,}000^\circ$, $n = 1$,

$p_b = 1.5 \times 10^{-7}$ cm and $p_a = 2.2 \times 10^{-2}$ cm, giving $\ln(p_a/p_b) = 11.8$.

We thus find:

$$\frac{E(\nu)}{4\pi} = 4.2 \times 10^{-40} N_e^2$$

In order to obtain the brightness, this must be multiplied by the path length ℓ :

$$B(\nu)_{cgs} = 4.2 \times 10^{-40} N_e{}^2_{(cm^{-3})} \ell \text{ cm}$$

or in M.K.S. units,

$$B(\nu)_{MKS} = 1.30 \times 10^{-24} N_e{}^2_{(cm^{-3})} \ell \text{ (parsec)}$$

Strömgren has called the quantity $N_e{}^2_{(cm^{-3})} \ell$ (ps) the "emission measure." Its more accurate definition is $\int N_e{}^2 ds$. The value of this integral can now be obtained more precisely from Strömgren's measurements of the Hα and Hβ emission in the galaxy than from a crude assumption about the electron density and the effective path length. Strömgren's measured values for the "emission measure" are, for example, 8×10^6 for the center of the Orion nebula, 6×10^3 for some diffuse nebulae, and 400-2000 for the Struve-Elvey regions. The use of these measurements has one objection, namely that the interstellar extinction (by dust clouds) may have cut out an appreciable part of the visual radiation. The emission measure in any direction may therefore be higher than the values given by Strömgren. The very strong galactic concentration further reduces the chance of measuring gaseous emission in low resolution surveys. Westerhout and Oort have estimated for a typical direction that the brightness temperature due to gaseous emission would be 715° in the galactic plane but only 100° when averaged over Bolton and Westfold's antenna beam. This is only a fraction of the total observed temperature of 900° in this direction. Such considerations indicate that the interstellar gas probably contributes measurably to the intensity observed in the meter range. It is certain however that it cannot account for the whole of the intensity, since brightness temperatures at the longer wavelengths exceed the actual temperature, and also the space distribution is wrong.

why didn't vdH consider non-thermal mech's?
— he doesn't cite Kiep. 50 a + Alfvén + Hulst. 50 or p. 170, in connection w/ RS's

Westerhout and Oort decided to neglect the gas component in their consideration of the general intensity distribution.

Extended emission regions at longitudes away from the galactic center are probably important. The anticenter peak found in the radio surveys may be due to the greater Orion nebula (i.e. the outer loop of the nebula). Also the peak in Cygnus at $l = 50°$, which Bolton and Westfold suggest is evidence of a spiral arm, might well be a "gaseous emission arm," rather than an arm of "point sources."

Width of the Milky Way. As was expected from the better resolving power, the Milky Way seems to be narrower at higher frequencies. But Reber (Sky and Telescope 8, April, 1949) and Williamson (J.Roy.Astr.Soc.Canada 44, 12, 1950) have claimed that this effect cannot be dismissed as being wholly due to the resolving power. Williamson tabulates the widths measured across the Milky Way at longitude 0° between the half-power points at southern and northern latitude. After a rough correction for resolving power they range from about 40° near 60 Mc/s to 8° at 480 Mc/sec. This is also illustrated by the figures in Bolton and Westfold's paper (AJSR 3, 251, 1950). If this result is substantiated by further investigations, two explanations (or a combination of both) may be thought of:

(a) Absorption of the waves in and near the galactic plane would reduce the brightness at the equator relative to the higher latitudes and thus cause an apparent widening. This explanation seems unlikely because also the brightness distribution with galactic longitude would be strongly affected by the absorption. In addition, no physical reason for such absorption is known; the free-free absorption is too weak.

(b) The sources away from the galactic plane may have spectra that differ systematically from those in and near the galactic plane. This point is fully unexplored.

8.4 Interpretation of the point sources (F. Q. Orrall, R. B. Dunn)

Distance and luminosity.* We have seen that headway may be made with the problems of galactic structure if we attribute the general galactic radiation to unresolved point sources. We must now see whether any answer can be given to the question "what are the point sources?" The distribution of galactic radiation suggests some common type of star as the origin, yet the fact that 7 out of 17 of Stanley and Slee's list of point sources are within latitudes of $\pm 5^{\circ}$ leads us to believe that these objects at least are at distances of the order of 1000 parsec. It is indeed remarkable that the two strongest sources (Cassiopeia and Cygnus) are very close to the galactic equator. We might consider that there are two types of point sources, "common" and "giant" types.

Although the total number of known sources is small, it is worth trying a statistical treatment, however rough. We follow here the work of Westerhout and Oort.

First, let us assume that all radio stars have the same brightness temperature T_s and disk area $A(p\ c)^2$. Then T_sA will be proportional to the absolute luminosity. Let there be $n(r)$ stars per $(p\ c)^3$, where r is the distance, and let n be proportional to the mass distribution in the galaxy. Put $n = n_o$ in the vicinity of the sun. Then for the brightness temperature observed in a galactic survey, we have:

$$T_b = \int_o^\infty nAT_s dr.$$

Now Oort's mass model gives $\frac{n}{n_o}$ for any point in the galaxy, so that the integral $M = \int_o^\infty (n/n_o)dr$ is known in all directions. We can then rewrite the above as:

$$T_b = n_o A T_s M.$$

*A quite different approach has been suggested by Shklovsky (Doklady Ak. Nauk 73, 479, 1950). If some bursts of radio stars are real, then time delays between different frequencies due to different group velocities in interstellar space may give estimates of distance.

Putting in values of T_b observed and N computed we find as the constant of proportionality in any direction of the sky:

$$n_o AT_s = 0.08\ ^oK\ (pc)^{-1}.$$

The volume of a sphere enclosing N sources is

$$\frac{N}{n} = \frac{4}{3}\pi r_N^3$$

so that the probable distance of the Nth brightest source is

$$r_N = \left(\frac{3N}{4\pi n}\right)^{1/3}$$

and it will subtend a solid angle $\Omega = \frac{A}{r_N^2}$ whereupon

$$\Omega T_s = AT_s n^{2/3}\left(\frac{4\pi}{3N}\right)^{2/3}.$$

From the list of known sources, we can estimate with reasonable accuracy the flux density S for the 10th brightest source, and hence find ΩT_s since $\Omega T_s = \frac{\lambda^2}{2k} S$. We can thus find the numerical value of $n^{2/3} AT_s$, which yields the second equation for n_o and AT_s:

$$n^{2/3} AT_s = 1.14\ ^oK.$$

Solving the two equations Westerhout and Oort find the following values:

$$n_o = 0.00032\ (\text{parsec})^{-3}, \text{ or } n^{-1/3} = 15 \text{ parsec}$$

$$AT_s = 260\ ^oK\ (\text{parsec})^2$$

distance of tenth source,

$$r_{10} = 20 \text{ parsec}.$$

These may be compared with the results obtained by Bolton and Westfold (unpublished), using a similar method, but a less detailed model of the galaxy. Their values were: $n_o = 0.00020$, $AT_s = 1000$.

The two sets of results agree quite well, and are in keeping with the assumption that radio stars produce the background.

Westerhout and Oort went on to consider the case in which there is a Gaussian dispersion of absolute magnitudes, to take account of the evidence that the brightest point sources are probably not the closest. They used a value of $\sigma = 1$ for the dispersion in $\log_{10} AT_s$, since a value of σ exceeding 1.5 would give too great a galactic concentration of the brightest sources. The resulting values are (for the tenth brightest source)

r_{10}, logarithmically averaged over all possibilities, = 39 parsec,

r_{10}, linearly averaged over all possibilities, = 75 parsec,

AT_s corresponding to the "average source" (I_o) = 5^o K $(\text{parsec})^2$

With a dispersion of unity, one sixth of all the sources would be of luminosity $< (1/10) I_o$, and one sixth $> 10\ I_o$.

The analogy with visual stars is of interest. There the situation is very similar. In any volume element by far the greatest contribution is from the dwarf stars, but among the brightest stars we find only the giants.

Physical discussion of the point sources. The most important result from the standpoint of the physical nature of the point sources is not their distance but their "absolute magnitude," AT_s. In the uniform-source analysis the value $AT_s = 260^o$ K $(\text{ps})^2$ was found. With dispersion introduced the same value will still be correct for quite a few giant sources. From this value we find the following estimates for T_s:

if radius = sun's radius, then $A = 32 \times 10^{-16}$ ps^2 and $T_s = 8 \times 10^{16}$ °K,

if radius = 800 x sun's radius, like largest known stars, $T_s = 1.2 \times 10^{11}$ °K.

An even more direct estimate may be made for one of the known bright sources, e.g., the Cygnus source, which has $\Omega T = 42$ (at 100 Mc/sec). Assuming that the angular diameter is 0".036 like Antares we have $\Omega = 2.4 \times 10^{-14}$, which gives

$T_s = 2 \times 10^{15}$ degrees K. And assuming it to be a giant star of 800 times the sun's radius at the distance 1000 ps, which seems probable for this source, then $T_s = 2 \times 10^{16}$ degrees K.

The suggestion has been made that the point sources are stars with a continuous activity of unusually strong flares. This could just be reconciled with the strongest known outbursts on the sun, which have given $T_a = 10^{13}$ (referred to the entire sun's disk) so probably $T_b = 10^{15}$ to 10^{16} (average brightness temperature over the emitting area). A serious objection is that it is hard to see how a star could spend the energy to have so strong flares all over its surface all the time.

It is interesting to compare the energy output in the radio domain with that in the optical (ultraviolet, visual and infrared) domain that we may reasonably expect. The solar constant is 1300 watt/m^2. Scaling down by 36.6 magnitudes, which is 4 x 36.6 = 146 decibels (a factor $10^{14.6}$) we obtain for the optical output of a 10th magnitude star the flux density

$$\int S \, d\nu = 3 \times 10^{-12} \text{ watt/m}^2.$$

Similarly, by direct integration of the measured flux density of the Cygnus source over all frequencies from 0 to 300 Mc/sec (i.e. $\lambda > 1$ meter) we find the output in the long-wave radio domain

$$\int S \, d\nu = 4 \times 10^{-14} \text{ watt/m}^2, \text{ or somewhat higher.}$$

The comparison may be made more clear by the following little table:

Flux density received at the earth in units of 10^{-12} watt/m^2

	integrated over vis + uv + ir	integrated over $\lambda > 1$ meter
Sun	1.3×10^{15}	quiet: 0.2 average: 100
Tenth magnitude star of sun's type	3	average: 2×10^{-13}
Tenth magnitude star or source of any type	3	(?????)
The Cygnus source	(?????)	0.04

We thus are forced to conclude that either the Cygnus source is a so peculiar object that the greatest amount of its energy output is spent in the radio domain or that its radio output is a small fraction, say 1 per cent, of its optical output but that means that we must be able to find it with an ordinary telescope as a tenth magnitude star. The first alternative now seems more likely.

Suggested explanations of point sources.--In view of the highly speculative nature of these explanations we shall not more than mention them.

(1) They may be stars, very cool or very hot or very old or very young, but at any rate very large; ~~if they are~~ they would have to have an extremely high radio brightness temperature and a radio output comparable to or larger than the visual.

(2) They may be regions of interstellar matter with special properties. The special mechanisms that have been suggested are roughly these.

A. The radiation might be emitted by cosmic-ray electrons when moving in spirals in a magnetic field, that might be either stellar or interstellar (this mechanism was proposed independently by Alfvén and Herlofson and by Kiepenheuer).

B. The currents associated with the smallest size turbulent elements in interstellar gas clouds might emit this radiation (an idea proposed by Teller).

C. Something might happen in the wake of a star moving through a gas cloud, in the general vein of Hoyle and Lyttleton's accretion theory (an idea proposed by I. King).

Several authors have played with the possibility of an intimate connection between cosmic rays and radio stars. This is suggested by the high temperatures that apparently are reached in radio stars. (Note that a temperature

7/9/07 - Ivan told me that he left in on this course as a Junior Fellow at Harvard

of 10^{14} degrees K. = 10^{10} eV, which is a typical cosmic ray energy.) Also the intensities have a remarkable proportionality. According to Unsöld, the ratio of the short-period radio emission by a solar flare to the short-period cosmic ray increase associated with a flare is equal to the ratio of the ever-present level of galactic radio noise and the ever-present intensity of cosmic rays.

8.5 The hydrogen emission line at 21 cm (H. C. van de Hulst, Discovery.--On a colloquium in 1944 Van de Hulst suggested a search for the interstellar hydrogen emission line at 21 cm wave length.[1] His prediction was backed up by a similar analysis made by Sklovsky.[2] Both authors assumed that the interstellar gas would be optically thin in this line (which now seems incorrect) and Shklovsky computed the transition probability (which now seems off by a factor 4). Shklovsky also made clear that no other atomic lines, but perhaps some molecular lines, might be capable of detection.

This line was first discovered in the sky by Ewen at Harvard University on March 25, 1951. Ewen employed a horn antenna having a beam width of 12°, so no great angular resolution could be obtained. He proceeded, however, to measure accurate line shapes and frequency shifts during the times that the line could be observed. Calibration of his equipment gave the antenna temperature $T_A = 35^{\circ}$ at the center of the line at the time when the fixed antenna was in the direction of Ophiuchus, close to the galactic center. The discovery was confirmed six weeks later, when Muller observed the line at Kootwijk, Holland. He employed a parabola with a diameter of seven meter, which gives a beam width of about 3 degrees. Sweeping the antenna across the galactic equator at various declinations he obtained a first outline of the distribution of

this line emission over the sky. In the center and near-center regions the intensity fell to half the maximum intensity at 8° to 10° galactic latitude. This shows that Ewen's antenna pattern was almost completely filled; the value of $T = 35^{\circ}$ may therefore be interpreted as the brightness temperature at the center of the line in the direction of Ophiuchus-Sagittarius. This does not include the T_b of the continuous galactic radiation, which at this wave length is about 15° in the direction of the galactic center (see section 8.1).

Wave length and transition probability.--A neutral hydrogen atom has a $^2S_{\frac{1}{2}}$ ground state. The electron spin can assume two positions in the very weak field of the nuclear magnetic moment. This causes a splitting of the ground state into two hyperfinestructure levels, the upper one of which is degenerate of 3 levels that can be split by Zeeman effect. The Zeeman effect in interstellar space is negligible so that only one frequency is emitted. Laboratory measurements of the frequency by magnetic resonance have been made by Rabi and others since 1946. The most recent value[3] is 1420.405 Mc/s. The emission is possible by magnetic dipole radiation. The line strength in the sense defined by Condon and Shortley is 1 when referred to one sublevel and 3 when referred to the entire upper level; the matrix element of the magnetic moment is simply the Bohr magneton. This gives a transition probability for spontaneous emission, $A = 2.84 \times 10^{-15}$ sec^{-1}. In the absence of other processes this means that the average life time of an atom in the upper level is $1/A = 3.5 \times 10^{14}$ sec = 11 million years.

First results, astronomical importance. As of June 11,* hardly any results have been obtained except those already described. The Dutch measurements gave a brightness in Cygnus stronger and more closely concentrated towards the galactic

*Early July also Christiansen and Hindman in Australia detected the line.

plane than in the center region. The wide extent in latitude in the Ophiuchus region can be explained only by assuming that nearby clouds, rather than distant regions of the galaxy are responsible. This is also made plausible by Purcell and Ewen's computation of the opacity depth based on assumed density and temperature. The temperature needed in this computation refers to the ratio of the population of the upper and lower hyperfinestructure level. This T is of direct importance because the ratio of the "effective absorption" to the "pure absoprtion" (see section 3.1) depends on it. A consequence is that T_b can never exceed this temperature. The T_b of 35° measured by Ewen comes close to the value of 50° that may be theoretically estimated for the temperature characterizing the population ratio. This again makes a high optical depth plausible.

The greatest astronomical importance of these investigations will be in the measurements of line widths and line shifts. The shifts due to galactic rotation may go up to 0.5 or 1 Mc/sec for distant regions of the galaxy. It is quite likely that so strongly displaced lines will not be affected by absorption in the nearby clouds.* At present neither the Harvard nor Kootwijk instruments are equipped to make accurate comparison of spectral intensities at frequencies more than a few 100 kc apart. Further observations will undoubtedly tell us a great deal more of galactic structure. Both the fact that here neutral hydrogen is observed and the fact that we can measure velocities of parts of the galaxy near the center are complete novelties in galactic research.

* This prediction has since been confirmed.

1. H. C. van de Hulst, Ned. Tijdschrift voor Natuurkunde 11, 201, 1945.

2. I. S. Shklovsky, Astronomicheskii Zjournal 26, 10, 1949.

3. P. Kusch and A. G. Prodell, Physical Review 79, 1009, 1950.

→ Letters to Nature will appear summer 1951.

8.6 Extragalactic radio waves

The analysis by Westerhout and Oort of the intensity distribution of galactic noise in relation to the mass distribution in the galaxy indicated clearly the existence of a residual brightness temperature of 600° (at 100 Mc/s), with a spherical distribution. This residual most probably comes from unresolved external galaxies, since theoretical considerations about the interstellar gas indicate that absorption will be small inside our own galaxy. An early estimate by Van de Hulst (1) has shown that it is possible that the combined effect of all external galaxies gives a measurable brightness in radio-frequencies while in photographic light the combined effect is unobservable. The difference is that the red shift brings into the photographic region the ultraviolet parts of the spectrum, that are intrinsically weak. But the same red shift brings into the radio frequency region the parts of the spectrum at higher radio frequencies that are quite as bright. These preliminary computations were based on the models of the universe studied by Hubble and Tolman (2).

The first external galaxy detected was M31, the Andromeda nebula, which was observed by Hanbury Brown and Hazard (3) at 158 Mc/s, using the 218 ft. antenna at Manchester (beamwidth about 2°). Their measured position was $\delta = 40^{\circ}55' \pm 20'$, $\alpha = 0^{h}40^{m} \pm 2^{m}$, compared with the visual position of the center of $40^{\circ}59'$, $0^{h}40^{m}$. The dimensions of the source were found to be $25' \pm 10'$ in declination and $45' \pm 10'$ in right ascension, which are of the correct size and orientation for the source to be identified with M31. The observed intensity was 4×10^{-25} watt / m^2 (c/s), leading to a brightness temperature of about 1000° K. Oort and Westerhout computed that our galaxy, when placed at

the distance of the Andromeda nebula (210,000 parsec) = ⅓ median distance would give a flux density of 7×10^{-24} at 100 Mc/s, or probably 2×10^{-24} at 158 Mc/s. This is five times the observed value, which shows that the Andromeda nebula and our galaxy are sources of radio waves of about the same strength.

More recently, the radiation from three more external galaxies has been detected at Cambridge, England (4). These are M33, M51, and M101. The last two of these are believed to be three or four times as far away as are M31 and M33. Ryle (unpublished) states that: "The angular diameters of the nearest ones are too large to [measure the field flux density] produce the correct amplitude on our interferometer system, but the recorded powers (81.5 Mc/s) are: 4.0, 8.0, 7.5 and 3.5 watt/m^2(c/s) for M31, M33, M101, and M51 respectively. By making reasonable assumptions about the probable distribution across the Andromeda nebula we would suggest that the total flux is probably about 25×10^{-25} watt/m^2 (c/s), a figure somewhat higher than the Manchester one."

The possibility that most known point sources might be external galaxies has been suggested at various times. In possible support of this is the fact that the "common" point sources (although not the "giant" point sources) appear to have an approximately spherical distribution, which could imply that the objects concerned are either very near or very distant. However, apart from the four galaxies mentioned above, there is no evidence that any of the 300 nearest galaxies can be related to radio stars (Ryle, c.s. 5), so that the possibility can be regarded as very unlikely.

NB

1. H. C. van de Hulst, Ned. Tijdschrift voor Natuurkunde 11, 210, 1945.

2. E. Hubble and R. C. Tolman, Astrophys. J. 89, 302, 1935.

3. R. Hanbury Brown and C. Hazard, Nature, 166, 901, 1950.

4. Note in B.A.A. Journal, March, 1951. 5. See ref. 13 on page 158.

Chapter 9. RADAR ASTRONOMY (F. J. Kerr)

Introduction. This is the branch of radio astronomy which deals with the application of the echo technique of radar. The first detection of echoes from the moon is of some philosophical interest, in that it was the first true "astronomical experiment." It was also the first direct measurement of an astronomical distance, albeit rather a crude one.

The radar method in astronomy has two special characteristics:--

(i) it provides the possibility of measuring distances,

(ii) it provides a controlled signal, hence eliminating some of the uncertainties inherent in working with a source of unknown characteristics.

The method is, however, restricted to fairly short distances, in consequence of the inverse fourth power dependence of intensity on distance, as compared with the inverse square law for a single direction of propagation as in visual or radio-noise astronomy. Echoes have so far been detected from meteor trails and the moon, and it should be possible to receive echoes from the sun and the inner planets. The outer planets might be reached one day, but it is difficult to imagine echoes ever being received from any of the stars.

Radar Equation. The flux density incident on a body at distance d from a transmitter radiating a power P_T from an antenna of gain G_T is:

$$S_I = \frac{P_T G_T}{4\pi d^2}$$

If the body has an "echo cross-section" σ, the flux density in the scattered wave at the receiver is:

$$S_R = \frac{\sigma S_I}{4\pi d^2},$$

which, with an antenna of gain P_R, produces an available power at the receiver of:

$$P_R = S_R \frac{G_R \lambda^2}{4\pi} .$$

Thus,

$$P_R = \frac{P_T G_T G_R \lambda^2}{(4\pi)^3} \frac{\sigma}{d^4} .$$

For a large body σ may be put equal to $A\rho D$,

where A = projected area of the echoing body,

ρ = reflection coefficient,

D = directivity of the scattered energy in the echoing direction, with in some cases an additional factor due to frequency dispersion of the echoed energy.

The sensitivity of a system to echoes depends on the signal-to-noise ratio, since the echo must be detected against a background of noise from the receiver, the galaxy, and possibly the sun.

In order to have free time to receive the echo, the transmitter is pulsed. The greatest sensitivity is obtained by using long pulses, as a smaller bandwidth can then be used, and less noise will be received. This is the basis of the high sensitivity obtainable in astronomical radar. In normal radar applications, a typical pulselength might be 1 microsecond, and bandwidth 1 Mc/s. (The minimum pulse length for a given bandwidth $\Delta\nu$ is $1/\Delta\nu$.) In astronomical radar, on the other hand, where the distances involved are considerably greater, pulselengths measured in seconds can be used, and correspondingly small bandwidths, limited only by the difficulty of obtaining sufficiently stable frequencies.

For precise range measurement, however, a larger bandwidth must be used, since a small receiver response time (inversely related to bandwidth) is necessary

when the commencement of the echo-pulse is to be precisely timed. The increase of bandwidth increases the noise and thus reduces the sensitivity, so that more powerful equipment is required for ranging than for merely detecting an echo.

Moon. Echoes from the moon were first detected by the U. S. Army Signal Corps in January, 1946 (1,2), and shortly afterwards by Bay (3) in Hungary. The Signal Corps group worked on 111.5 Mc/s (2.7 meters), with a pulselength of $\frac{1}{4}$ second, bandwidth 57 c/s, and transmitter power 3 kw. peak. The antenna, comprising 64 dipoles and a reflecting screen, was directed horizontally, so that observations were made near moonrise and moonset.

Theoretically, echoes were expected at an intensity 100 times noise, but it was found that on many occasions echoes could not be detected at all, while only rarely did the intensity reach the theoretical strength.

This result, together with some anomalies in observations of solar and galactic noise, led the Australian group to undertake moon echo studies to obtain information on transmission through the ionosphere (4, 5). The upper portion of the F region cannot be studied by the normal ionospheric technique, in which energy is reflected back to ground, but only by the use of an extra-terrestrial source or reflector.

For an ionospheric study a low frequency (around 20 Mc/s) was preferred and a short-wave broadcast transmitter was used in its free time. The use of existing transmitting equipment has obvious advantages, but the system was severely limited in its time of operation, since the moon passed through the antenna beam, in the period of the day when the station was available, only on 30 days a year.

In this system, taking the moon's reflection coefficient as 0.15, and neglecting ionosphere effects, the theoretical echo-to-noise ratio was 1000. However, the echo strength was usually considerably below this level, and moreover the echoes always appeared later than might be expected as the moon rose through the antenna beam.

Although far removed from curves based on standard theory, the observed figures for median intensity and altitude at first detection in the various tests were found to correlate quite well with the values for the critical frequency of the F_2 region of the ionosphere. Hence the anomalies must have arisen in the F_2 region indicating that there are shortcomings in the present theory relating oblique and normal incidence propagation conditions. It appears likely that the effects might be due to irregularities in the F region, so that a new oblique incidence theory which takes these into account is required.

The relative motion of the observer and the moon leads to a Doppler shift in the frequency of the reflected wave, amounting to about 300 and 50 c/s respectively under the conditions of the American and Australian experiments. This shift is due principally to the earth's rotation, but also in part to the radial component of the moon's orbital motion. It was thought at one stage that a precise measurement of the Doppler shift might lead to a new determination of the earth's radius and/or the velocity of light. However, uncertainty about the phase effects produced by scattering in the ionosphere would severely limit the accuracy obtainable (6).

In addition to the day-to-day effect due to the ionosphere, the 20 Mc/s echo intensities showed two types of more rapid variations. One type, with periods of the order of minutes, was apparently the same as the "twinkling" observed with radio stars, originating in the non-uniformities of the F region. The rapid variations, with periods of the order of seconds, was shown to be an effect of the moon's libration.

From the point of view of an observer on the earth, the moon undergoes an apparent oscillation of 1°-2°/day, about some (variable) axis through its center. At any instant one half of the moon is approaching the observer and the other half receding. At the limb, the radial velocity arising from libration amounts to about one meter/second, so that components of the 20 Mc/s echo coming from the limb and from the center will differ in frequency by about 0.1 c/s.

Thus, if the moon reflects as a rough body, with echo components coming from all portions of the disk, the originally monochromatic transmitted wave will be Doppler-broadened on reflection. Since the receiver bandwidth accepts the whole of the broadened "line," the instantaneous echo intensity will be the resultant of a very large number of components with random amplitudes and phases. Libration of the moon will vary the phases of these numerous components, so that the resultant intensity will vary from instant to instant, the rate of intensity variation depending on the rate of libration. From a study of the echo intensity record, a statistical measure was derived for the "fading speed" in each experiment. These values were found to correlate well with the libration rates on the various days, and to approximate to the figures obtained theoretically by attributing the fading to the libration of a rough moon.

No system yet used has been sufficiently powerful for making accurate distance-measurements. Hagen, of N. R. L., is proposing to use his new 50 ft. mirror at a wavelength of 3 cm. to make a precise measurement of the moon's distance, and hence to obtain a new determination of the solar parallax.

Sun. The sun is, after the moon, the simplest major astronomical body to reach by radar, and it is a more interesting object for study than the moon or the

planets. An order of magnitude for the power required to produce a detectable echo can be obtained by the following simplified argument:--

The angular diameter subtended by the sun at the earth is nearly the same as that of the moon ($\frac{1}{2}^{o}$), so that it will receive about the same power flux from a terrestrial transmitting antenna. Neglecting for the moment the differing reflection coefficients of the two bodies, the decrease in the sun's echo relative to that from the moon will be due solely to the greater distance traveled by the diverging reflected radiation. Since the sun is about 400 times as distant as the moon, the sun's echo will be smaller by a factor of $(400)^2$.

A more detailed discussion must of course take additional factors into account, such as the partial absorption of the incident energy in the solar atmosphere, the greater size of the sun's radio disk, the necessity of overriding the sun's own radiation ("solar noise"), and the variation of these and other factors with frequency.

It appears (7) that the detection of usable echoes from the sun is technically possible on present knowledge, though an engineering project of great magnitude would be involved. There is an optimum frequency range, at around 25-30 Mc/s, principally determined by the variation of the sun's reflection coefficient with frequency. Actually the use of a low frequency is indicated on astronomical grounds also, since interest centers largely in obtaining echoes from the outer corona and surrounding clouds or streams of charged particles because of the difficulty of optical observation in these regions.

Planets. An equipment capable of receiving echoes from the sun could also detect echoes from Venus and Mercury, provided that integration over a sufficiently

long period were used in the receiver output. Mars would be just beyond reach. Such echoes would not, however, be strong enough for precise distance measurement, so that little of astronomical value could be learnt, except for the rotation period of Venus, which could be determined from the Doppler-broadening of the echo.

For planetary work it would be better to use a considerably higher frequency, where it would be somewhat easier to obtain strong echoes, since in this case the reflection coefficient would not be expected to vanish as the frequency went up, and also there would be no question of having to compete with solar noise.

Solar noise reflected from the moon. Visually, substantially all the energy received from the moon is reflected solar radiation. It is therefore of interest to consider the corresponding radio circumstances. Also it might be found possible to detect some of the major increases of solar noise occurring at nighttime by observing the moon.

The flux density received by reflection from the moon is equal to that received directly from the sun, multiplied by a factor $\frac{\sigma'}{4\pi d^2}$, where σ' is the reflection cross-section of the moon in the direction considered, and d is the distance of the moon from the earth. This factor is approximately equal to 10^{-6}. Thus, since the moon's own brightness temperature is in the region of 300° K, an increment due to reflected solar noise might be detected on occasions when the sun's apparent temperature exceeds 3×10^8 °K, as it does during many outbursts.

Such reflected solar noise does not yet appear to have been detected. Steinberg and Zisler (8) reported having observed some bursts by reflection, but

from the quantitative evidence presented it appears that some other phenomenon, possibly a terrestrial effect, must have led to an erroneous interpretation.

Meteors. The radar method already has contributed greatly to the domain of meteor observations, since it operates equally well by day or night, and with or without clouds. Also it is more sensitive to faint meteors than the visual method, and has the additional advantage that it can be made semi-automatic in recording.

The importance of meteoric ionization in contributing to the ionosphere has been suspected for a long time, but the first good correlations of radio echoes with visual meteors was not obtained until work was carried out at shorter wavelengths. Hey and Stewart (9), in 1946, working at 5 meters, demonstrated the correlation by the simultaneity of individual phenomena, the diurnal variation and the association with meteor showers.

Below 8 meters, transient echoes are observed, with durations usually a fraction of a second, rarely exceeding a few seconds. The strongest of these echoes correlate well with visual meteors. Above 8 meters, the phenomena are more complicated, apparently due to the effects of residual ionization from previous meteors, or ionization from other sources.

The echo is obtained from the ionized trail left by the meteor, being strongest in a direction normal to the trail. Only occasionally is an echo seen from a meteor itself. The radiant of a stream can be determined from the time variation of the activity rate on a fixed antenna, i.e. by making use of the directed character of the reflection from the cylindrical trail. The velocity of a meteor may be found from a direct measurement of the time variation of the distance to the meteor, or more precisely from a study of the varying diffraction

pattern produced by the advancing head of the train, as evidenced by echo intensity variations.

Herlofson (10), using a theory which accounts satisfactorily for the phenomena observed by radar, has suggested that the kinetic energy of a typical meteor is divided into heat, light and ionization in the ratios $10^4:10^2:1$. On this theory, a meteor of visual magnitude +1 produces ionization to the extent of 10^{12} electrons per centimeter of path.

Possibly the most interesting result to date from radar work on meteors has been the discovery of great daylight meteor streams, extending through the greater portion of the northern summer.

Extensive reviews of meteoric work have been published by Lovell et al. (11), Hey (12), and McKinley and Millman (13).

Aurora. - Many early observations of the abnormal increases in ionization in the E region of the ionosphere, especially those made in polar regions, probably included cases of the reflection of radio waves from auroral formations, but the first specific investigation appears to have been made by Harang in 1940. Auroral echoes have since been obtained at frequencies ranging from 3.5 Mc/sec to 72 Mc/sec. References may be found in a paper by Aspinall and Hawkins (14) of the Manchester group.

References

1. Mofenson, J., Electronics, 19, 92, April, 1946.
2. DeWitt, J. H., Jr., and E. K. Stodola, Proc. I. R. E., 37, 229, 1949.
3. Bay, Z., Hungarica Physica Acta, 1, 1, 1946.
4. Kerr, F. J., C. A. Shain, and C. S. Higgins, Nature, 163, 310, 1949.
5. Kerr, F. J., and C. A. Shain, Proc. I. R. E., 39, 230, 1951.
6. Thomas, A. B., Aust. Journ. Sci., 11, 187, 1949.
7. Kerr, F. J., Proc. I. R. E., in press. 40, 650 (1952) (Auch. 10/51)
8. Steinberg, J. L., and S. Zisler, C. R., 229, 811, 1949.
9. Hey, J. S. and G. S. Stewart, Proc. Phys. Soc., 59, 858, 1947.
10. Herlofson, N., Reports on Progress in Physics, 11, 444, 1948.
11. Lovell, A. C. B. et al., Reports on Progress in Physics, 11, (389) 415 1948.
12. Hey, J. S., M. N., 109, 185, 1949.
13. D. W. R. McKinley and P. M. Millman, Proc. I. R. E., 37, 364, 1949.
14. A. Aspinall and G. S. Hawkins, Journal of the B.A.A., 60, 130, 1950.