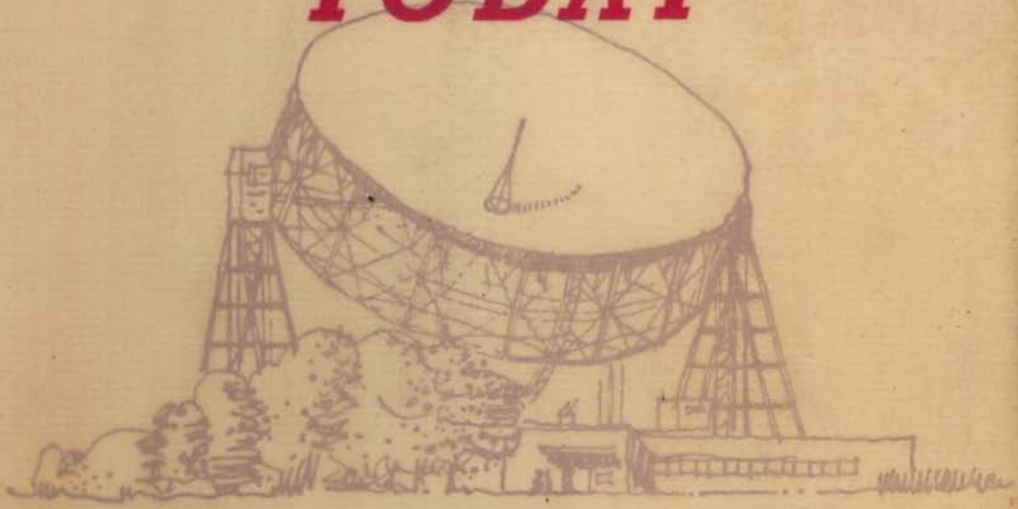
RADIO ASTRONOMY TODAY



edited by

H. P. PALMER

R. D. DAVIES & M. I. LARGE

with a foreword

by

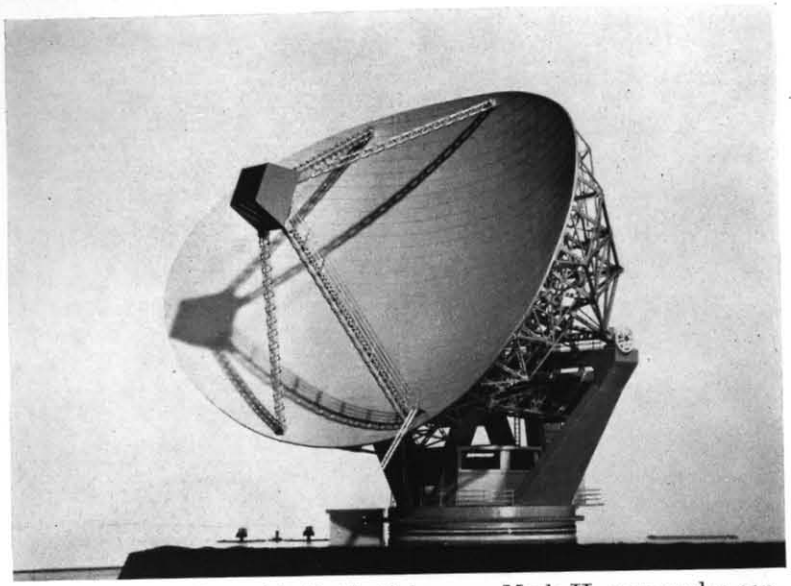
SIR BERNARD LOVELL

RADIO ASTRONOMY TODAY

Participants in the Summer School at Jodrell Bank.

Summer School students gaining practical experience in the analysis of records.

A 'Brains Trust' at the Summer School: R. C. Jennison, G. Burbidge, R. D. Davies, M. I. Large, H. P. Palmer and Professor Sir Bernard Lovell.



A model of the Steerable Radio Telescope, Mark II, now under construction at Jodrell Bank.

RADIO ASTRONOMY TODAY

Papers, designed to present Radio-Astronomy
in all its aspects, given at the
Jodrell Bank Summer School, 1962

edited by

H. P. PALMER, R. D. DAVIES
and M. I. LARGE

with a Foreword by

SIR BERNARD LOVELL

MANCHESTER UNIVERSITY PRESS

Published by the University of Manchester at
THE UNIVERSITY PRESS
316–324 Oxford Road, Manchester, 13

Printed in Great Britain by Butler & Tanner Ltd, Frome and London

FOREWORD

THE proposal to hold an international Summer School in Radio Astronomy was made by Mr. W. Burmeister when he was the Director of Extra-Mural Studies in the University of Manchester. It is believed that this School was the first of its kind anywhere in the world. The list of lecturers and their subjects shows that, with the help of the many distinguished visiting lecturers, we were able to arrange a general introductory course to Radio Astronomy, which, none the less, took participants close to the frontiers of knowledge in most of the main subdivisions of the subject.

The School was held at Jodrell Bank, mainly in the lecture room of the control building of the large radio telescope. The participants were therefore able to obtain a clear impression of the experiments in progress, and the facilities now available here. These have grown from nothing, fifteen years ago, to the present complex of observational instruments. In addition to the 250-foot fully steerable telescope (for wavelengths $\geqslant 21$ cm), these include two 30-foot dishes for the same wavelength range, and several other aerials for particular projects here and elsewhere in Cheshire. Telescopes of 50 foot diameter (polar mounted) ($\lambda \geqslant 1$ cm) and 80 $\times$ 125 foot elliptical ($\lambda \geqslant 3$ cm) are under construction. Those attending the Summer School were shown all these instruments, and they were able to spend some time studying each of ten observational programmes in the general fields of Meteor, Lunar and Planetary radar; the instrumentation of artificial satellites; observations of the celestial positions, radio spectra, and angular diameters of discrete radio sources; and galactic and extragalactic hydrogen line observations.

The lectures given during the course are reported in this book, in most cases by the lecturers, and in a few cases by the editors from tape recordings. Towards the end of each week of the course there was a discussion meeting, the first on 'Future Instruments in Radio Astronomy', and the second on 'Radio Astronomy and Cosmology'. These discussions were very interesting and lively, and are also reported here. It is hoped that this volume will be of interest to all those engaged in

observational or theoretical work in Astronomy or Radio Astronomy, and indeed to all scientists who retain an interest in our growing knowledge of the Universe.

A. C. B. LOVELL

Jodrell Bank
November 1962

ACKNOWLEDGEMENTS

THE editors gratefully acknowledge permission to reproduce diagrams and tables used by contributors which had previously appeared in:

Annales d'Astrophysique, Astrophysical Journal, Journal of the British Institute of Radio Engineers, Monthly Notices of the Royal Astronomical Society, Nature, Proceedings of the Kyoto Conference 1961, Proceedings of the National Academy of Sciences, Proceedings of the Paris Symposium on Radio Astronomy, 1958.

CONTENTS

PLATES

1

THE RADIO EMISSION OF THE SUN

J. P. WILD

The discovery of radio emission from the Sun in the early 1940's provided the terrestrial observer with the means of observing solar radiation in a vast new band of the solar spectrum. The band extends over four decades, covering wavelengths between a few millimetres and about 50 metres, and is limited only by the transparency of the Earth's atmosphere. Extra-terrestrial observations by rockets and satellites, now being introduced, will give further extensions.

Our previous knowledge of the Sun had come mostly from optical observations, which provided a ready-made framework of information on the Sun's structure, composition, and physical conditions, on which the radio studies could build; and here we shall be constantly referring to this framework. But much of the general pattern of radio phenomena, which has now emerged from 15 years of study, could never have been predicted from previous knowledge and it is the striking, unpredictable phenomena which have yielded the most important results. Extreme variability in the total flux of solar radio emission—up to 10^4 or 10^5 : 1, as compared with less than 1·01 : 1 in the optical spectrum—and extreme maxima in surface brightness—up to 10^{10} or 10^{11} °K, or millions of times higher than optical brightness temperatures—these properties have shown the solar radio spectrum to be largely unrelated to the optical and have revealed the existence of important non-thermal processes.

The significance of solar radio observations has emerged by degrees, and their contributions as we see them are several. They are complementary to optical observations in delineating the structure (notably density and temperature) of the steady and slowly varying atmosphere of the Sun; they are revealing the nature of physical processes in fully ionized plasmas previously unknown in the laboratory. They are making contributions to the understanding of eruptive processes in the solar atmosphere; and they form one of several new techniques which

are revolutionizing our knowledge of solar terrestrial relations.

In these two papers I shall be discussing these four contributions in that order. But I should mention a fifth aspect, namely the development of observational techniques, which has conditioned the whole growth of solar radio astronomy up to the present time. The gap in angular resolution between radio and optical observations has been the main problem, and the narrowing of this gap has provided the greatest challenge. The existence of this problem, itself a great drawback to radio observations, has perversely acted as a kind of stimulus towards a continuous succession of new kinds of observation and new approaches: it has been impossible to stagnate in a pool of routine observations such as have sometimes handicapped solar research in the past.

THE STRUCTURE OF THE SOLAR ATMOSPHERE

When pen recordings are made of the total flux of solar radiation at a particular radio wavelength, it eventually becomes evident that one is dealing with three main components with characteristic time scales. First, there is an apparently constant background of emission which varies only slightly with the 11-year solar cycle; this is due to thermal radiation from the 'quiet' solar atmosphere. Secondly, there is a slowly varying component of enhanced emission from centres of activity around sunspots with a time scale conditioned by the life-time of sunspots ($\sim$1 week to months) and by the period of solar rotation ($\sim$27 days). Thirdly, there is an intermittent and highly variable component ('bursts') with a time scale of seconds to hours related to flares and other eruptive phenomena in the Sun's atmosphere. Radio emissions contributing to our knowledge of the structure of the solar atmosphere belong mainly to the first two categories, which in some respects are almost inseparable.

Quiet Sun (see references 1, 2, 3)

In a later section we shall discuss the methods used to isolate the quiet from the variable components. For the time being let us assume that this process of subtraction has been effectively done.

The spectrum of the quiet component is shown in Fig. 1/1. At centimetre wavelengths the apparent disk temperature is about

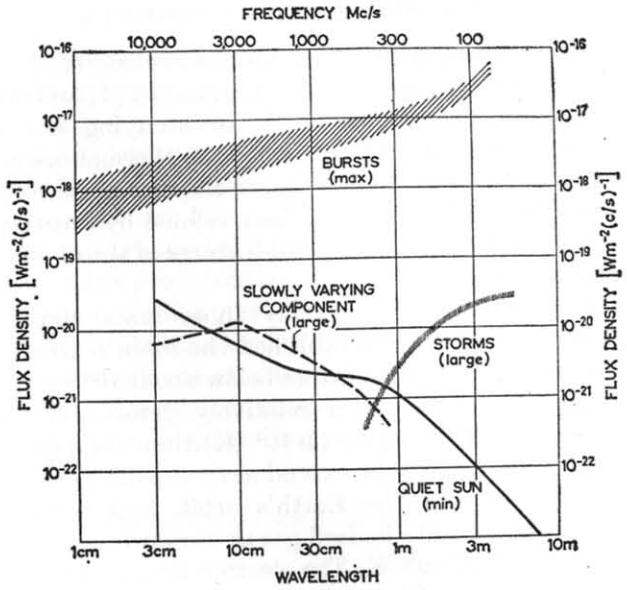
10^4 °K, rising to about 10^6 °K at metre wavelengths. As soon as this spectral pattern began to emerge, Martyn [1] and Ginzburg [1], applying concepts developed for studying the Earth's ionosphere, recognized independently the thermal origin of the component and laid the foundation to the thermal theory of the quiet Sun. The theory has since been refined by Smerd [1] and others, and explains all the general features of the observations, radio and optical.

Optical observations, especially those made during total eclipses of the Sun, have established the main features of the solar atmosphere. The lower strata below about 10^4 km comprise the *chromosphere* which is a relatively dense and partially ionized gas somewhat hotter ($\geqslant 10^4$ °K) than the photosphere (6000°K). The upper strata extend several solar radii at least, perhaps even beyond the Earth's orbit, and comprise the *corona*—a tenuous, fully ionized gas at curiously high temperatures of the order of 10^6 °K. The electron density in the corona decreases gradually with height according to a law originally derived from eclipse observations by Baumbach, and later refined by Allen and by Van de Hulst.

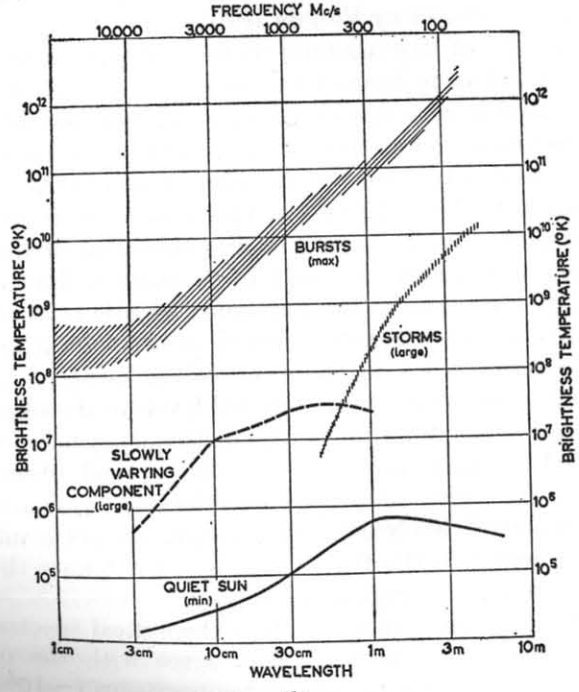
The theory of the thermal radio emission considers the transfer of radiation from one level in the solar atmosphere to the next, and is thus concerned with the emission, absorption and refraction of radio waves in an ionized gas. Emission of the radiation is due to free–free transitions as electrons execute hyperbolic orbits round protons. The absorption is determined as in classical radiation theory by an application of Kirchhoff's law; and the resultant intensity of emission is limited, as for black-body radiation, to the maximum kinetic temperature of the electrons. Refraction phenomena in the solar atmosphere, as in the terrestrial ionosphere, cause the escaping waves to follow curved paths and define critical levels in the solar atmosphere below which radiation of a given frequency cannot escape.

Detailed calculations, based on the optical observations, show that the short, centimetre wavelengths originate almost entirely in the relatively dense chromosphere and the metre and decametre waves in the tenuous corona; the longer the wavelength the higher the region of emission.

It is not surprising therefore that theoretical spectra of the total emission from the Sun's disk agree with the observations in predicting chromospheric temperatures ($\sim 10^4$ °K) for



(*a*)



(*b*)

FIG. 1/1.

centimetre waves, coronal temperatures ($\sim 10^6$ °K) for metre waves, and intermediate temperatures for decimetre waves.

A striking result of the theory was the prediction that at centimetre and decimetre wavelengths the solar disk should exhibit pronounced limb brightening owing to the positive temperature gradient in the lower solar atmosphere and the increased opacity offered by higher levels when these are viewed tangentially. No such limb brightening was predicted at metre wavelengths, partly because no positive temperature gradient was expected in the outer corona, and partly because the corona refracts or reflects metre waves more effectively than it absorbs them. Widespread efforts were thenceforth made to test these experimentally and obtain brightness distributions of the radio disk.

The observation demanded angular resolution (~minutes of arc) far better than any previously encountered in radio science. The resolution was beyond the scope of conceivable aerial arrays, and eclipse techniques proved rather inconclusive. The challenge was met by instrumental developments at Cambridge, where the concepts of Michelson's stellar interferometer were adapted to the radio case, and at Sydney where multi-element interferometry was developed. The Sun's brightness distribution has now been studied at many wavelengths, though the observations are still limited by resolution, by the time required to synthesize a single picture (often extending over several days), and by assumptions of quadrant symmetry adopted to simplify the analysis and smooth the data. ('Pencil beam' scans of the Sun now available have not yet proved sensitive enough to show the features of the quiet Sun, at least during the recent period of sunspot maximum.) Nevertheless, excluding an early

FIG. 1/1.—The spectrum of the general components of solar radio emission. The quiet sun refers to data recorded at sunspot minimum; the slowly varying component to average levels during sunspot maximum (the I.G.Y. period). The intense variable emissions are divided into 'storms' and 'bursts', for which the shaded curves indicate approximate maximum values.

In (*a*) the spectrum is plotted as flux density versus frequency; in (*b*) the spectrum is given in temperature units so as to give a very rough indication of the brightness temperatures of the different components. For the latter purpose the emitting area is assumed to be: the optical disk for the quiet sun; the sunspot area for the slowly varying component; and estimated mean emitting areas based on available observations for the storm and burst component (S. F. Smerd).

observation which may have been confused by the presence of centres of activity, the observed brightness distributions have verified all the main predictions of the theory, including limb brightening at centimetre and decimetre wavelengths (e.g. Fig. 1/2*a*) and limb darkening at metre wavelengths. The observations also showed up marked departures from circular symmetry, which, though adopted by the original theories for simplicity, was clearly not to be expected on the basis of eclipse photographs.

The quiet Sun has been found devoid of any observable degree of polarization, consistent with the now known smallness of the Sun's general magnetic field.

However satisfactory it appeared at first sight, the general agreement between prediction and observation seemed to lead to the disappointing conclusion that the radio observations contributed nothing substantially new to our knowledge of solar physics. It is conflict rather than agreement that stimulates interest and leads to new ideas. The role of radio studies has so far been to test existing models rather than derive new ones. To what extent can the radio observations be exploited? What is most needed from the radio observation is a means of determining unambiguously the distribution of electron density, N, and kinetic temperature, T, in the chromosphere and the transition region, and the distribution of T in the corona.

Using an ingenious mathematical treatment, Piddington [1] has demonstrated the feasibility of deriving N and T in the chromosphere (where refraction effects are negligible) under a restricted set of assumptions including circular symmetry; but the present state of solar physics requires much more refined results. Two main obstructions show themselves at present: the lack of observations of sufficient accuracy, resolution and frequency range; and the analytical difficulties in separating the effects of N and T. The former may be overcome by increasingly elaborate instrumentation, the latter perhaps by the solution of sets of simultaneous equations with high-speed computers.

The Sun's radio emission provides data on the solar corona out to heights of 1 or 2 solar radii. Important radio astronomical explorations have however been made far beyond these heights by studying the refraction effects imposed by coronal irregularities upon radio waves transmitted through the corona from distant radio sources [3]. The most important of these

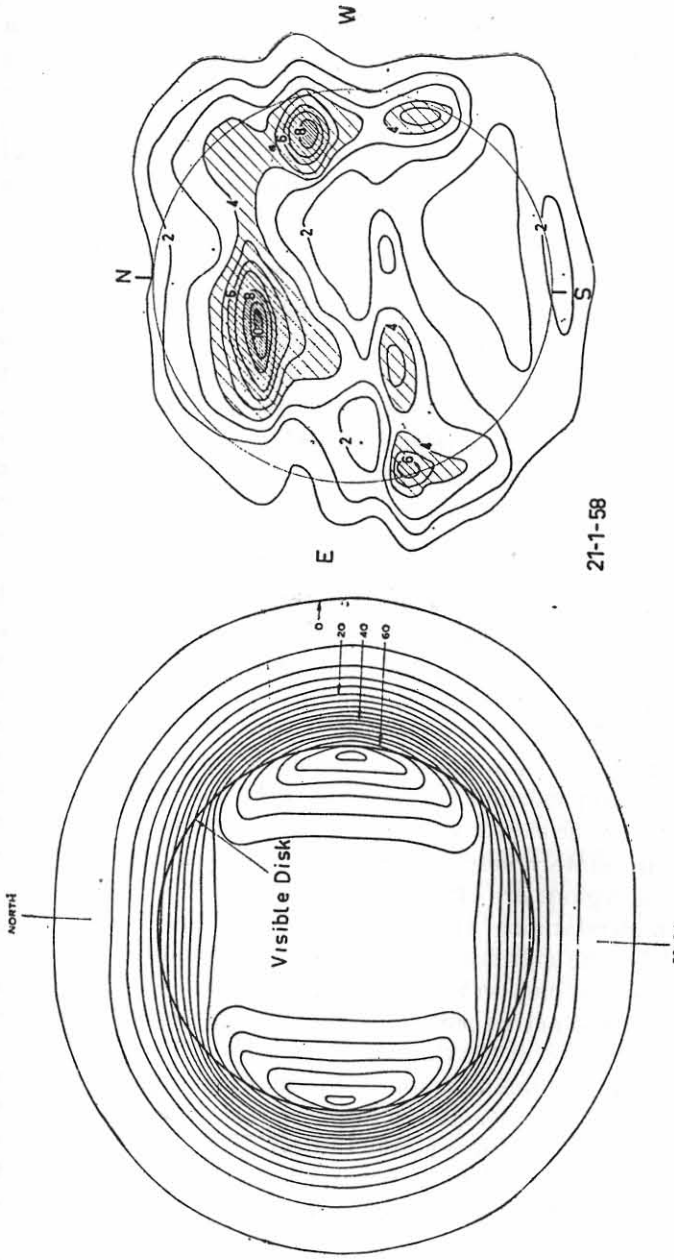


FIG. 1/2.—The distribution of radio brightness across the Sun's disk at a wavelength of 21 cm: (*a*) Quiet Sun at sunspot minimum, (*b*) the active Sun with radio plages (W. N. Christiansen).

sources is the Crab Nebula which is occulted by the corona during June each year. Many workers, notably Hewish in Cambridge and Vitkevich in Moscow, have used this technique to study the irregularities out to heights of about 20 R_0. More recently Slee, in Sydney, succeeded in observing the coronal occultation of a number of different radio sources and was able to detect the influence of the corona out to nearly 100 R_0.

The slowly-varying component (see references 1, 2, 3, 5)

The character of total-flux records changes markedly over the radio spectrum. At centimetre and decimetre wavelengths, the quiet and slowly varying components are prominent, the burst component rather rare. At metre and decametre wavelengths, the quiet component is relatively feeble, the slowly varying component obscure or non-existent, while the burst component is dominant and rapidly variable—that is at times of eruptive activity. Thus to resolve the quiet and slowly varying components, we turn first to the short wavelengths.

Isolation of the slowly varying component at centimetre and decimetre wavelengths may be achieved by plotting daily fluxes against sunspot area. One finds a high correlation between the two quantities, and the large residual flux which remains at zero sunspot area may be defined as the quiet component. There appears to be an even higher and simpler correlation between radio flux and calcium plage areas, and it is now generally believed that the slowly varying component is due to hot, dense regions intimately associated with calcium plages. For this reason the regions are often known as *radio plages*.

The understanding of these phenomena has been greatly clarified by the cultivation of highly directional observations. Indeed the first clue on the relation between the slowly varying component and active centres came from a directional observation by Covington at the time of a solar eclipse. The radio plages were first located unambiguously by simultaneous observations at several points on the Earth during eclipses, and were found, as expected, to lie above active centres. Since then a series of beautiful instrumental developments, pioneered by Christiansen and his colleagues at a wavelength of 21 cm, have resolved the regions, first in one dimension and later in two.

The first of these developments was the multi-element interferometer [1], comprising 16 equally spaced aerials in a row;

this instrument gave a fan beam some $3\frac{1}{2}'$ wide which was allowed to scan the Sun by the Earth's rotation. The radio plages showed themselves clearly on these records, and their day-to-day movement due to solar rotation was evident. When many such strip scans were superimposed on one another, a clearly defined lower envelope, representing the quiet component, was recognized.

The second important instrumental development was responsible for two-dimensional pictures of the Sun [3]. A pair of multi-element interferometers each 1400 ft long were arranged to form a cross and were connected together in the fashion of a Mills Cross to yield a pencil beam 3′ arc in diameter. This beam was scanned across the Sun's disk to form, in a period of the order of 1 hour, a television-like image of the radio Sun. Pictures thus obtained (Fig. 1/2*b*) showed the radio plages as quasi-circular regions, typically 3–5′ arc in size, and confirmed their general coincidence with calcium plages.

Higher resolution studies have been made in the 3–10 cm wavelength range by Kundu [3] and others using long-base simple interferometers, and a fine structure has emerged consisting of a concentrated nucleus of strongly circularly polarized radiation of size commensurate with the underlying sunspot, and a wider, less polarized structure corresponding to the calcium plage.

The physical properties of the plage emissions have now been determined from observations of various kinds over most of the decimetre and centimetre wavelength ranges. Their mean spectrum is indicated in Fig. 1/1. Determinations of height, made by studying the displacement in position between radio plage and optical sunspot for different positions on the disk and bursts of the Sun, lie in the range 0–50,000 km; these measurements are difficult and approximate.

Optical observations, in the coronal emission lines, from the limb of the Sun, had shown the existence of dense dome-like regions over active centres, reaching heights of about 10^5 km which Waldmeier called coronal condensations. It became clear that the radio plage component referred to emission from these condensations or regions very clearly linked with them; and that the radio emission offered important possibilities of exploring their structure, especially in the optically inaccessible *transition region* between corona and chromosphere. One first needed

to know the physical process responsible for the emission. Since radio plages rarely or never exceed 10^7 °K, it was natural to make the simplest assumption, that the emissions were thermal and originated in free–free transitions. Next came the problem, as with the quiet sun, of interpreting the radio observations unambiguously in terms of electron density (N) and temperature (T) as functions of height above the photosphere. The first attempts to perform this operation with the rather inadequate data available relied heavily on assumption and plausible models, but were sufficient to demonstrate that very high densities existed in these regions. It is now clear that the *unambiguous* interpretation of these quantities requires high-resolution observations over a wide frequency range, sufficient to yield the variation of both brightness and height of the region with frequency. A commendable attempt [5] to demonstrate the feasibility of this complicated operation was made during the I.G.Y. by pooling the data of four stations (Nançay, Sydney, Toyokawa and Washington) for the passage of a selected active centre across the Sun's disk. Although the data, especially of height, lacked the accuracy to be desired, the model which emerged gives a first-order picture and demonstrates a principle for future investigations. The desired model (Fig. 1/3) reflects the characteristics of the current view now held, that these regions represent a highly overdense region ($\simeq 3 \times 10^9$ cm^{-3}) up to a height approaching 10^5 km, merging at greater heights into coronal streamers. Temperatures appear to be normal for the solar atmosphere.

In this discussion no mention has been made of magnetic fields, which are the most important single factor in determining the structure of active regions. Magnetic effects show themselves directly at centimetre wavelengths as partial circular polarization in the general plage emission. They also show themselves indirectly by the 'turn-over' of the flux spectrum at a frequency between 5 and 10 cm (see Fig. 1/1). The existence of this maximum is highly significant; it casts doubt on the hypothesis that the plage emission is due entirely to free–free transitions, since it can be shown that the flux from a hot, ionized slab of gas must never decrease with frequency. Japanese and Russian workers have interpreted this effect in terms of a component due to radiation from electrons precessing round magnetic lines of force, emitting low harmonics of the gyro frequency. The addi-

tion of this complexity in the interpretation of the plage radiation is at first sight rather perplexing, but it could eventually be of great value in exploring the magnetic field in the Sun's

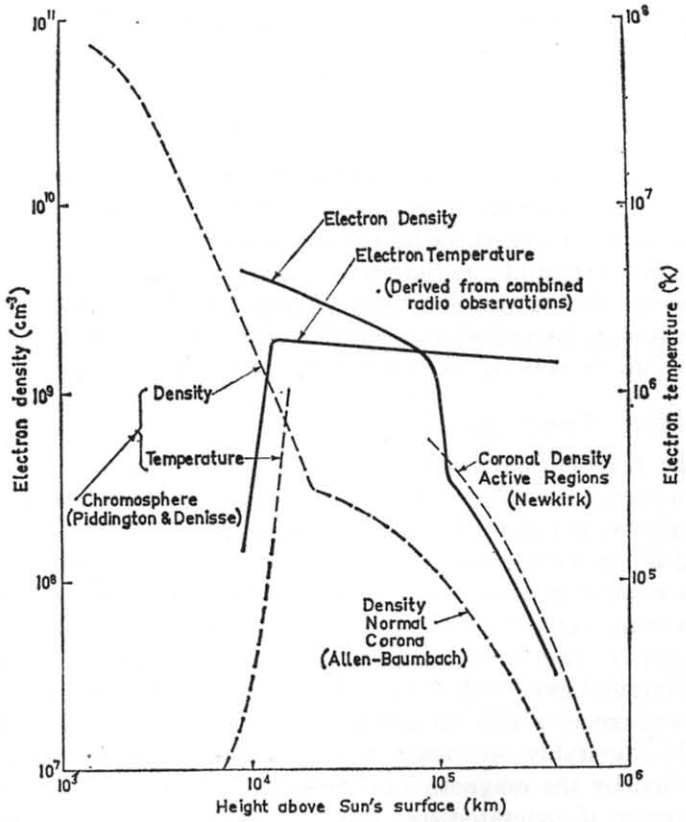


FIG. 1/3.—The distribution of electron density and temperature, with height in the solar atmosphere, over a selected plage region. The model is derived purely from radio observations at several stations across the world (see reference 5).

chromosphere. Preliminary investigations of this effect indicate fields of the order of 600 gauss.

The plage regions are of special interest because they now appear to be the main source of continuous X-ray emission from the Sun, an effect recently demonstrated directly from rocket-borne photographs. Likewise the regions control the ionization of the ionosphere, an effect found by Denisse and Kundu who

demonstrated a high correlation between the flux of the slowly varying component and the ionization of the E-layer.

Finally the dense regions above sunspots are the breeding grounds of the great eruptions to be discussed later, and a knowledge of the physical conditions within them is essential to an understanding of the eruptive processes.

SOLAR-ERUPTIONS

In the previous section we discussed the radio emissions whose variations occur over long periods ($\geqslant$days) and whose brightness temperatures ($\leqslant 10^7$ °K) fell within the black-body limit of thermal-like radiation appropriate to the physical conditions of the solar atmosphere. We now turn to the spasmodic, rapidly varying emissions ('bursts') whose brightness temperatures (up to 10^{11} °K) far exceed the reasonable thermal limit.

Non-thermal processes

The non-thermal processes of emission responsible for the radio bursts are of great interest in themselves; these, to our knowledge, fall into the following classes:

(*a*) *Electron gyro radiation.* In the presence of a magnetic field, H, a moving electron executes circular or helical orbits at the gyro frequency $eH/2\pi mc$, where e and m denote the electronic charge and mass (in general the relativistic mass). However, the fundamental gyro frequency radiated by electrons in the Sun's atmosphere is unable to escape outwards because to do so it would inevitably encounter a region of non-propagation, as specified by the magneto-ionic theory. Harmonics of the gyro frequency, if generated (see (*b*)), may escape; hence the term 'gyro radiation' usually refers to the emission of the low harmonics of the gyro frequency.

(*b*) *Synchrotron radiation.* In the case where the velocity of the gyrating electrons is close to the velocity of light, the radiated energy is received (by a distant observer situated near the orbital plane) in the form of pulses rather than waves, owing to the doppler bunching which takes place once per period during the phase of the electron's rapid approach. The spectrum of this train of pulses will clearly consist of a multiplicity of harmonics of the gyro frequency, which under natural conditions becomes smeared into a continuum. This radiation is generally called

synchrotron radiation, after the machine by which it was first accidentally and perhaps unwillingly observed. The emission due to a single electron can be exactly computed, but absorption effects in an assembly of radiators are not yet completely understood. Roughly, the brightness temperature from such an assembly is limited to the 'kinetic temperature' of the electrons. It is useful to remember that following this crude argument, a brightness temperature of 10^{10} °K requires electrons with energy of about 1 Mev.

(*c*) *Plasma radiation.* It would perhaps have been tempting to invoke the synchrotron process to explain all non-thermal emissions in radio astronomy were it not for the observation that certain solar bursts show a spectrum with 2 : 1 harmonics. A characteristic frequency must be emitted, and this frequency is apparently not the gyro frequency since in that case the fundamental would not be observed. Rather, in these cases, is the frequency identified with the plasma frequency, indeed the internal consistency of the evidence on harmonic types of solar bursts seems to place this identification beyond reasonable doubt.

The plasma frequency is the frequency at which natural oscillations take place in an ionized gas when the positive and negative charges are separated and then released. It is given by

$$f_p = \frac{e}{\pi m}\sqrt{N},$$

where N is the electron density. Longitudinal plasma waves—travelling waves of electron density—can be excited in an ionized gas by streams of charged particles; the waves are reminiscent of the bow wave of a ship. The waves can propagate freely in a restricted range of frequencies immediately above the plasma frequency and are highly dispersive. Their wavelength is invariably shorter than the wavelength (in the medium) of the electromagnetic wave of corresponding frequency; hence they do not couple directly with the electromagnetic field, i.e. they do not radiate by themselves. The extra physical circumstances needed for the generation of radiation is a question which is receiving much attention at the present time, and various solutions have been suggested. One, of special interest because no magnetic field is required, is the scattering theory of Ginzburg and Zhelezniakov [3]. These workers have shown that, when

pervaded by travelling plasma waves, irregularities in electron density (such as must exist in the corona) will act as incoherent radiators at frequencies near the plasma frequency; though the coupling process is inefficient, the flux seems to be sufficient to account for the observed intensity of the fundamental emission band. These authors further note that certain of the coronal irregularities will themselves consist of plasma wavelets, oscillating at the plasma frequency, and that scattering of the original travelling wave at these oscillatory irregularities can lead to direct emission of electromagnetic waves at twice the plasma frequency: by this means Ginzburg and Zhelezniakov interpret the 'second harmonic' emission. The theory is especially attractive since it explains why only the 1 : 2 bands, rather than the 1 : 2 : 3 : 4 : . . . bands, are observed. Other more direct means of coupling, involving the existence of magnetic fields, have been proposed to explain rather different phenomena, notably by Denisse (see Boischot [4]).

The solar flare and its widespread effects

An association between radio bursts and solar flares was recognized by Hey in 1942 at the time of his discovery of the variable radio emissions. Since then it has become clear that nearly all the different classes of radio phenomena occur at the time of flares, or belong to a class of events characteristically associated with flares, or are triggered by flares, or occur in flare-producing regions. So before considering the radio observations in detail, let us briefly review the relevant physics and effects of flares.

A flare is seen optically as a brightening, notably in the line of H_α, over part of a centre of activity, usually close to a sunspot group. It usually develops suddenly and decays slowly, the whole event lasting some five minutes to one hour. Its shape is variable: small ones, which may occur at the rate of several per hour during sunspot maximum period, are often seen merely as small blobs; large ones, much rarer, often show filamentary structure. Out of the flare, rising jet-like surges can often be seen in either emission or absorption, and the eruption of existing arch-like structures or prominences may follow. The optical emission may be accompanied by a burst of X-rays which causes temporary, but immediate, excess ionization in the Earth's lower ionosphere and consequent absorption of communication

radio waves passing through it (the Sudden Ionospheric Disturbance). Large flares cause the emission of solar protons in a wide energy spectrum between 10^3 and 10^{10} e V; the harder particles possess the rigidity to penetrate the Earth's magnetic field and cause an increase in the flux of cosmic rays recorded by ground-level detectors; while the softer particles, which require rockets or balloons for their direct detection, penetrate the ionosphere in the polar regions and cause ionization by collisions again leading to absorption of communication radio waves (Polar Cap Absorption). Depending on their energy the proton events arrive at the Earth tens of minutes to several hours after the flare; the delay is attributed to magnetic effects in the interplanetary medium.

Other effects occur on the Earth some 30 or more hours after the flare—aurorae, magnetic storms and decreases in the cosmic ray flux. These effects are attributed to the influence of corpuscular streams or magneto-hydromagnetic waves travelling from Sun to Earth with mean velocities of the order of 1000 km/s; the magnetic storm requires a major disturbance in the solar corona and involves a significant fraction of the total coronal matter lying above an active region. There is some evidence that the effects of flares extend beyond the Earth, as far as Jupiter.

The radio emission from flares is proving to be critical in understanding the origin of these effects, and perhaps of the nature of the eruption itself.

The energy of a flare, revealed predominately in the emission of light, X-rays and protons, amounts to as much as 10^{32} ergs in exceptional cases and is currently believed to be stored in the Sun's atmosphere in the form of magnetic energy. Most modern theories envisage a flare as the sudden release of part of this magnetic energy by the merging together and consequent partial destruction of a pair or complex of differently polarized field configurations such as may exist above a sunspot group.

The radio emissions

The character of the variable component of solar radio emission varies greatly with frequency (Fig. 1/4). The greatest intensities and most rapid fluctuations are seen at metre wavelengths: bursts lasting seconds or minutes and storms lasting hours or days. At centimetre wavelengths the variations are

slower, the bursts, lasting from about 1 minute to 1 hour, less frequent, and the storms absent. The emissions, originally observed only with single frequency radiometers, are now investigated with respect to their spectrum, polarization, position

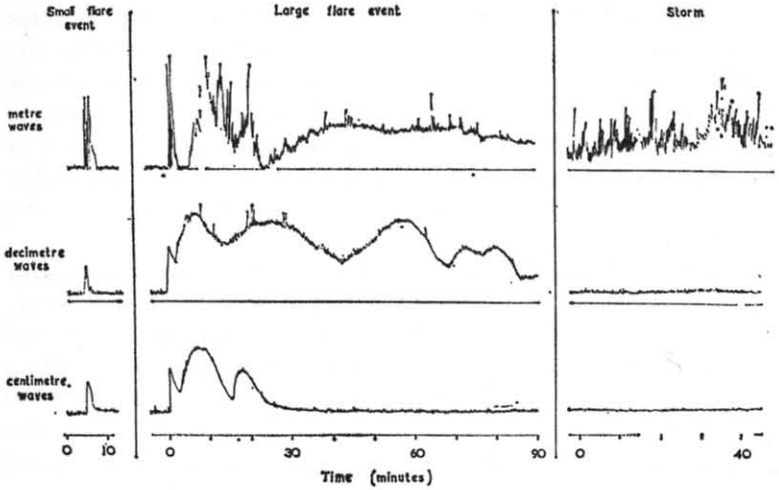


FIG. 1/4.—The appearance of records at different wavelength ranges while bursts and storms are in progress.

on the disk, and size. They are most conveniently classified in terms of their spectrum—their so-called 'dynamic spectrum', which gives the variation of intensity in the frequency-time plane—which also provides considerable physical insight into the nature of the phenomena.

Storms

On a time-scale of days the most prominent variable component, restricted to metre wavelengths, consists of storms of short-lived, narrow-band bursts (spectral Type I), sometimes superimposed on a broad-band slowly fluctuating background (see Fig. 1/4). The storm radiation is often strongly circularly polarized, and emanates from regions above certain sunspots. The radiation from one region may persist for several days while the spot traverses the central longitudes of the Sun—i.e. the emission is restricted to a rather narrow cone. Storms sometimes start shortly ($\leqslant 2$ hours) after flares. The radiation is certainly non-thermal in origin, but the actual emitting process is controversial. Current theories tend to favour an origin in plasma oscillations, though the possibility of gyro radiation cannot be excluded.

Small flare events

The commonest type of isolated burst, known as spectral type III, occurs at metre wavelengths and can accompany the smallest of optical flares, normally near their start (Wild [4]). They occur in groups lasting about one minute, each individual burst lasting about 10 seconds. The dynamic spectrum shows a progressive delay of lower frequencies after high frequencies—i.e. they appear in the time-frequency plane as bands sloping slightly away from the vertical. In favourable circumstances one can clearly identify simultaneous pairs of bands showing the 2 : 1 harmonic relationship. Groups of type III bursts are sometimes followed by a short period of continuum radiation, designated spectral type V.

We have seen earlier that the presence of 2 : 1 harmonics indicates plasma oscillations as the origin of emission. This is the most conclusive reason for believing that each type III burst is caused by the outward passage through the corona of a stream of electrons which excites plasma waves of progressively lower frequency as it passes through regions of progressively lower electron density. With the aid of the electron density models of the corona (e.g. Fig. 1/3), we find that the rate of frequency drift in type III bursts corresponds to outward stream velocities of about $0{\cdot}2c$ to $0{\cdot}5c$, where c is the velocity of light. This *plasma hypothesis* has now been confirmed by direct observation of the relative positions of different frequencies emitted by individual bursts near the limb of the Sun: low frequencies were found to originate at greater heights in the atmosphere than high frequencies. Investigations of this kind have given information on the variation of electron density with height in the corona, and the results show higher than normal densities, consistent with those found in coronal streamers (e.g. Newkirk's curve in Fig. 1/3). It appears that the electron streams flow out along coronal streamers.

While the metre-wave type III burst is in progress, one commonly observes a coincident event at centimetre wavelengths consisting of a simple smooth burst with one or a small number of maxima, as depicted in Fig. 1/4. Striking cases have

been found (Kundu [4]) in which the detailed time profile of these bursts corresponds closely to that of the bursts of X-rays, revealed with balloon-borne apparatus. The solar X-rays are believed to be generated by free–free transitions (Bremsstrahlung) as fast electrons and other particles pass through the chromospheric matter; likewise the centimetre-wave burst may be explained by the generation of synchrotron radiation as the same stream of electrons passes through the magnetic field of the active region in the chromosphere. And it is natural to infer that these electrons are derived from the same explosion as that responsible for the type III bursts (the absence of sharp bursty structure in the chromospheric phenomena may be due to the effect of the strong chromospheric magnetic fields which trap and hold the electrons, so smoothing the time profile of the emissions). We conclude from these studies that most if not all flares eject fast electrons in a series of short sharp puffs; and all the indications are that the instant of time when these ejections take place coincides with the time at which the flare region erupts and ejects the much slower particles seen optically as surges.

Large flare events

Turning now to the very large and completely developed flare event, we find (Figs. 1/4–6) radio phenomena of extraordinary complexity, still in the process of being sorted out and understood. The final scheme developed here (Wild [4]) is tentative and undoubtedly over-simplified.

The complete radio flare event has two distinct phases. The first phase is similar to that described above for small flares—i.e. a group of type III bursts at metre wavelengths accompanied by a burst at centimetre (and shorter decimetre) wavelengths referred to as a 'microwave early burst'. After the type III bursts have subsided, the level of metre-wave emission remains quiet for several minutes before the start of the second. Let us first consider the second phase in the metre-wave region (Fig. 1/5).

The second phase begins with a complex but remarkably characteristic type of burst, known as spectral type II. It lasts for 5–30 minutes, shows sharp bands and nebulous structure which drift in frequency slowly (~200 times slower than the type III drift) across the metre-wave spectrum. It usually has

pronounced 2 : 1 harmonics, and so again its emission is attributed to plasma oscillations, a hypothesis strongly supported by directional evidence. Indeed the type II burst is interpreted, like the type III burst, as an outward-moving disturbance, but at a velocity much slower than the type III—about 1000 km/s is the derived radial component of velocity. This speed is several

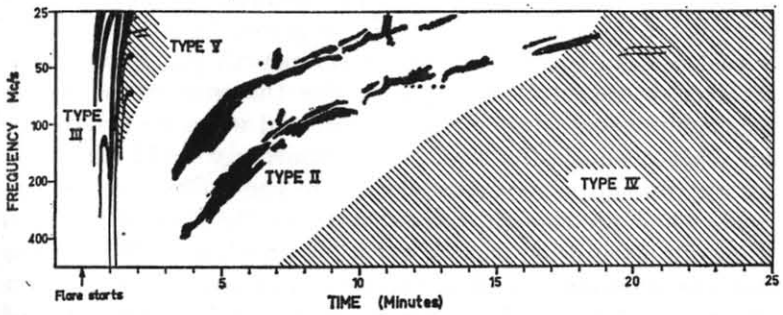


FIG. 1/5.—Idealized dynamic spectrum record showing the appearance of a large solar outburst at metre wavelengths.

times greater than the speed of sound in the corona, and the source of emission is believed to be a magneto-hydrodynamic shock wave passing through the corona; the evidence is consistent with the shock wave emanating from the original explosion at the same time as the ejection of the fast electrons responsible for the first phase.

In very large events, the decline of the type II burst merges with, or is followed by, the onset of a long persistent (~hours), smooth, broad-band continuum known as spectral type IV, whose significance was first recognized in Paris by Boischot and Denisse [3, 4]. Using a high resolution, multiple interferometer these workers showed that during the first half-hour of its lifetime, the type IV source can move away from the flare region at speeds of the order of 1000 km/s and can attain heights of several solar radii. Observations of similar phenomena in Sydney showed furthermore that, unlike the type II and type III 'plasma' emissions, the type IV source emitted a wide band of frequencies from the *same position*. These characteristics seem to preclude plasma oscillations as the process of emission, and strongly favour the synchrotron mechanisms, proposed by Denisse. At a later stage, the source of the continuum radiation becomes strongly circularly polarized and is located above the

original active centre where it may persist for hours or days. The French workers have given evidence, based mainly on the narrow cone of emission during this latter stage, that plasma rather than synchrotron radiation is then responsible.

The notion that the type IV continuum consists of more than one component is intensified when we consider the decimetre and centimetre accompaniments. On the macroscopic scale, the complex radio spectrum of a major flare event is dominated by a vast area of continuum, whose component parts cannot be resolved by spectral data alone. For this reason the whole continuum is called spectral type IV, and when we believe we can isolate physically separate components we add suffixes—e.g. m, dm and μ, for metre, decimetre and microwave components which are thought to be distinct on the basis of observations of polarization, cone of emission, position on disk, source size, etc. A current model, arrived at during a recent symposium at Kyoto [4] in consultation with representatives from most of the research groups concerned, is shown in Fig. 1/6.

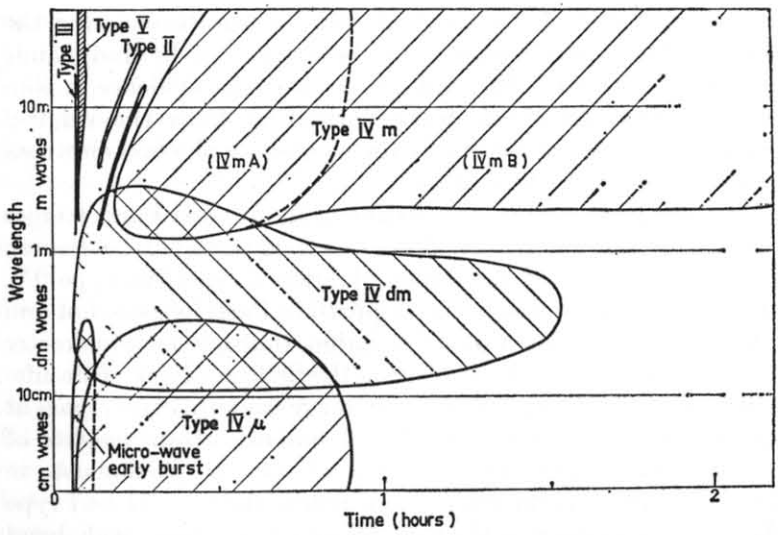


FIG. 1/6.—Idealized model of the spectrum of a complete radio outburst, showing the different possible components of the type IV burst (Kyoto conference, 1961[4]).

The decimetre and microwave components of the type IV burst remain extremely tentative, but the work of Takakura, Tanaka, Kundu and others is beginning to disentangle the phenomena. Present indications are that these emissions originate mainly in synchrotron radiation generated by electrons accelerated and trapped in the magnetic fields above the flare region.

SOLAR-TERRESTRIAL RELATIONS AND THE INTERPRETATION OF THE FLARE

The first indications that the solar phenomena being observed by radio emission were to throw light on unsolved problems in the solar-terrestrial relationships, came from the study of type II bursts. The derived velocities ($\sim$1000 km/s) of the type II sources were of the correct magnitude to explain the time delay ($\geqslant 1\frac{1}{2}$ days) between great flares on the Sun and great magnetic storms and aurorae on the Earth; it was suggested that in a type II burst, we were witnessing the launching of the disturbance linking the two phenomena. Statistical evidence partially supported this view, but now it has become clear [4] that the probability of a type II burst being followed by a geomagnetic storm becomes high only when a type IV burst is also present.

A second relationship of great interest concerns the solar protons which accompany certain large flares and which very occasionally produce cosmic-ray increases on the Earth at ground level. The proton events are found to be almost invariably accompanied by type IV bursts (though the converse is not true).

A further close relationship exists between centimetre bursts and sudden ionospheric disturbances, a relationship which further supports the close connection between the generation of X-rays and of centimetre waves in the solar chromosphere.

From the foregoing discussion we can begin to speculate upon the picture of the invisible accompaniments of a major flare event—it is the invisible rather than the visible processes which are fundamental, because we are dealing with highly energetic phenomena in an almost completely ionized gas; and the power of the radio observation lies in their reference to free rather than bound electrons.

The initial flare centre appears to disrupt in two ways simultaneously (Fig. 1/7*a*). Firstly it disrupts by the emission of sharp bursts of fast electrons. Some of these travel outwards

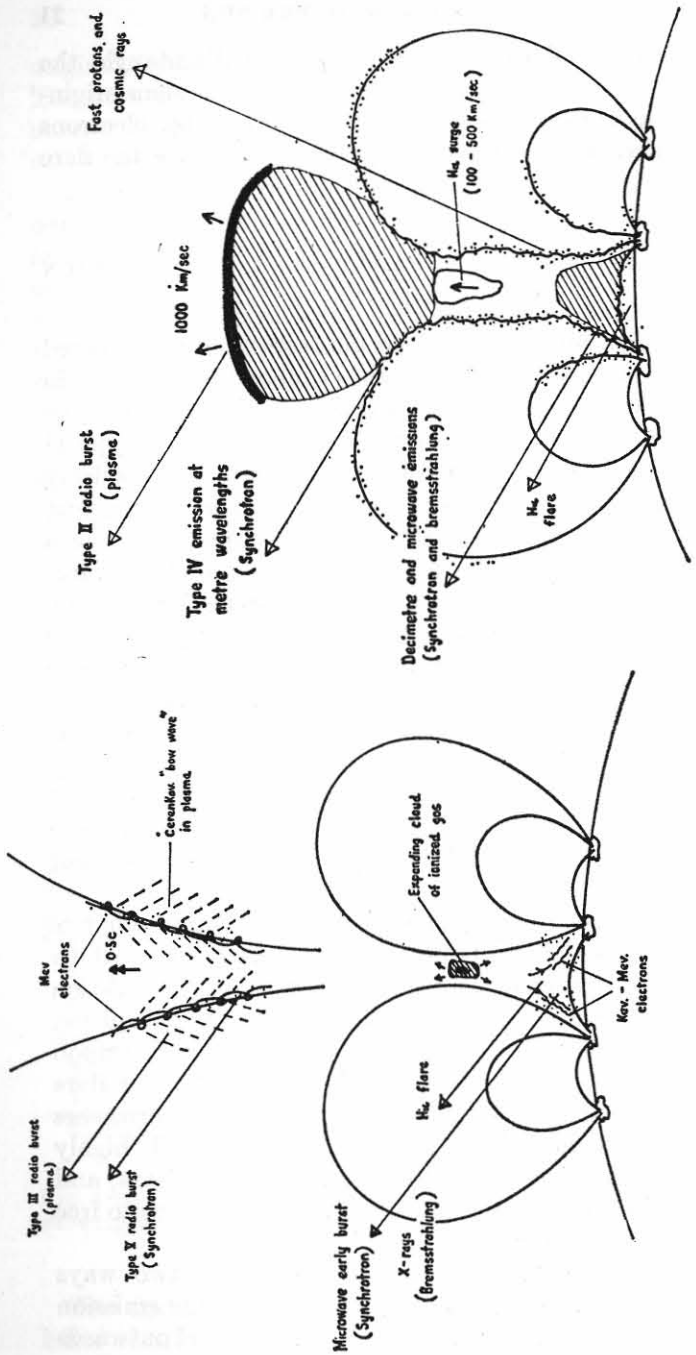


FIG. 1/7.—The possible origin of the two phases of a solar outburst, assuming the initial flare event to be due to the partial self-annihilation of a pair of opposing magnetic fields above a sunspot group. The first phase is ascribed to bursts of high-speed electrons, the second to a simultaneously emitted shock wave.

along coronal streamers delineating lines of force, and escape from the Sun within a few seconds; these excite plasma oscillations in transit, to generate type III bursts. Other electrons encounter strong magnetic fields in the chromosphere and emit synchrotron radiation in the centimetre-wave spectrum and Bremsstrahlung in the X-ray spectrum. These effects persist for a minute or two.

Secondly (Fig. 1/7*b*), the disruption occurs by the ejection of ionized gas (ions and electrons), whose flux is accompanied by a magneto-hydrodynamic shock wave. In some way a portion of this matter reaches the Earth a day later to produce the magnetic storm. The shock wave itself travels out at a speed of about 1000 km/s and is revealed, by the plasma oscillations set up at its front, as the type II radio burst. Electrons contained in the ejected matter entwine and trap themselves in the neighbouring magnetic fields to generate the type IV synchrotron emissions. At the same time the protons, moving in changing magnetic fields, are subject to Fermi acceleration and may attain energies sufficient to escape and produce the solar protons detected on Earth. When the kinetic energy of the gas clouds is great enough, it will be free to move bodily away from the Sun, dragging with it and stretching the magnetic field in its path; by this means travelling synchrotron sources develop far from the Sun.

During the next solar cycle more sophisticated and complete methods of observation will surely be forthcoming. We may hope then for a deeper understanding of the phenomena and perhaps for the discovery of a variety of new facts.

REFERENCES

1. *The Sun*, ed. G. P. Kuiper, Chicago, 1952.
2. *Handbuch der Physik*, ed. S. Flügge, **52**: *Das Sonnensystem* (mainly in English), Springer Verlag, Berlin, 1959.
3. *Paris Symposium on Radio Astronomy*, ed. R. N. Bracewell, Stanford University Press, 1959.
4. *Proceedings of the International Conference on Cosmic Rays and the Earth Storm* (Kyoto, 1961), II Joint Sessions, *Jnl. Phys. Soc.*, Japan, Vol. 17, Supplement A-II.
5. *Annales d'Astrophysique*, **23**, 75, 1960.

2

PLANETARY EMISSIONS

F. T. HADDOCK

THE first observation of microwave emission from the planets and satellites was the detection by Dicke and Beringer (1946) in 1945 of 1 cm wavelength emission from the Moon using a 3-ft diameter paraboloid. This pointed to the feasibility of detecting the thermal emission from the solar planets using large aerials and sensitive receivers at microwave frequencies. However the next report was that of the unexpected outbursts from Jupiter observed at 22 Mc/s by Burke and Franklin (1955). This impulsive radiation consisted of bursts with a 1 Mc/s bandwidth and 1 second duration. Since the radiation was strongly circularly polarized it was inferred that the radiation was near the local gyro frequency which would make the Jovian magnetic field at least 7 gauss.

The steady microwave emission expected from the planets was observed at the Naval Research Laboratories by Mayer, Sloanaker and McCullough (1957) from Venus at 3·15 cm and 9·4 cm and from Mars and Jupiter at 3·15 cm. Saturn was observed by Drake and Ewen (1958) at 3·75 cm and by Cook, Cross, Bair and Arnold (1960) at 3·45 cm. Howard, Barrett and Haddock (1961) first detected the weak emission from Mercury.

Observations of the planets to the present date may be summarized as follows. Jupiter has been observed in the wavelength range 60 m to 3 cm while Venus has been observed in the range 20 cm to 4 mm. Only measurements near 3 cm have been made on Mars, Saturn and Mercury and no observations have been made of Uranus, Neptune or Pluto. It is interesting to note that the two planets observed at more than one frequency have produced unexpected results; Venus is more than 100°C hotter than expected and Jupiter has two separate non-thermal components.

The thermal emission from the planets gives rise to a flux density spectrum proportional to the inverse square of the wavelength and accordingly a wavelength (λ_1) can be specified for

each source at which the flux density equals 10^{-26} w m^{-2}(c/s)$^{-1}$. This wavelength and the flux density of the planets at 10 cm wavelength are presented in Table I.

TABLE 2/I

Thermal emission from the planets

Planet	λ_1 cm	Flux at 10 cm (units of 10^{-26} w m^{-2}(c/s)$^{-1}$)
Pluto	0·08	$6{\cdot}4 \times 10^{-5}$
Neptune	0·30	9×10^{-4}
Uranus	0·65	$4{\cdot}2 \times 10^{-3}$
Mercury	3·0	9×10^{-2}
Saturn	4·0	0·16
Mars	6·5	0·42
Jupiter	11·0	1·1 (thermal)
		6 (non thermal)
Venus	30	9

Many different types of observation have been made of the radiation from the planets. Spectra may be obtained from observations at different frequencies and the relative thermal and non-thermal contributions can be determined. The two brightest sources both show variations in intensity with time; Venus shows a minimum in observed disk temperature at the time of inferior conjunction and Jupiter shows variability at centimetre wavelengths which may be associated with solar activity. The low frequency variability of Jupiter has been studied by Warwick (1962) using a dynamic spectrograph. Other observations of interest which have been carried out at Caltech are the measurement of linear polarization and the radio angular diameter of Jupiter.

A comment will be made at this point on the accuracy of the measurement of T_D, the disk temperature of a planet.

$$T_D = \frac{\lambda^2 T_A}{A_e \Omega}$$

where λ is the wavelength, T_A the aerial temperature due to the planet, A_e the effective area of the aerial which is equal to the product of the physical area and the efficiency and Ω is the solid angle subtended by the planet. The value of the

efficiency of the aerial can be determined from an absolute determination of the aerial gain or from a comparison with a source whose flux has already been related to an absolute calibration. In the case of emission from a solid surface a correction needs to be made to the derived disk temperature $T_{\tau=1}$ to allow for the radio-frequency albedo A_{RF} of the planetary surface and so

$$T_D = T_{\tau=1}(1 - A_{RF})$$

For the Moon this correction is about 10 per cent and the corresponding disk temperature refers to a level several centimetres below the surface. When calculating T_D for a gaseous body Ω may be uncertain and it may also be a function of received wave polarization and time.

Saturn

Saturn has been reliably measured with a 3·45 cm wavelength four-level ruby maser at the University of Michigan using the 85-ft radio telescope. The overall system noise was 100°K and the bandwidth was 8 Mc/s giving an r.m.s. noise of 0·015°K when using a 12 sec time constant. The observed aerial temperature of Saturn was 0·095 ± 0·02°K corresponding to a disk temperature of 106 ± 21°K after making a 2 per cent correction for atmospheric absorption. This value is in good agreement with the infra-red temperature of 123°K found by Menzel, Coblentz and Lampland (1926). This planet requires measurements at longer wavelengths to see if it is similar to Jupiter and perhaps has a non-thermal spectrum. Moreover observations at a later date when the projected angle of the rings is small will show if the rings themselves radiate.

Mars

Mars has only been observed at the Naval Research Laboratory, in Washington, at wavelengths near 3 cm. In September 1956 70 observations yielded a disk temperature of 218 ± 50°K, while in November 1959 a temperature of 211 ± 20°K was obtained using a maser (Giordomaine, Alsop, Townes and Mayer, 1959). These values can be compared with 245 ± 9°K measured by Menzel, Coblentz and Lampland (1926) and 260° found by Pettit and Nicholson (1924). It appears that the radio disk temperature is 40°C lower than the infra-red value. This difference can be attributed to the fact that the infra-red

measurement refers to the sunlit surface whereas the radio measurement is an average value for the temperature a small distance beneath the solid surface. Alternatively, the discrepancy could be due to a finite radio-frequency albedo. Thus the results at 3 cm wavelength can be interpreted as thermal emission from the planet.

Mercury

Mercury has been detected by Howard, Barrett and Haddock (1961) at the University of Michigan using a maser at 3·45 cm and a travelling wave tube at 3·75 cm. The mean aerial temperature observed with the 85-ft telescope was 0·05°K which corresponds to a mean disk temperature of 400°K. Since the planet only presents its partially sunlit surface to the Earth the temperature of the subsolar point (T_0) will be greater than this value. On the assumption that the planet is smooth, with uniform reflectivity, is not rotating and has no atmosphere or internal heating then the temperature at a point of longitude angle θ from the subsolar point will be $T(\theta) = T_0 \cos \frac{1}{4}\theta$. The resultant value of T_0 is 1100 ± 300°K.

A number of effects may modify this estimate of T_0. If there were surface irregularities such as mountains which cast shadows near the sunlit edge of the planet the calculated value of T_0 would be increased. On the other hand if the planet were at a uniform temperature throughout the sunlit hemisphere the estimated subsolar temperature would be decreased by 30 per cent. The 23·7° libration of Mercury in its elliptical orbit will reduce the calculated value of T_0 if the surface layers are similar to those of the Moon and cause a thermal lag in the heating at the depths responsible for 3 cm radiation. This reduction may be as much as 15 per cent. On the assumption that the radioactive heating of Mercury is similar to that of the Earth it can be estimated that the temperature of the dark side is 25°K while at 84 cm, the depth at which the optical depth of dry sandy soil is unity at 3 cm wavelength, the temperature will be 43°K. This again reduced the deduced subsolar temperature. The combined effect of these processes is to bring the 3 cm value of T_0 close to the optical value given by Kuiper (1952) of 600° to 700°K or the value given by Pettit and Nicholson (1924) of 613°K.

The major difficulty in obtaining accurate estimates of the

aerial temperature of Mercury is its low intensity and its proximity to the Sun (less than 28°). At 3 cm observations cannot be made when the Sun is below the horizon because the planet is at low elevations where atmospheric scintillation and absorption are troublesome. The planet has to be detected in the presence of the weak sidelobes of the Sun some 60 db down in intensity. The travelling wave tube radiometer because of its large bandwidth has the advantage over the narrow-band maser of smearing out the sidelobe positions into a rather uniform side-lobe response level.

Venus

The observations of the disk temperature of Venus as a function of wavelength are summarized in Table 2/II, compiled

TABLE 2/II

Wavelength	Average disk temperature (°K)
4 mm (Grant & Corbett, private communication)	350 ± 50
4 mm (Kisliaker, Kuzmin & Salomonovich, private communication)	390 ± 120
8·0 mm (Kuzmin & Salomonovich, 1960)	315 ± 70
8·6 mm (Gibson & McEwan, 1959)	410 ± 160
3·15 cm (Mayer, McCullough & Sloanaker, 1958*a*, *b*)	595 ± 55
3·37 cm (Alsop, Giordomaine, Mayer & Townes, 1958, 1959)	575 ± 58
3·4 cm (Mayer, McCullough & Sloanaker, 1960*b*)	575 ± 60
9·4 cm (Mayer, McCullough & Sloanaker, 1958*a*, *b*)	580 ± 160
10·2 cm (Mayer, McCullough & Sloanaker, 1960*a*, *b*)	600 ± 65
21 cm (Lilley, private communication)	580

by C. H. Mayer. Venus was the first planet to be observed at millimetre wavelengths where the temperature is found to be about 400°K and is significantly less than at centimetre and decimetre wavelengths (600°K). These values may be compared with that of 330°K obtained by Menzel *et al.* (1926) in the infra-red band 12·5 to 15 microns and with that obtained by Sinton and Strong (1960) of 240°K in the range 8 to 14 microns. It is implied in these results that the centimetre-wave value of 600°K refers to the solid surface of the planet while the millimetre and infra-red results come from the cooler atmosphere.

There are several theories in circulation purporting to explain the reasons for this large temperature difference between the Venusian surface and its atmosphere. Only the briefest comments will be made on them. The first of these (Wildt 1940, Sagan 1961) invokes the greenhouse effect caused by a dense carbon dioxide atmosphere. However in order to obtain a surface temperature as high as 600°K almost all the solar radiation must be trapped and only a small fraction can escape, which brings this theory into difficulties.

Another theory by Sagan, Siegel and Jones (1961) envisages a very dense Venusian ionosphere at 600°K which lies above cool regions at about 300°K. A value of electron density N_e is required which makes $\int N_e^2 dl$ throughout the ionosphere equal to 10^{26} electrons cm^{-5}. This corresponds to an electron density of 10^9 electrons cm^{-3} which is a thousand times greater than that of the Earth's ionosphere. Such an explanation also has its difficulties. Other explanations of the observed phenomena are the aeolosphere theory of Opik (1961) and the theory requiring a redistribution of the emission due to particle charge fluctuations proposed by Tolbert and Straiton (1962).

A further observation of some interest that can be made on Venus is the search for a variation of the disk temperature with phase. Such a variation if found would provide evidence for the direction and possibly the rate of rotation of the planet and with aid in the choice of a proper model for the radio emission at centimetre wavelengths.

Jupiter

The evidence for a non-thermal component in the decimetre radiation from Jupiter is presented in Table 2/III, also compiled by C. H. Mayer. The temperatures measured near 3 cm wavelength (170°K) are very similar to those found in the infrared. However as the wavelength increases up to 60 cm the temperature increases to above 10,000°K indicating a strong non-thermal component. Evidence at present available also shows that this non-thermal radiation is variable and there is an indication that the variability is associated with solar activity.

A most interesting advance in the study of Jupiter has been the measurement of the polarization and radio emission diameter of the planet at 30 cm wavelength, by Radhakrishnan

TABLE 2/III

Wavelength	Average disk temperature (°K)
3·15 cm (Mayer, McCullough & Sloanaker, 1958*b*)	145 ± 18
3·03 cm (Giordomaine, Alsop, Townes & Mayer, 1959)	171 ± 20
3·36 cm (Giordomaine, Alsop, Townes & Mayer, 1959)	189 ± 20
3·75 cm (Drake & Ewen, 1958)	200
10·3 cm (Sloanaker, 1959) July 1958	640 ± 57
10·3 cm (Sloanaker & Boland, 1960) October 1959	315 ± 45
21 cm (McClain, 1959)	2,500 ± 450
31 cm (Roberts & Stanley, 1959)	5,500
68 cm (Drake & Hvatum, 1959)	50,000

and Roberts (1960). This work showed that Jupiter's radiation at this frequency could be as much as 43 per cent polarized and that it came from a region 2′ in diameter suggesting an origin in synchrotron emission from relativistic electrons trapped in Jupiter's van Allen type belts.

The low-frequency radiation near 20 Mc/s appears to originate in regions fixed relative to the solid surface of the planet because the observed rotation period of the radio emitting areas has been observed to have a constant value of $9^h\ 55^m\ 29{\cdot}37^s$. The source of the emission is of interest since the characteristics of the bursts are different in important respects from those observed from the Sun or from lightning flashes in terrestrial thunderstorms. The narrow frequency bandwidth, the change of the central frequency from pulse to pulse, the pulse duration and the circular polarization suggest that the radio waves are produced in an ionized gas excited at its plasma frequency. The method of excitation is not clearly understood; large mechanical or electromagnetic energies of 10^{18} erg per pulse are required to give the observed total radiated radio energy.

Conclusion

Radio studies of the planets have made a number of specific contributions to our understanding of planetary conditions.

1. The discovery of a high surface temperature on Venus has led to a re-examination of theories of planetary atmospheres.

2. The decimetre wavelength radiation from Jupiter suggests a van Allen type belt around the planet with an associated magnetic field of at least about 7 gauss.

3. Long period studies of the metre-wavelength radiation from Jupiter have localized emitting regions which are probably associated with the solid surface and thus give the rotation period of the solid body of Jupiter.

4. Large dynamical phenomena must occur on Jupiter to give the metre-wave outbursts.

5. There are on the planets a variety of sources of radio emission and propagation mechanisms which can be studied close at hand.

6. These new results provide important new data on the conditions on the planets which is a great help in planning the space probe programmes.

7. Planets are radio sources with accurately known positions and provide a means of calibrating the pointing of microwave radiometers.

REFERENCES

Alsop, L. E., Giordomaine, J. A., Mayer, C. H., and Townes, C. H., 1958, *A.J.*, **63**, 301.
1959, *Paris Symposium on Radio Astronomy*, ed. R. N. Bracewell, Stanford University Press, p. 69.
Burke, B. F., and Franklin, K. L., 1955, *J. Geophys. Res.*, **60**, 213.
Cook, J. J., Cross, L. G., Bair, M. E., and Arnold, C. B., 1960, *Nature*, **188**, 393.
Corbett, H. H., 1962, private communication.
Dicke, R. H., and Beringer, R., 1946, *Ap. J.*, **103**, 375.
Drake, F. D., 1962, *Pub. Nat. Radio Ast. Obs.*, **1**, 165.
Drake, F. D., and Ewen, H. I., 1958, *Proc. I.R.E.*, **46**, 53.
Drake, F. D., and Hvatum, H., 1959, *A.J.*, **64**, 329.
Epstein, E., 1959, *Nature*, **184**, 52.
Gibson, J. E., and McEwan, R. J., 1959, *Paris Symposium on Radio Astronomy*, p. 50.
Giordomaine, J. A., Alsop, L. E., Townes, C. H., and Mayer, C. H., 1959, *A.J.*, **64**, 332.
Giordomaine, J. A., Alsop, L. E., Mayer, C. H., and Townes, C. H., 1959, *Proc. I.R.G.*, **47**, 1062.
Grant, C. R., 1962, private communication.
Howard, W. E., Barrett, A. H., and Haddock, F. T., 1961, *A.J.*, **66**, 287.
Kislaiker, A. G., 1962, private communication.
Kuzmin, A. D., and Salomonvich, A. E., 1960, *A.J.* (U.S.S.R.), **37**, 297.
Mayer, C. H., McCullough, T. P., and Sloanaker, R. M., 1958*a*, *Ap. J.*, **127**, 1.
1958*b*, *Proc. I.R.E.*, **46**, 260.
1960*a*, *A.J.*, **65**, 349.
1960*b*, paper read at XIIIth General Assembly, U.R.S.I., 1960.
Mayer, C. H., Sloanaker, R. M., and McCullough, T. P., 1957, N.R.L. Rep. Nos. 4998 and 5021.

McClain, E. F., 1959, *A.J.*, **64**, 339. Paper read at the XIIIth General Assembly U.R.S.I., September 5–15, 1960.
Menzel, D. H., Coblentz, W. W., and Lampland, C. O., 1926, *Ap. J.*, **63**, 177.
Opik, E. J., 1961, *J. Geophys. Res.*, **66**, 2807.
Pettit, E., and Nicholson, S. B., 1924, *Pop. Astr.*, **32**, 601 and 614.
Radhakrishnan, V., and Roberts, J. A., 1960, *Phys. Rev. Letters*, **4**, 493.
Roberts, J. A., and Stanley, G. J., 1959, *Pub. A.S.P.*, **71**, 485.
Sagan, C., 1961, *Science*, **133**, 849.
Sagan, C., Siegel, K. M., and Jones, D. E., 1961, *A.J.*, **66**, 52.
Sinton, W. M., and Strong, J., 1960, *Ap. J.*, **131**, 459 and 470.
Sloanaker, R. M., 1959, *A.J.*, **64**, 346.
Sloanaker, R. M., and Boland, J. W., 1960, paper read at XIIIth General Assembly U.R.S.I.
Tolbert, C. W., and Straiton, A. W., 1962, *J. Geophys. Res.*, **67**, 1741.
Warwick, J. W., 1962, private communication.
Wildt, R., 1940, *Ap. J.*, **91**, 266.

Further general references:

The Solar System, ed. G. P. Kuiper and B. M. Middlehurst, Vol. III. *Planets and Satellites*. Chicago. University of Chicago Press, 1961.
The Atmosphere of the Earth and Planets, ed. G. P. Kuiper. 2nd edition. Chicago. University of Chicago Press, 1952.

3

INTERPLANETARY RADAR

J. H. THOMSON

THE subject of this lecture is interplanetary radar and it will be confined to this, leaving out any detailed consideration of the allied but at present much more sophisticated subject of Moon radar. However it is useful to summarize the chief results of Moon work, as they give some indication of the probable behaviour of the planets as radar reflectors.

There are two extreme cases of reflection by a sphere large compared with the wavelength when it is illuminated by a point source. In the first case when the sphere is smooth compared with the wavelength a single bright spot is seen in the centre, which is an image of the source. At optical wavelengths this is how a ball bearing behaves. The second case, when the surface of the sphere has irregularities comparable with a wavelength in size, gives appreciable reflection from the whole sphere, but with limb darkening. At optical wavelengths this may be thought of as the way a ball of chalk would reflect.

The Moon is found to be an intermediate case between these two, but more like the first: the Moon by radar is closer to a ball bearing than a ball of chalk. Most of the energy returned comes from a circular area about a fifth of the Moon's diameter in size at the centre of the disk. The power flux returned to Earth is as though about 5 to 10 per cent of the total power intercepted by the disk were reradiated isotropically. This behaviour seems to be largely independent of wavelength over a wide range from metre to centimetre wavelengths.

The main scattering properties of the Moon for radar may be satisfactorily explained on the assumption of surface irregularities large compared with the wavelength, having slopes of up to about 10 degrees (*e.g.* Hughes, 1962). The proportion of those reflecting specularly then falls off rapidly away from the centre of the disk, giving the observed brightness distribution.

There are two extreme cases of radar, first when very short pulses are used, and second when very long pulses or CW is

used. The first gives good discrimination in depth and range. A short pulse successively illuminates circular annuli of a spherical target; these annuli have the property that the angle of incidence is the same all round a given annulus. Hence a display of echo amplitude against time allows the law relating the back-scattered power with angle of incidence to be derived. This is the so-called angular power spectrum. The range can also be determined by measuring the time from the transmission of a pulse to the return of the echo.

In the second case the target is illuminated with CW. Range information is no longer available, but instead the shift of frequency between that transmitted and that received can be measured with very great accuracy to give the line of sight velocity of the target. Usually the target will be rotating about some axis not necessarily perpendicular to the line of sight. If velocities due to rotation are resolved along the line of sight, points on the disk with equal resolved velocities lie on lines parallel to the projection of the axis of rotation on the apparent plane of the disk. Thus the returned signal will have a finite spread in frequency, and the power spectrum gives directly a strip distribution of power across the disk.

In both these cases the primary quantity, range or line of sight velocity may be determined whether or not the whole of the echo can be observed. In both cases the secondary quantity, concerned with observing the spread of the echo, in range or in frequency, is usually difficult to measure completely, as a large dynamic range is required (of the order of 20 db or more for the Moon). This can be very important in the second case where the complete width of the spectrum may be needed to deduce the rotation rate of the target. It is possible to combine these two basic methods of investigation by the use either of coherent pulses as at the Lincoln Laboratories (Pettengill, 1960) or by some form of modulation (Victor, Stevens and Golomb, 1961) of a CW transmission. Such methods are likely to be the basis of much future astronomical radar work.

THE RADAR EQUATION

The equation giving the signal/noise ratio to be expected in a radar system will now be developed and its application to the problem of obtaining echoes from the planets considered.

If P is the power radiated from the aerial and G is its gain, the flux of power at a planet at a distance R is

$$\frac{PG}{4\pi R^2} \text{ watts}/m^2$$

The power intercepted by the disk of a planet of radius a is this multiplied by the disk area πa^2 or

$$p = \frac{PG}{4\pi R^2} \pi a^2 \text{ watts}$$

Conventionally a fraction ρ of this intercepted power is considered to be reflected and $(1 - \rho)$ absorbed. The power ρp is reradiated with a gain g in the direction of the Earth. Then the power flux at the Earth's surface is

$$\frac{(g\rho)p}{4\pi R^2} \text{ watts}/m^2$$

The term $(g\rho)$ is placed in brackets to indicate that g and ρ cannot easily be separated. At the Earth's surface an effective area of aerial A_e is used to collect the returned power. So

$$\text{Signal Power} = \frac{PA_e^2 f^2}{4\pi c^2 R^4} (g\rho)\pi a^2 \text{ watts}$$

where $G = \dfrac{4\pi A_e f^2}{c^2}$ has been used.

This received signal power must be detected against a noise power made up of receiver noise and sky noise. If a system temperature of T_E°K is assumed, then

$$\text{R.F. Signal/Noise} = \frac{PA_e^2 f^2}{4\pi c^2 R^4} (g\rho)\pi a^2 / T_E k b$$

where b is the receiver bandwith.

This bandwidth is an important factor in determining the signal/noise. In a pulsed radar b must at least $\dfrac{1}{\tau}$ where τ is the pulse length. Hence if $\tau = 5$ ms (a rather long pulse to explore say Venus with a radar depth of 41 ms), $b \sim 200$ c/s. If resolution is increased by reducing τ, b goes up in proportion and the signal/noise falls. Thus in the pulsed case b is determined by another constant τ of the system, rather than by the nature of the target. In a CW system the intrinsic bandwidth of the

transmissions is likely to be very small indeed, but a correspondingly narrow bandwidth cannot be employed because the target planet itself, due to rotation, imposes a bandwidth on the echo.

Limb to limb this is $$4a\left(\frac{2\pi}{T}\right)\frac{f}{c}$$

where T is the period of rotation with the axis assumed perpendicular to the line of sight, and c the velocity of light. This would be rather wide to use for b if the system is to be optimum for detection as only a portion of the disk near the centre is likely to be an effective reflector. Hence we shall put

$$b = \alpha 4a\left(\frac{2\pi}{T}\right)\frac{f}{c}$$

where α expresses the linear fraction of the disk diameter effectively returning power. Reasonable values of b for a planet always turning the same face to the Sun (such as Mercury) are a few cycles, up to a few hundred for say Mars, rotating as fast as the Earth. In the case of the trapped planets an increase in signal/noise of the order of 20 db can thus be gained on bandwidth alone, over the pulsed case.

Substituting for b and rearranging, we obtain for the CW case

$$\text{RF signal/noise} = \left(\frac{1}{32\pi kc}\right)_1 (A_e^2)_2 (Pf)_3 \left(\frac{1}{T_E}\right)_4 \left(\frac{g\rho}{\alpha}\right)_5 \left(\frac{aT}{R^4}\right)_6$$

The numbered factors in brackets are determined respectively by

1. Constants of nature.
2. Available radio telescope.
3. Transmitter power and frequency.
4. System and sky noise.
5. Properties of planet as a radar target.
6. Astronomical quantities for planet.

Factor 5 is unknown till experiments have been successfully made, but for the Moon is of the order of unity. Evidently factor 6 is the one to examine to decide the relative detectabilities of the planets. Table 3/I gives this on a db scale, which has been chosen so that Venus at closest approach is zero db. The slow rotation of Mercury and Venus (assumed trapped also) gives them a great advantage as CW targets over the fast-spinning

Mars and Jupiter. The two figures given for each of the planets is the effect of the change of R due to orbital movement.

TABLE 3/I

R.F. Signal/Noise in db for the planets on a scale giving a ratio of unity for Venus at closest approach

	Closest	Furthest
Mercury	−20	−34
Venus	0	−23
Mars	−38	−66
Jupiter	−62	−68
Saturn	−74	−78

SPECIAL PROBLEMS OF RADAR ASTRONOMY

Diplexing

The problem of transmitting and receiving from the same radio telescope can be avoided by having two radio telescopes (Victor, Stevens and Golomb, 1961) in bistatic operation. This solution has many advantages; it doubles the observation time and eliminates the possibility of delicate receiving equipment being damaged. Otherwise some arrangement to stop the transmitter power from getting to the receiver must be made. Fortunately the time of the transmit-receive cycle is large enough to allow even relays to be used. The Jodrell system exploits the change in sense of circular polarization on reflection to gain more than 40 db rejection by a hybrid ring system, followed by a receiver protection relay. This use of circular polarization also eliminates troublesome fading experienced in linearly polarized systems caused by Faraday rotation of the line of polarization in the Earth's ionosphere.

Changing range

In pulsed systems the range is changing at rates of the order of km/s because of orbital motion. As only a small portion of the complete time base will be observed some system must be used to follow the echo as it changes range according to the expected planetary movement. This may be done electronically in real time by a special purpose computer (as in the first Jodrell

system) or in a specially programmed digital computer in real time or later, as in the usual M.I.T. systems (Pettengill, 1962).

Changing frequency

In both pulsed and CW systems the changing relative velocity of the two planets causes a varying doppler shift of the order of a kc/s in a few hours, further complicated by the observer's motion on the rotating Earth. Receiver tuning must thus be continually adjusted according to a previously calculated ephemeris. In a pulsed system relatively crude hand tuning to about 5 c/s may suffice, but in a CW system the maintaining of receiver tuning to an accuracy of the order of 0·1 c/s becomes one of the major technical problems. Some automated method is the best solution, and is employed in the new Jodrell CW system. The only other possibility is to lock a closed loop system on to the signal as has been done by J.P.L., but this is difficult at the low RF signal-to-noise levels expected.

Narrow bandwidths

CW systems need to exploit bandwidths of a few cycles or less and to make meaningful spectral measurements within them. Such bandwidths are difficult to achieve stably by conventional means. Digital methods however are potentially limited only by the frequency stability of the system as a whole and may be applied in two ways, either by recording the data (the receiver IF noise) and later making a frequency analysis of it in a computer, or by analysing the data digitally in real time—as is done in the present Jodrell equipment. Another approach is to make use of the fact that the power spectrum is the Fourier transform of the auto-correlation function. This method was successfully applied in real time by J.P.L. in their recent Venus radar experiment, using a special digital computer fed directly from the receiver.

INTEGRATION TECHNIQUES

As in radio astronomy, so in radar astronomy, signals well below noise at RF must be detected. The usual way of doing this in radio astronomy is to compare the result of integrating signal plus noise for a given time with the result of integrating noise alone for the same time. The way of getting these two levels may

be simply by pointing the aerial successively on and off target, or by substituting a dummy load for the aerial (Dicke system). In radar astronomy two more ways of obtaining a comparison suggest themselves. In pulsed systems regions of the time base where the target is not present may be compared with the region where it is. This was the method used in the first Jodrell equipment (Evans & Taylor, 1959). Here a digital form of the detected receiver noise is selected from eight adjacent portions of the time base and fed into counting channels (which may be regarded as infinite time constants). As time goes on both the integrated signal power and the integrated noise power increase in direct proportion to time. The ripple on both increases in absolute size only as the square root of the time. Hence the signal to ripple—which is what counts in detecting a signal on top of the noise—improves as the square root of the integration time. Roughly, the summation of 100 independent echoes in this system gives 10 db, 10,000 gains 20 db and so on. At 1 pps, about 20 db improvement over the RF signal-to-noise can thus be obtained in a day's working.

In CW systems comparison can be made in frequency; this is done in the new Jodrell equipment. Here the echo is expected to occupy several channels of frequency whilst the others provide a reference level. In CW systems the improvement—as in radio astronomy receivers—is given by

$$\sqrt{\text{time of integration} \times \text{RF bandwidth}}$$

so again 10,000 'samples' and an improvement of 20 db can be achieved in about a day's working with a 1 c/s bandwidth.

HISTORY OF ATTEMPTS TO OBTAIN VENUS ECHOES

The most obvious parameter to use to compare results of Venus radar work is the value of the astronomical unit determined from a reduction of range data. Astronomers know well (to the order of 1 in 10^6) the movements and relative distances of the planets in terms of the A.U., but the value of the A.U. itself in kilometres is not known so well—only to 1 in 10^3 or 10^4. There is also an unsatisfactory spread of results beyond their limits of error (de Vaucouleurs, 1961). By radar the range—a time delay of about 300 seconds—need only be measured to 30 ms to give 1 in 10^4.

Pioneers in this field were the workers at the Lincoln Laboratories of M.I.T. Observations were made at the close approach in February 1958, and after an extremely sophisticated computer analysis a positive result was announced together with a result for the A.U. (Price *et al.*, 1959). At the next close approach in September 1959 an attempt was made at Lincoln to repeat the work. No significant echoes could be achieved on this occasion, and after exhaustive analysis the workers felt doubt also about the validity of the previous result (Pettengill and Price, 1961). At Jodrell Bank on the same occasion an attempt was made to obtain echoes. The choice was made to examine values of range as predicted from the 1958 Lincoln result. After 58 hours of useful observation, corresponding to the integration of about 100,000 echoes, a statistically significant rise was observed, in a range channel giving good agreement with the American value. It was noted however in the publication (Evans and Taylor, 1959) of the result that there was an 8 per cent chance that noise alone produced the rise interpreted as an echo.

It was generally felt that these two radar results for the A.U. were interesting but rather inconclusive, and preparations were made in several parts of the world to obtain more conclusive echoes at the close approach of Venus in April 1961.

In the event five groups obtained echoes, three in the U.S.A., Jodrell, and one in the U.S.S.R. All five finally agreed to within 2200 km on a value of the A.U. which differed substantially from the old radar value, and also from the best estimate of the optical astronomers. These results are shown in Table 3/II. The Russian result was at first ambiguous owing to the use of a high P.R.F. of 256 ms. and as announced by Tass (1961) was in agreement with the old rather than the new western values. However, successive revisions (Kotelnikov, 1961*a*, 1961*b*) have now brought it into good agreement. All workers used pulsed systems except J.P.L. who pioneered CW with great success, being the first to obtain echoes by some weeks. Range information was obtained by modulation, to give the A.U., which was also measured independently by Doppler methods.

The radar results for the A.U. thus exhibit strong agreement and small errors of the order of 1 in 10^5 but seriously disagree with the optical methods. The agreement between results on widely different frequencies would seem to rule out any significant effect of the interplanetary medium on the velocity of propagation. However the disagreement between the radar results and the mean of the optical methods of more than 60,000 km indicates a serious systematic error somewhere.

TABLE 3/II

April 1961 determinations of the Astronomical Unit

Place	Reference	Frequency (Mc/s)	Power (kw)	A.U. in km
J.P.L.	Muhleman *et al.* (1962)	2388	13 CW	149,598,845 ± 250
M.I.T.	Pettengill *et al.* (1962)	440	2500 peak	149,597,850 ± 400
Jodrell Bank	Thomson *et al.* (1961)	408	60 peak	149,600,000 ± 5000
U.S.S.R.	Kotelnikov (1961*b*)	700	*	149,599,500 ± 800
Moorestown (R.C.A.)	Maron *et al.* (1961)	—	—	149,596,000 ± (not given)

Note: The best value from optical astronomical methods given by de Vaucouleurs 1961 is

149,536,000 ± 3000 km

* 'Power flux density was 250 megawatts/steradian.'

Turning now to spectral results, agreement is not nearly so good. When J.P.L. first observed the return, it was found to have a width of only a few cycles. Both J.P.L. and Lincoln Labs were able to produce good spectra, J.P.L. recording 8 c/s at 2388 Mc/s and Lincoln about ½ c/s at 440 Mc/s. These results reduced to the same frequency do not agree to a factor of about two. The Russians however observed a width of more than 400 c/s at 700 Mc/s—in strong disagreement. Jodrell was unable to measure a spectrum.

The most likely interpretation of the American results is that Venus always turns the same face to the Sun.[1] Pettengill (1962) has shown that if this is so, the apparent rotation rate, and hence also the spectral width, should show a marked peak at conjunction, rising then to about three times the value a month before or after. This effect has not been observed: J.P.L. actually reports a definite stability of the spectral width throughout the observations. Kotelnikov (1961*a*) interprets the Russian results[1] as implying a rotation period of about ten days, and

[1] But see Postscript 1, p. 43.

observes variations of strength and frequency of the reflected echo as differently reflecting areas are carried round by the planet's rotation.

As well as these spectral results, the Lincoln workers have been able to measure the response of Venus to short pulses (Pettengill, 1962). Using 4 ms pulses no evidence of lengthening was detected in the return and using 0·5 ms a barely detectable lengthening was seen. This is strong evidence that Venus does indeed reflect rather like the Moon, most of the power coming from a small region in the centre of the disk, though the Lincoln results show this to be relatively smaller and that the power returned from the limbs is less.

Another interesting discovery made by both the Lincoln and J.P.L. workers is that Venus detectably does not follow the best available ephemeris. Thus not only the measurement of the A.U., but the study of orbital motion, is made possible.

CONCLUSIONS

The observations of Venus in 1961 have greatly increased knowledge of several aspects of solar system studies. Results of great interest concerning the value of the A.U. and the motion of the planet have been obtained. A good start has been made to the radar study of Venus in similar terms to that of the Moon; first results, particularly on the rotation period, are not in good agreement.

One can look forward confidently to further work on Venus, and to the extension of similar measurements to the other inner planets. Detection of more members of the solar system by these techniques will largely depend on the effort and resources devoted to the work in future years.

Postscript added in proof (*January 1963*)

In a recent communication Kotelnikov states that further spectral analysis of the recordings made during the Venus observations in 1961 has revealed the existence of a narrow feature in the centre of the spectrum of the returned echoes. This is less than 4 c/s wide, compared with 400 c/s for the broad component. The two components contain about equal energy. It is now believed that it is not possible to determine the rotation rate from the available data.

At Jodrell in November/December 1962, Venus was successfully detected with the new CW system. First results confirm both the value of the astronomical unit, from the measured doppler, and the extreme narrowness of the returned spectrum, of the order of $\pm\frac{1}{2}$ c/s at 410 Mc/s.

At the same time echoes were obtained by J.P.L. Preliminary results (N.A.S.A. release No. 62-275) are that Venus may be rotating slowly, with a period of perhaps 250 days in a *retrograde* direction. This appears to have been determined by the use of both CW and coherent pulses, and by observing the changing spectral width over the period of the observations, as briefly discussed above. Confirmation of this was obtained by observing the apparent motion of a singularity of the radar echo, which was tentatively identified with a surface feature.

REFERENCES

Evans, J. V., and Taylor, G. N., 1959, *Nature*, **184**, 1358.
Hughes, V. A., 1962, *Proc. Phys. Soc.* (in the press).
Kotelnikov, V. A., 1961*a*, *J. Brit. I.R.E.*, **22**, 293.
Kotelnikov, V. A., 1961*b*, Paper presented at 12th International Astronautical Congress, Washington, D.C.
Maron, I., Luchak, G., and Blitzstein, W., 1961, *Science*, **134**, 1419.
Muhleman, D. O., Holdridge, D. B., and Block, N., 1962, *Astron. J.*, **67**, 191.
Pettengill, G. H., 1960, *Proc. I.R.E.*, **48**, 933.
Pettengill, G. H., and Price, R., 1961, *Plan. and Space Sci.*, **5**, 70.
Pettengill, G. H., 1962, Report presented at Third International Space Science Symposium of COSPAR, Washington, D.C.
Pettengill, G. H., *et al.*, 1962, *Astron. J.*, **67**, 181.
Price, R., *et al.*, 1959, *Science*, **129**, 751.
Tass announcement, May 1961.
Thomson, J. H., Ponsonby, J. E. B., Taylor, G. N., and Roger, R. S., 1961, *Nature*, **190**, 519.
de Vaucouleurs, G., 1961, N.A.S.A. Memorandum R.M.-2944-N.A.S.A.
Victor, W. K., Stevens, R., and Golomb, S. W., 1961, J.P.L. Technical Report, No. 32–132.

4

PRESENT-DAY TECHNIQUES IN RADIO ASTRONOMY

E. J. BLUM

In this paper we shall discuss the idea of radio telescopes as a whole from the point of view of physics and astronomy, the information we can get from them and the sources of error. It seems to me that the greatest progress in this field recently has been in gaining a deeper insight into the nature of the measurements themselves. I shall illustrate my points by referring to actual or projected instruments.

VERY LOW NOISE RECEIVERS

You have all heard of the principle of masers and parametric amplifiers, with which overall noise temperatures as low as a few tens of degrees have been obtained. Up to now this type of instrument has posed difficult problems of maintenance. The gain is not very stable and the price is very high. Finally, the majority of present-day aerials introduce a ground noise of several tens of degrees making the use of the very low noise maser pointless. It is certain that we can expect in a very short time the appearance of radio telescopes with masers conceived as a unit. Their use on centimetre waves will perhaps be limited in part by troubles of tropospheric origin, but on decimetre waves, where the sky temperature is low, they will be extremely useful.

The receiver temperatures which can be attained with parametric front-ends are much higher and present-day aerials are very suitable. As in the case of masers, stability is difficult to obtain. It must be noted, however, that the necessary gain stability in an instrument is independent of receiver noise temperature and has a limit fixed by the equation $\sqrt{\frac{1}{B\tau}}$ (B being the bandwidth and τ the time constant). In fact, the smallest

detectable signal is $T = \frac{T_R}{\sqrt{B\tau}}$ (T_R is the receiver temperature) and in order to measure it with a direct receiver it is sufficient that $\frac{\Delta g}{g}$ should be equal to $\frac{\Delta T}{T_R}$, that is to say, to $\frac{1}{\sqrt{B\tau}}$.

In the same manner, for a correlation receiver or a receiver using a Dicke switch, the gain stability requirement is less stringent. The useful stabilities are therefore not greater than those which can be obtained with present-day receivers. Thus with the use of correlation receivers we can hope to obtain the theoretical sensitivity for some considerable time. In many places studies are in progress to develop parametric amplifiers especially suitable for radio astronomy. Diode systems are the most common but attempts are also being made to use Adler tubes. For the moment, the temperatures of the systems we have discussed are of the order of 200°K. It appears probable that this will be lowered quite rapidly to 100°. The diodes themselves are in progress of rapid evolution and some of them already allow of a theoretical receiver temperature of the order of 50°K.

COLD AERIALS

For a long time aerials were designed to obtain a high forward gain with a reduction of the first side lobes to acceptable limits. But it is the contribution of the complete set of far side lobes which determine the effect of the ground on the temperature of the aerial. In the case of a parabola illuminated by a horn, for example, the ground radiation gives a contribution of approximately 50°K to the aerial temperature.

In order to improve this situation we can endeavour to minimize the side lobes of the parabola by choosing a rapidly decreasing illumination, but then the width of the principal lobe is increased and the output is lowered for the observation of small diameter sources. Besides, the primary horn remains responsible for an important part of these diffused side lobes. With a reasonable compromise we can bring down the ground contribution with such a procedure to about 30°K.

At the present moment two procedures have been employed to improve this situation. In one the parabola is replaced by a horn of large aperture but of acceptable length—this is the

Hogg horn. In the other a second reflecting element is added with the primary horn at the centre of the principal reflector. This is the well-known optical Cassegrain system. We can then hope to reduce the contribution from the ground to ten degrees or so when the aerial is pointing towards the zenith as there is a great reduction in the backward lobe. But if the aerial is pointed to an average elevation the ground contributes more to the aerial temperature.

By using this type of aerial with very rapidly decreasing illumination such as we have just mentioned we can hope to gain still further, but the forward gain diminishes. The choice is very difficult. It depends on the projected measurement and to my knowledge no satisfactory criterion has been determined up to the present moment. We must not however consider that a conventional aerial is such a bad thing. The radio telescope at Nançay consists of a plain meridian mirror of variable declination which reflects the energy received from the sky on to a convergent mirror. The rays are close to the ground for several hundreds of metres. Nevertheless, we have in the course of preliminary measurements noted a ground temperature varying between approximately 60° and 40° according to the declination.

FURTHER DEVELOPMENT OF BASIC TECHNIQUES

We can increase the receiver bandwith and thus gain in sensitivity by a factor $\sqrt{B}$. In the United States a receiver has been constructed with a bandwidth of 1000 Mc/s. I think that this method has very little future for the radio spectrum is already extremely crowded and this state of affairs can only get worse. On the other hand, radio astronomy can hope in the near future for the efficient protection of some frequency allocations. If we wish to go further than this it is to risk more and more interference and the increase of sensitivity has every chance of being purely theoretical.

The development of transistor devices appears on the other hand of great interest for radio astronomy. The installation and maintenance of great arrays of aerials with separate amplifiers is much simpler than with valves. In the scheme for the Benelux cross project in particular, C. Murray has developed transistor preamplifiers of which the performance is analogous to the best existing valve systems. We are thinking for the next solar cycle

of modifying the multiple element interferometers which we have at Nançay in order to adapt them for use with transistorized preamplifiers.

TYPES OF RADIO TELESCOPE AND SPATIAL BANDWIDTH

i. *General remarks.* I would like to remind you of the idea of using spatial frequencies. This method was developed for radio astronomy some years ago simultaneously by Arsac and Bracewell.

An aerial is a linear system which gives a response proportional to a signal received from space. It can be considered therefore as a filter, obviously linear, which transmits a signal composed of spatial frequencies with a certain gain and bandwidth.

Normally we study the polar diagram of an aerial, that is to say, its response to a point source at infinity. This is the equivalent to the response of a circuit to an impulse. As for the circuit, we can define a bandpass which is the Fourier transform of the response to a point source at infinity.

For example, the polar diagram of a uniformly illuminated square aperture of side a is of the form $\left(\frac{\sin au}{au}\right)^2$ in each dimension. Its spatial passband, the Fourier transform of this expression, is a triangle. The highest frequency transmitted is equal to $\frac{a}{\lambda}$ and is related more fundamentally than the beamwidth to the geometry of the aerial system.

ii. *Total power instruments.* The detector produces the square of the voltage appearing at the aerial terminals. The spatial passband always has an absolute maximum at the origin. Sensitive to a continuous background, these instruments give at the same time a continuous component due to the receiver noise. In switching systems the latter is eliminated at the cost of a loss of sensitivity. In direct systems the receiver is stabilized sufficiently to determine the continuous level by regular checks.

iii. *Correlation (or phase switching) instruments* include in particular all true interferometers. On the other hand we must class amongst total power instruments interferometers comprising a single receiver joined to the terminals of two aerials working in parallel. The possible combinations in correlation

apparatus are extremely numerous. Since the Mills cross, complex antenna of one or two dimensions are now seen: crosses of aerial arrays, spaced aerial arrays, aperture synthesis, etc. All these instruments are insensitive to the continuous component of the sky or receivers, i.e. the gain is equal to zero at zero spatial frequency. The shape of the bandpass in general decreases less smoothly than for total power instruments towards the upper frequencies. Put another way, the side lobes are larger for interferometers.

iv. *Partial coherence instruments* have not yet been described. At the present moment we are studying this at Meudon. They form an intermediary class between the two preceding groups. The zero frequency component no longer has a privileged role but otherwise the general shape of the spatial bandpass is similar to other instruments. The noise of the receivers can give a negligible continuous component at the cost of a few precautions.

We can think of these instruments *either* as being derived from total power systems, *or* as being derived from correlation apparatus. As an example of this system, between the sky and the two elements of a correlating interferometer one places a diffracting screen. The plane wave front coming from a source at infinity is diffracted but there is coherence between the diffracted waves received by the two aerials.

More practically we can place at the focus of a parabola and in the first maximum of its diffraction pattern a system of concentric horns. The coupling between these horns is extremely weak and therefore the background component due to receiver noise is small. Nevertheless, the shape of the spatial bandpass according to calculations can be very near that of a total power system.

We have experimentally verified the possibility of employing this type of apparatus and we expect to use it on the great radio telescope at Nançay for measurements on 21 cm.

THE AVAILABLE INFORMATION AND ITS USE

i. *Instruments with a single aerial*

It is not customary in optics to limit oneself to use the diffraction pattern on the axis of an instrument. Measurements are always made in a great number of directions close to the axis.

Now in radio astronomy the opposite is true at the present time. There are several reasons for this; for a long time, small instruments were used, of which the price was of the same order as that of the receiver. It was a simple matter to reconstruct the instrument completely, and even to use it to collect all the available information. For mechanical reasons the focal ratio, f/d, of radio telescopes is generally less than unity and coma aberration limits the field considerably. We think that this situation will rapidly change. For instance, the Cassegrain system allows a simple lengthening of the focal length and therefore a greater field. On the other hand, in order to explore all the sky and even limited regions with a resolving power of a few minutes would become impossibly long. To make simultaneous measurements on several declinations might be easy and useful. Large aerials are extremely costly when compared to receivers. We can then envisage the employment of several of the latter.

To conclude, for reflecting telescopes the available information is fixed by the field, itself determined by the aberrations. For maximum speed of observation it is necessary to have as many receiving horns as there are beam solid angles in the field.

ii. *Instruments using several aerials*

The information from each aerial is coupled with a certain phase. Let us take, for example, an interferometer with two aerials EW pointing to the meridian. Usually, we are content to record the amplitude of the fringes with a single phase setting; according to the sampling theory it is necessary to record simultaneously with two phase settings to obtain almost all the available information.

This procedure has been used at Nançay in the NS network, using eight aerials functioning on 169 mc/s. If all the aerials are in phase the lobe is on the axis. By displacing the central aerials by $\lambda/16$, the two following by $\lambda/8$ and the extreme ones by $\lambda/4$, we can shift the pointing direction by half a lobe and this operation may be repeated sixteen times. Sixteen detectors thus record simultaneously the signals from sixteen lobes. This system is used to advantage for solar measurements.

The authors of the Benelux cross project are thinking of using multiple phase settings for their radio telescopes, analogous to

those which we have just described, and J. P. Wild likewise for his radio heliograph.

iii. *Aperture synthesis*

This type of interferometer using two aerials at variable separations has for each spacing a different spatial passband. The collection of results is recombined by a digital Fourier transform in order to give an image of the observed regions. It is an extremely powerful method but somewhat difficult in practice, for the restoration necessitates a knowledge not only of the amplitude but also of the phase of the observed signal.

iv. *Limits set by local conditions*

Even with a perfect instrument it may happen, at least for very great separations, that accidental out of phase components due to the ionosphere or to the troposphere will occur. Recent observations show that apparent displacements reaching in extreme cases a few minutes of arc take place on metre wavelengths. Commonly, refractions of the order of a minute are produced. On decimetre waves the situation is perhaps better, but no conclusion can be drawn according to present-day observations.

For the moment, we are sure that resolving powers of the order of 10″ can be achieved for a great part of the time, but we cannot exclude that the terrestrial environment determines, as in optics, the greatest aperture of the apparatus beyond which the resolving power will no longer increase.

THE PROBLEM OF MEASUREMENT—IMAGERY AND RESTORATION

i. *Isolated regions*

The Sun is the principal representative of this category. Here it is a question of an isolated object and one limited in space from which the space signal has no component below a certain frequency. We know that it is enough in order to study them to make observations on multiples of a minimum spatial frequency. This is the principle of arrays of multiple aerials regularly spaced out, used in the first place by Christiansen. In general, the illumination of these arrays is chosen in order to have very weak side lobes but this is not an absolute necessity. In fact, the image given by an instrument possessing a certain

passband with an appreciable gain on the high frequencies, that is to say, with important side lobes, can always be restored by a numerical convolution containing a function (amounting to artificial filtering). In this manner we can obtain an image closer to the idea that we have in mind, but the intermediate image contains more information on the higher frequencies and can give more precise information on position.

Recent very high resolution measurements on a few radio sources use the same method but the observations are sequential instead of being simultaneous. An interferometer with two reflectors is employed at variable distances. The space band of the instrument, which can be called synthetic, has a rectangular envelope and the restoration method will allow of eventual restoration of the image. This very powerful procedure has been applied here by Palmer, and almost simultaneously with less resolving power but closer points at Cal. Tech. and at Nançay.

ii. *Point sources*

If we know *a priori* that a point source is isolated, then the two-aerial interferometer gives the best accuracy on its position, or better still, three aerials in order to remove the ambiguity.

Unfortunately, this ideal case is rare and very often we are concerned with other relatively close sources which can modify the fringe pattern. We approach therefore the question of confusion which is very far from being treated in a satisfactory manner.

iii. *The establishment of a complete sky map*

In this case, it is highly necessary to have an instrument with a continuous space band with no zeros except at frequencies close to zero on the one hand, yet sufficiently sensitive at high frequencies in order to obtain the desired resolving power on the other hand. Further, we have little information *a priori* on the spectrum of the space signal. Finally, the image obtained by filtering the signal in the bandwidth of the instrument will not reproduce the real distribution of luminosity in the sky except with certain attendant errors. Different optimizations are possible. We can try to obtain the average minimal quadratic error, then the best spatial passband is rectangular. This method is not often used for it gives large side lobes, and as a few very bright sources exist in the sky their contribution extends over

some distance. Other criteria can be employed, i.e. minimal error in temperature or position measurements.

Finally, the determination of the spatial bandwidth of an instrument is a compromise, and I again insist on the fact that this compromise must take account as much of the criterion of accuracy as of the type of objects observed.

Problems associated with measurement

i. Confusion appears in fact in the idea of error which we have just described. Here it is a question of the appearance in the record of fictitious sources due, for example, to the simultaneous presence of several weak sources in the side lobes in phase.

To avoid this, it is admitted for the moment that it is necessary to have, according to the aerial, between 20 and 100 beam areas for each detected source but these figures have not a very solid basis. Once more we must bring in the statistics of sources as well as the characteristics of the instrument. The only systematic study of this question has been made here at Jodrell Bank by Hazard who clearly showed the inferiority of interferometers when compared to total power instruments from this point of view.

ii. The exposure time is determined most frequently by the necessary sensitivity. In particular, for objects with low brightness temperatures (e.g. external galaxies), the integration time even with large radio telescopes and the best of present-day receivers still remains relatively long. To make a complete survey with an instrument having a resolving power of 1 minute requires 20 years if we allow the sky to cross in front of a system employing a single lobe. Interest in instruments using multiple lobes is then obvious.

Finally, this question has no meaning except when the observations are not jammed. The allocation of frequencies for radio astronomy appears more vital each day with the increase in sensitivity, in interference and in the number of transmitters.

CONCLUSION

We have endeavoured during this sketch to show that radio astronomers are now beginning to know how to choose the best instrument for a specific measurement, or can, with an existing apparatus, prejudge the value of any observations.

During these few days the astronomical questions which are being asked at the present day will be examined here. You will see that they are very varied and it is very unlikely that one radio telescope can be employed to resolve many problems.

Nevertheless, we must not go too far. You have here at Jodrell Bank the example of a radio telescope used to advantage for high resolution measurements, for survey work and for planetary radar studies. It is certain that the very great instruments of the future must also be suitable for a variety of researches.

5

RADIO ASTRONOMY FROM SPACE VEHICLES

R. C. JENNISON

In recent years man's knowledge of the universe has been very considerably extended through observations in the 'window' of the electromagnetic spectrum whereby radio waves in the range from about 15 to 15,000 Mc/s may propagate through the upper atmosphere to aerials situated on the ground. Above 15,000 Mc/s the molecular absorption of the Earth's lower atmosphere renders observation of extra-terrestrial radiation extremely difficult, whilst below 15 Mc/s the Earth's ionosphere and terrestrial interference limit observation at the lower end of the spectrum.

THE MILLIMETRE SPECTRUM

The difficulties above 15,000 Mc/s could be avoided by mounting equipment in satellites or very high flying balloons, but other problems would then arise in the techniques of mounting and steering highly accurate aerial systems. At these very high frequencies it is likely that the sky will appear as a system of thermal sources, the regions of hot ionized gas in such objects as the Orion nebula standing out more than the non-thermal radio sources which are familiar at metre wavelengths. The Sun would be expected to appear as a reasonably well behaved hot body at a temperature of about 6000°K, whilst the planets would also radiate with temperatures close to their thermal temperatures; this is indicated by Table 5/I which lists the temperatures of three of the planets determined from the surface of the Earth by various observers at centimetre wavelengths. It may be noted that Venus and Jupiter are slightly anomalous in that their temperatures differ from those determined at infra-red wavelengths. In the case of Jupiter the apparent temperature increases at decimetre wavelengths to about 100,000°K but is likely to be normal in the millimetre band.

TABLE 5/I

The temperature of the planets at centimetre wavelengths

Planet	Wavelength (cm)	Temperature (°K)
Venus	0·86	410 ± 160
	3·15	560 ± 73
	3·4	518 ± 47
	3·37	575 ± 58
	9·4	580 ± 160
Jupiter	3·15	145 ± 26
	3·2	177 ± 22
	3·75	210 ± ?
	10·3	640 ± 85
Mars	3·14	211 ± 28
	3·15	218 ± 76

The planet Venus is permanently covered by a complete layer of cloud which effectively prevents optical observation of either its surface or lower atmosphere. It is not known whether water vapour and oxygen are present in this atmosphere. One of the most useful experiments which might therefore be performed from a satellite would be to fly a spectrometer covering the water vapour and oxygen bands at $1\frac{1}{2}$ cm and 5 mm respectively, and to search for these substances in absorption against the radiation from the surface of the planet.

This experiment presents a considerable technical difficulty in the deployment of a suitable aerial. To compute the minimum size of this aerial let us assume that the temperature of Venus is 500°K and that the receiver has a noise temperature of 1000°K. The spectrometer will have to operate without manual adjustment, and as we shall be searching for changes in the temperature with frequency relative to the aerial temperature of the planet, we shall require an aerial temperature from Venus of, say, 50°K to give a usable spectrum. This, in turn, implies a beam-width of the order of $\sqrt{500/50}$ times the diameter of Venus. As the latter is about one minute of arc the aerial beam required is about 3′ arc or 1/1000 radian, requiring an aperture of 1000 wavelengths or 5 metres for a wavelength of 5 mm. This aperture would require an accuracy equal to or better than 1 mm according to the design of the aerial. It might be possible to produce such an instrument by unfurling an aluminium foil

or coated mylar parabola. A horn aerial is attractive but would be excessively long and a slotted sandwich is probably preferable. By far the tidiest solution is to use a much smaller aerial mounted on a space probe travelling close to the planet; in these circumstances an aperture of a few centimetres would suffice.

EXPERIMENTS AT LONG AND MEDIUM WAVELENGTHS

Certain features of the radio sky at frequencies of the order of a few megacycles are to be expected by extrapolation from surveys at higher frequencies. It is likely that radiation from the corona of the galaxy may predominate whilst the galactic plane will appear as a dark absorption band. The well-known discrete radio sources will be insignificant, as well as the quiet Sun. It is not known if other features exist which do not appear on surveys at higher frequencies, and in this connection the accidental discovery of emissions from Jupiter at 20 Mc/s may be quoted as an example of the possibility of observing unexpected phenomena in the many unexplored octaves of lower frequency.

Reber and Ellis (1956) have endeavoured to perform surveys of the sky at medium wavelengths by ground observations through 'holes' in the ionosphere. The results from these experiments are not easy to interpret though they yield an order of magnitude for the total flux from the Galaxy.

It is clearly of importance to perform a survey of the sky at low frequencies with aerials of high resolving power and without the need for the rare, difficult and complicated propagation to ground stations through the earth's ionosphere.

A high gain aerial at frequencies of the order of 1 Mc/s would be an enormous structure many miles across and could not be deployed from a satellite in the foreseeable future. Interferometer systems between parts of a satellite connected by long wires and between separate satellites can be conceived, but the separation would have to be many miles before fine beams could be achieved. Problems ensue in controlling the orientation of such systems during the lifetime of the satellite.

The situation appears very different when consideration is given to the effect of placing the aerial in the region of decreasing electron density above the F layer of the ionosphere. Even a

short dipole may be shown to acquire a high gain and narrow beamwidth in these circumstances.

THE DIRECTIVITY PATTERN OF AN AERIAL ABOVE THE IONOSPHERE

The gradual decrease in the electron density modifies the directivity pattern of an aerial immersed in the medium. This situation was first investigated by Haselgrove, Haselgrove and Jennison (1961*a*) who showed that the beam pattern is focused into the direction of the vertical and is modulated by fine interference fringes which increase in amplitude towards the edge of the pattern where the termination of the beam is extremely

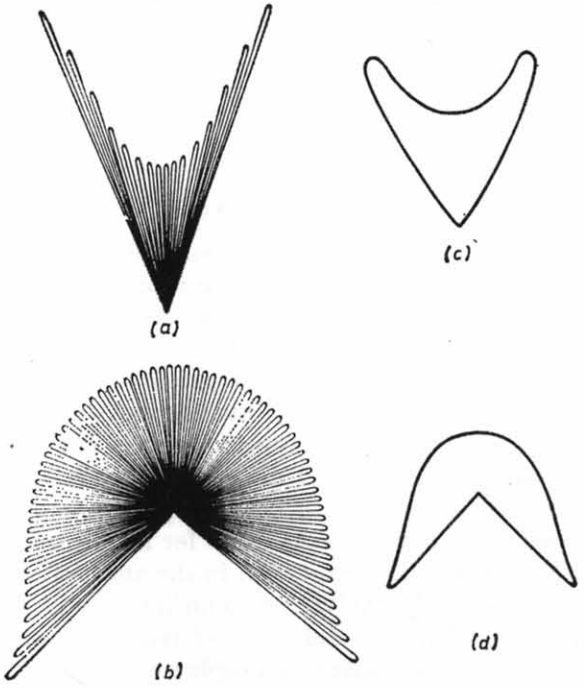


FIG. 5/1.

(*a*) The directivity pattern corresponding to the conditions in Fig. 5/2(*a*).
(*b*) The directivity pattern corresponding to the conditions in Fig. 5/2(*b*).
(*c*) The effect of a finite bandwidth upon the directivity pattern of Fig. 5/1(*a*).
(*d*) The effect of a finite bandwidth upon the directivity pattern of Fig. 5/1(*b*).

sharp, as shown in Fig. 5/1 (*a*) and (*b*). The above paper indicated how the variations of signal strength could be obtained for any orbit or ionosphere by digital computation; it originally contained a number of errors and obscurities which have since been amended (Haselgrove *et al.*, 1961*b*). A more rigorous wave treatment including an investigation of the modification of the directivity pattern of an aerial inclined at various angles above the *F* layer has since been performed by Budden (1961).

These papers indicate that if radiation is emitted or received at an angle θ to the vertical by an aerial in the ionosphere, the corresponding angle far beyond the ionosphere will in general be different due to refraction. If this second angle is θ' then, assuming only that the ionosphere has symmetry about the vertical axis through the satellite, the relative power gain of the aerial at angle θ is given by

$$\frac{\sin\theta\, d\theta}{\sin\theta'\, d\theta'}$$

which may be written

$$\frac{\cos\theta'}{n\sqrt{(n^2 - \sin^2\theta')}}$$

where n is the refractive index of the ionosphere at the satellite.

If, as an approximation, the Earth is considered to be flat, the aerial beam is confined to a cone whose semi-angle θ'_0 is determined solely by the refractive index in the vicinity of the satellite, the semi-angle θ'_0 being given by $\sin\theta'_0 = n$. As shown in Fig. 5/2 (*a*) the radiation at the edge of the cone is that associated with the horizontal radiation at the aerial. This approximation is reasonable for an aerial deeply immersed in the ionosphere and operating on a frequency just above the plasma frequency but it breaks down for higher frequencies or more distant orbits. The limitation in the angle of radiation or reception is naturally equivalent to an increase in the overall gain of the aerial and useful gains of ten or twenty may be obtained in this way before ionospheric irregularities cause serious distortion.

The average gain of an isotropic aerial when immersed in the ionosphere (assumed to be flat and plain stratified) is given by the ratio of the solid angle of the reception cone of the combined aerial-ionosphere system to the solid angle, 4π, of an isotropic system in free space.

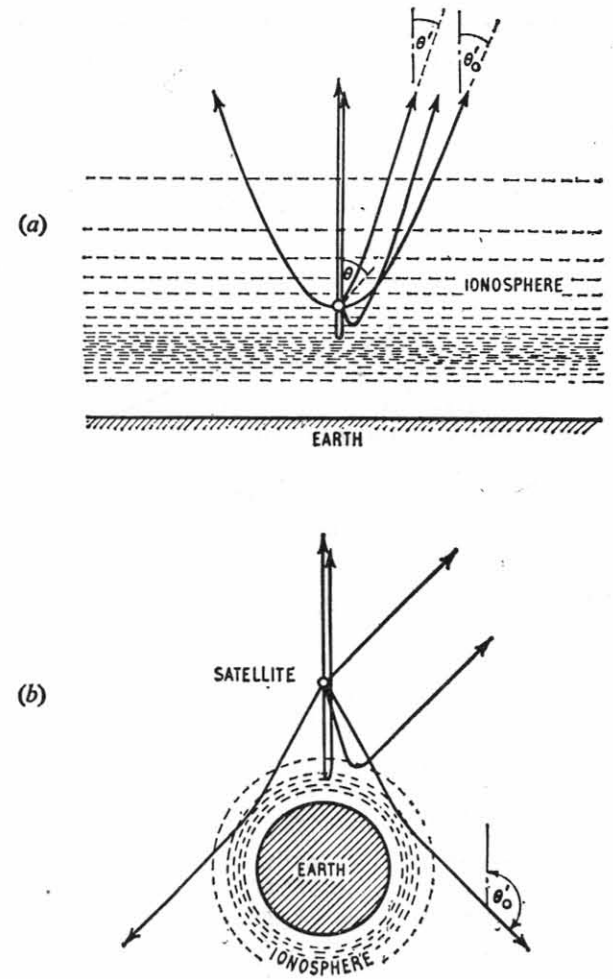


FIG. 5/2.

(*a*) Ray paths to a satellite aerial slightly above the maximum of the *F* layer when the receiver is tuned to a frequency slightly higher than the ambient plasma frequency (flat Earth and plane stratified ionosphere approximation). A ray leaving the satellite at angle θ' emerges at an angle θ' at infinity. The limiting ray emerging at θ_0' leaves the satellite horizontally ($\theta_0 = 90°$).

(*b*) The corresponding ray paths when the satellite is at a greater distance from the Earth and the flat Earth and plane stratified ionosphere approximations are no longer valid. The limiting ray emerging at θ_0' no longer meets the satellite horizontally.

Thus if θ'_0 is the semi-angle of the reception cone when the aerial is immersed in the ionosphere the average gain

$$G_{\text{av.}} = \frac{2}{1 - \cos \theta'_0}$$

but the refractive index $n_0 = \sqrt{(1 - f_0^2/f^2)}$ where f_0 is the ambient plasma frequency and f is the operating frequency, whence we obtain $\cos \theta'_0 = f_0/f$ and

$$G_{\text{av.}} = \frac{2}{1 - f_0/f}$$

Typical values of f_0/f, θ'_0 and $G_{\text{av.}}$ are given in Table 5/II which is similar to Table I in the original paper by Jennison (1961) but also lists the forward gain in the last column.

TABLE 5/II

$f_0/f = \cos \theta'_0$	$n = \sqrt{\left(1 - \frac{f_0^2}{f^2}\right)}$	$\theta'_0 = \sin^{-1} n$	Average gain $= \frac{2}{1 - f_0/f}$	Forward gain $= \frac{1}{n^2}$
1	0	0		
0·98	0·2	11° 31′	100	25
0·9	0·4359	25° 50′	20	5·26
0·8	0·6000	36° 50′	10	2·78
0·7	0·7141	45° 35′	6·7	1·96
0·6	0·8000	53° 10′	5	1·56
0·5	0·8660	60° 00′	4	1·33
0·4	0·9165	66° 25′	3·3	1·19
0·3	0·9539	72° 30′	2·85	1·10
0·2	0·9798	78° 30′	2·5	1·04
0·1	0·9950	84° 20′	2·2	1·01
0·0	1·0000	90° 00′	2·0	1

The forward gain may be obtained by substituting $\theta' = 0$ in the expression

$$\frac{\cos \theta}{n\sqrt{n^2 - \sin^2\theta}}$$

Hence the forward gain $= \frac{1}{n^2}$

This is listed in the last column of Table 5/II but it should be noted that it is not a good measure of the aerial efficiency as the

beam pattern is more intense at the edges. The criterion of forward gain is therefore inappropriate, as it is in the extreme case of a system formed by two antiphase apertures, in which, of course, the forward gain is zero.

The analysis by Haselgrove, Haselgrove and Jennison also showed that the aerial directivity pattern is modulated by interference fringes resulting from the two possible ray paths to the aerial, one direct and one reflected in the ionosphere beneath the satellite. Furthermore, the aerial gain tends to increase towards the edge of the pattern and the final edge is very sharply bounded. Ray treatments show an infinity at the boundary but Budden has shown that this is resolved by a wave analysis and the open circuit aerial voltage

$$V = 1K_2Ai(\zeta_0)$$

for broadside radiation from a horizontal dipole and

$$V = 1K_2Ai'(\zeta_0)$$

for end-fire radiation from a horizontal dipole where K_1 and K_2 are slowly varying functions involving the refractive index, the gradient of refractive index and the angle θ. 1 is the length of the aerial, Ai is an Airy integral and Ai' its vertical derivative.

If the aerial is at a large distance from the Earth such that the electron density, and therefore the refractive index, is constant or only slowly varying with distance, the directivity pattern is no longer dictated by the conditions in the immediate vicinity of the satellite, apart from the primary directivity pattern of the aerial. In these circumstances the flat Earth approximations no longer hold, the rays defining the edge of the reception cone no longer meet the satellite horizontally and the cone semi-angle can exceed 90 degrees. The Earth and its ionosphere still affect the directivity pattern as rays may still be reflected and refracted at the appropriate levels as in Fig. 5/2 (*b*), so that the pattern retains its fringe modulation and contains a sharply cut 'hole' or shadow in the direction of the Earth.

In Table 5/II a list is set out in which the semi-angle of the reception cone and the aerial gain are given as a function of the refractive index for a flat Earth and plain stratified ionosphere. In Table 5/III maximum reception angles are quoted for two sets of data on electron densities at large distances from the Earth when a fuller analysis has been used.

In each case a background electron density of 10^2 has been

TABLE 5/III

f(Mc/s)	2	3	4
θ'_0 degrees at 5000 km	123·7	128·1	130·6*
	124·2	126·1	127·4†
θ'_0 degrees at 10,000 km	141·7	144·7	146·5*
	144·0	145·0	145·6†

* Based on the distribution of electron density determined by Berning (1960).

† Based on the distribution of electron density determined by Al'pert *et al.* (1958).

used in extrapolation of the profiles. A background density of any value from zero to 10^3 produces negligible difference in the cone angles.

PRACTICAL SYSTEMS UTILIZING NON-MAGNETIC FOCUSING

The modification of the primary directivity pattern referred to in the previous section may be used in a number of practical systems to achieve a high resolving power at medium frequencies.

The principal properties of the combined aerial-ionosphere reception pattern are:

(i) The restriction of the radiation into a narrow cone when the refractive index approaches zero.

(ii) The modulation of the reception pattern by a system of interference fringes (approximately concentric circles in horizontal cross-section).

(iii) The very sharply bounded edge of the pattern preceded by an enhancement due to focusing at the limb.

Property (i) would enable gains and beamwidths at least comparable to those of Yagi aerials to be obtained at medium waves, but the beamwidth is very susceptible to variation in electron density and would continually change throughout the orbit of a satellite. It would be possible, in principle, to monitor the ambient electron density or plasma frequency and hence predict the cone angle. A more ambitious proposal which could only be applied to radiation having an approximately constant spectrum, as may well be the case for cosmic radiation, is to

control the frequency of the receiver by means of a servo system to keep it always operating at a frequency where the refractive index is a constant slightly in excess of zero. This system would preserve a constant beamwidth although the frequency of reception would be continuously varying. The frequency could be monitored and passed back, after reduction to a low data rate, by the telemetry system to give information on the electron density of the ionosphere along with the output of the receiver giving a crude sky survey on a further low data rate channel. The direct application of the overall focusing property has also been considered by Smith (1961).

Property (ii), the modulating fringe system, is only of use for CW transmission or reception, though for bandwidths up to a maximum of about 1 kc/s the fringes near the edge of the reception cone may still be distinct. The zero-order fringe is situated at the edge of the reception cone and whilst higher orders, further towards the centre, would rapidly integrate out over finite bandwidths, the position of the zero-order fringe itself moves and hence also smooths out the fringes to give a constant output with less sharp edges as a result of the change in the semi-angle of the cone with change in frequency. The fringe system appears, therefore, to be of little use to radio astronomy although it may be that it will find uses in the field of communications or for investigating the ionosphere itself.

(iii) The very sharply defined edge of the reception cone presents several possibilities for practical techniques. The intrinsic resolving power of this edge is typically a few minutes of arc. This figure would hold in practice if the system were monochromatic and the ionosphere were perfectly smooth. The effect of a finite bandwidth is to round off the edge to an extent which is simply determined by the variation in semi-angle of the reception cone over the same frequency range, as in Fig. 5/1 (*c*) and (*d*). If the satellite carrying the elementary aerial is a few thousand kilometres from the Earth and the operating frequency is about $2\frac{1}{2}$ to 3 Mc/s, the effect of a bandwidth of 100 kc/s is to reduce the resolving power to about 20 minutes of arc, whilst a bandwidth of 1 Mc/s would still give a resolving power of the order of three degrees.

The effect of ionospheric irregularities is not easy to assess as very little is known about the small-scale distribution of electron density above 500 km. It is probable that if the

operating frequency is sufficiently high and the distance from the Earth sufficiently great the effect of irregularities would not be troublesome. Even under more adverse conditions the possible effect of irregularities does not preclude the use of this principle in a satellite technique as the obtainable resolution would be very high and small sources, if such exist, might make their presence known by the production of scintillations in the received signal.

Two techniques have been proposed which make use of the sharp edge of the reception pattern (Jennison, 1961). In the first of these, the output of the receiver after detection is passed through a bandpass filter with a lower frequency limit of several minutes per cycle and an upper frequency limit adjusted to suit the fastest rate of change of signal that could be caused by radiation from a point source entering the edge of the reception cone as the satellite orbits. This frequency is a function of orbit and bandwidth and may lie between one cycle per second and one cycle per minute. The application of the bandpass filter removes the large background level of signal and the remaining higher Fourier components, corresponding to the edge of the reception pattern, may be further amplified prior to connection to a low data rate telemetry channel. The resolving power with this technique is determined by the bandwidth and the high pass limit of the filter whilst the large bandwidth and time constants reduce the background noise ripple and enable a high sensitivity to be obtained. The direct output of the detector, without differentiation or further amplification, may be monitored on a further low data rate channel.

The second proposal, making use of the sharp edge of the reception cone, is to commutate the receiver rapidly between two frequencies separated by a few tens or hundreds of kilocycles. The output of the detector is commutated in synchronism so that the reception in the two corresponding states is compared. The difference signal, corresponding for a flat spectrum to the signals from a narrow ring in the sky, is further amplified, and passed to the telemetry as a low data rate signal.

The latter technique could be combined with that employing the bandpass filter provided that a suitable switching rate were chosen, and both could independently feed separate telemetry inputs. Similarly the proposal making use of the narrow reception cone where the refractive index approaches zero,

could be incorporated in the same satellite by employing a separate receiver on a lower frequency.

THE USE OF *Z* MODE PROPAGATION

Ellis (1962) has suggested the use of Z mode propagation for the production of a pencil beam from an elementary aerial mounted on a satellite which is orbiting between the levels of critical frequency for the ordinary ray

$$\left(\text{defined by } f_0 = \sqrt{\frac{nNe^2}{\pi m^2}}\right)$$

and for the Z mode

$$\left(\text{defined by } f_z = \sqrt{\frac{f_H{}^2}{4} + f_0{}^2} - \frac{f_H}{2}, \quad \text{where } f_H = \frac{eH}{2\pi mc}\right).$$

In these circumstances rays can reach the satellite from the sky only over a narrow range of angle centred on the Z mode critical angle of incidence:

$$\theta_e = \sin^{-1}\left[\left(\frac{f_H}{f + f_H}\right)^{\frac{1}{2}} \sin\theta_0\right]$$

where θ_0 is the zenith angle of the geomagnetic field direction. The critical angle is determined by the condition that the wave normal of the incident rays must be aligned with the geomagnetic field direction at the plasma level defined by $f = f_0$. It has been shown that the angular diameter of the Z propagation hole is somewhat less than 1°. The hole is elongated in the direction of the magnetic meridian as a result of the second indirect Z ray, reflected by the ionosphere. Typical figures of $F_{\text{max.}} = 250$ km, operating frequency = 1 Mc/s and geomagnetic latitude = 50°, show that the satellite may be between 550 and 1200 km if $f_{\text{crit.}} = 3$ Mc/s and between 350 and 600 km if $f_{\text{crit.}} = 2$ Mc/s.

The antenna acquires a high effective angular resolution but does not have a correspondingly high gain. No details of the usable bandwidth or achievable signal to noise ratio are available in the reference.

REFERENCES

Al'pert, Ya. L., Chudesenko, E. F., and Shapiro, B. S., 1958, *Usp. Fiz., Nauk*, **65**, 2, 161.

Berning, W. W., 1960, *J. Geophys. Res.*, **65**, 9, 2589.
Budden, K. G., 1961, *Proc. Roy. Soc. A.*, **263.**
Ellis, G. R., 1962, *Nature*, **193**, 862.
Haselgrove, C. B., Haselgrove, J., and Jennison, R. C., 196*a*, *Proc. Roy. Soc.*, A, **261**, p. 423.
Haselgrove, C. B., Haselgrove, J., and Jennison, R. C., Corrigendum, 1961, *Proc. Roy. Soc. A.*, **261**, 424.
Jennison, R. C., 1961, *J. Brit. I.R.E.*, **22**, p. 205.
Reber, G., and Ellis, G. R., 1956, *J. Geophys. Res.*, **61**, p. 1.
Reber, G., 1958, *J. Geophys. Res.*, **63**, p. 869.
Smith, F. G., 1961, *M.N.R.A.S.*, **122**, 527.

6

THE USE OF COMPUTERS IN RADIO ASTRONOMY

J. G. DAVIES

ELECTRONIC computers can be applied to radio astronomy in a variety of ways. We immediately think of the calculation of ephemerides and co-ordinate conversion tables, which tell us in which direction to point our telescope to study a particular object. This is, however, a fairly straightforward way of using a computer, and I shall not discuss it further. To go a stage further, the movements of a telescope might be under the direct control of a computer.

The second major branch of the use of digital techniques in radio astronomy is in *data logging*. This is the application of automatic techniques to by-pass the recording of data on charts, and the later reading of these charts in numerical form, punching them out by hand on tape so that further work can be done in a computer.

The third branch of the application of digital techniques to radio astronomy is in data reduction. It is to this third branch that I wish to devote most attention in this paper, as it is this sort of work which is being carried out here by Mr. Haslam, Dr. Largé and myself. In general this involves the use of large computers. The data must be in the form of punched tape, punched cards, or possibly magnetic tape and it is put into one of these forms either automatically by a data logger, or by hand from the reading of paper charts. The computer is then used to reduce this data in volume, to analyse it, and finally to present it as nearly as possible in the form in which one wants it. Also the computer can be used for comparing different sets of data, which may have been obtained at different observatories. There has been a lot of work done here on the analysis of meteor orbits; the reduction of the orbits originally took $1\frac{1}{2}$ hours per meteor, but now it is done in a few seconds by automatic techniques.

THE BASIC DATA

Let us now restrict ourselves to the problem of the reduction of data. The case we shall take is where a region of sky has been surveyed with a total power receiver and we wish to produce from this a map showing contours of constant brightness temperature as a function of position in the sky. At Jodrell Bank the basic data consist of two punched paper tapes. The first of these contains the time, azimuth and elevation of the telescope, which can be punched out at intervals ranging from ten seconds to five minutes. The azimuth and elevation are read from repeaters on the telescope itself, and so represent the true position of the telescope, rather than the demanded position, which eliminates errors in the analogue computer and servo system which drive the telescope.

The second tape contains the receiver output voltage, again as a function of time. Thus, the two tapes are separate but are related because they each contain universal time as the independent variable. The problem is to eliminate the times on these two tapes within the computer, and to express the receiver output voltage as a function of some selected co-ordinate system.

THE CO-ORDINATE SYSTEM

The co-ordinate system will normally be a system which is fixed to the celestial sphere. The method which we have adopted is to take the region of sky, which we assume to be rectangular, and to cover this with a uniform grid of points in two co-ordinates. This grid can be defined in various ways, for example we could use right ascension and declination as our two co-ordinates; but latitude- and longitude-like systems are such that they cannot well be represented on a flat piece of paper. We have defined a co-ordinate system which is symmetrical in both co-ordinates and which gives the least distortion of the spherical co-ordinates over the region of sky in which we are interested. First we define the position of the centre of the region, either in celestial or galactic co-ordinates, and choose this to be our origin. Suppose that we are using celestial co-ordinates. We construct two orthogonal great circles through the origin. One of these is identical with the meridian circle through this point and represents a line of constant right ascension. The other

great circle is the y-axis: it is tangential at the origin to a small circle of constant declination, but diverges from it as x increases positively or negatively. The x and y co-ordinates of a point are now defined as the angular distances (along great circles) of that point from the y and x great circles, respectively. This system introduces no distortion at all along the axes, regardless of how far we go in either of these directions. The maximum distortion occurs almost equally in the four corners of the region, and is quite small for the size of region with which we are concerned at the present time. An effect of distortion which we must remember is that if the beam shape is represented at the origin by a circle of a certain size, the circle will be distorted if we try to represent the beam shape in a corner of the map, which is relevant if one is performing a convolution process in the computer. However, this is a second-order effect which can be ignored for most practical purposes.

CONSTRUCTION OF AN INTENSITY MATRIX

We have, then, a grid of points at uniform intervals in the x–y co-ordinates, and we wish to represent the region of sky by specifying its intensity at each point. How do we convert our two punched tapes into this form? First it is necessary to condense the data, because of the limited storage capacity of the Ferranti Mercury computer which we have been using. Generally the region of sky will have been explored by moving the telescope beam across it in a number of discrete scans. The computer converts the time-azimuth-elevation data points into time x–y data points, and stores those points lying within the region. Of course, we have to tell it the size and position of the region for it to be able to do this. It then takes all the points lying on a scan and fits to them, by a least-square process, two parabolic equations of motion of the form

$$x = a_0 + a_1 t + a_2 t^2$$
$$y = b_0 + b_1 t + b_2 t^2$$

Between specified times, these two equations represent the motion of the telescope across the region, so 10 or 20 data points, amounting to 30 or 60 discrete numbers, are condensed into six coefficients and two times. The computer now knows when the telescope was pointing within the region, and can select from

the tape containing time and receiver voltage those points representing intensities along a particular scan within the region. These intensities will be at arbitrary points within the region, and not, in general, at the points forming the regular grid, as we should like them to be. The process of converting from intensities at known but arbitrary points is as follows.

The intensities are corrected for linearity, gain and baselevel; the computer decides whether the scan is more nearly parallel to the x or the y axis. Suppose it is the x axis. It then constructs a *tabular scan*, which specifies the intensities along the scan at the points of intersection (*tabular points*) with lines of constant x. Imagine a suitable weighting function, centred at a tabular point, whose amplitude falls with increasing distance from the tabular point. Then surrounding data points are added into the tabular point with weights proportional to the amplitude of the weighting function. This amounts to a one-dimensional convolution process. If the weighting function is sufficiently narrow there need be no appreciable degradation of the telescope resolution by this process. We now have the intensities specified at tabular intervals of x but not necessarily of y. The intensities are normalized by dividing them by the sum of the values of the weighting function used in constructing them, and a second convolution is performed in the other dimension to find the intensities at tabular intervals of y. The intensities at these points must be normalized after the second convolution, again by dividing by the sum of the weights used at any given point. To summarize: the scans are converted into tabular scans, which are added, one by one, to the matrix of points which is stored in the computer.

Having obtained a map of a region of sky in this standard form, there is a variety of things which we can do. For example, to compare one's own survey with one made at a lower resolution it is a simple matter to perform in the computer a two-dimensional convolution which will degrade the resolution by an appropriate amount. I must also mention the process of *contour plotting*: we have a programme which will compute the co-ordinates of a chosen contour level by a two-dimensional interpolation between the matrix points. The co-ordinates can be punched in a form suitable for plotting on the automatic graph plotter at Jodrell Bank, so that contour maps can be produced literally untouched by human hand.

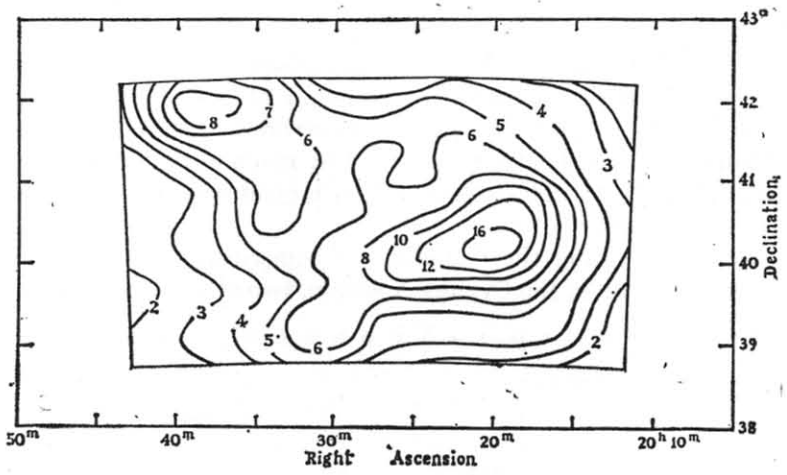


FIG. 6/1.—A survey of the Cygnus-X region at 408 Mc/s, completely reduced by hand (Mathewson, Large and Haslam, 1960).

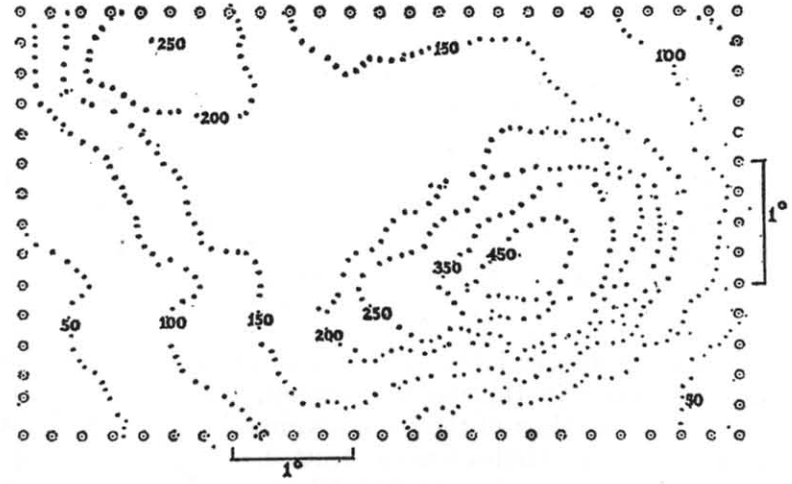


FIG. 6/2.—A survey of the Cygnus-X region at 408 Mc/s, reduced entirely in the Mercury computer. The slight differences between this map and that shown in Fig. 6/1 are due partly to the choice of contour levels, and partly because the spacing of scans in the automatic survey was rather too wide to give good correlation between scans (Haslam, Davies and Large, 1962).

THE RESULTS

Figs. 6/1 and 6/2 show for comparison two surveys of the Cygnus-X region, made at a frequency of 408 Mc/s, the first completely reduced by hand (Mathewson, Large and Haslam, 1960), the second entirely analysed by machine and automatically plotted. It will be seen that there is good agreement between the two maps.

Fig. 6/3 shows the technique of comparing three different surveys, again of the Cygnus-X region, two of which were made

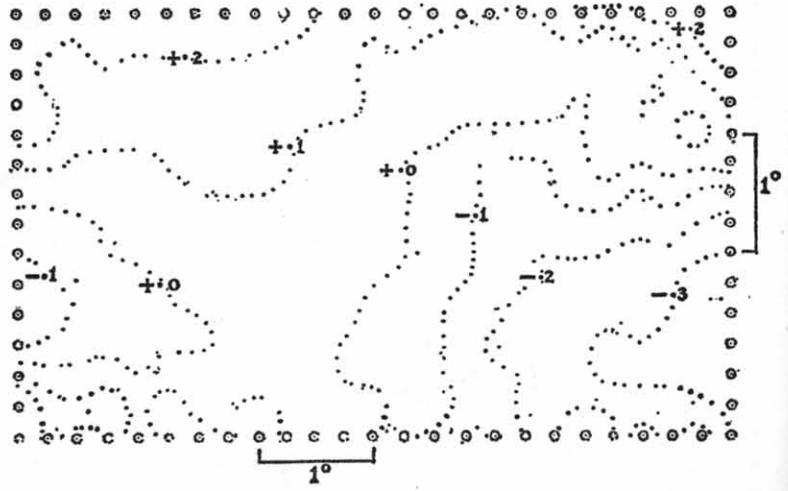


FIG. 6/3.—Showing contours of constant spectral index in the Cygnus-X region, obtained by comparison of maps at 240, 408 and 1390 Mc/s in the Mercury computer (Haslam, Davies and Large, 1962). The numbers do not represent absolute values of spectral index, but there is clearly an increase in the non-thermal component in the bottom right-hand corner of the map.

here, at 240 Mc/s and 408 Mc/s, and the third is at 1390 Mc/s and was made in Holland (Westerhout, 1958). Using point-by-point comparison of the three matrices in the computer, a matrix of spectral index was constructed, and contours of constant spectral index were plotted. The numbers do not represent absolute spectral index because the baselevels of the three maps are arbitrary, but the figure does show a real *variation* of spectral index, becoming more non-thermal in the bottom right-hand corner.

Fig. 6/4 is an entirely new map which has been reduced

automatically. It shows the North Galactic Spur at 240 Mc/s; the saving in time is very great indeed as I think this would have taken about a year to reduce by hand, but took only about thirty hours of computer time.

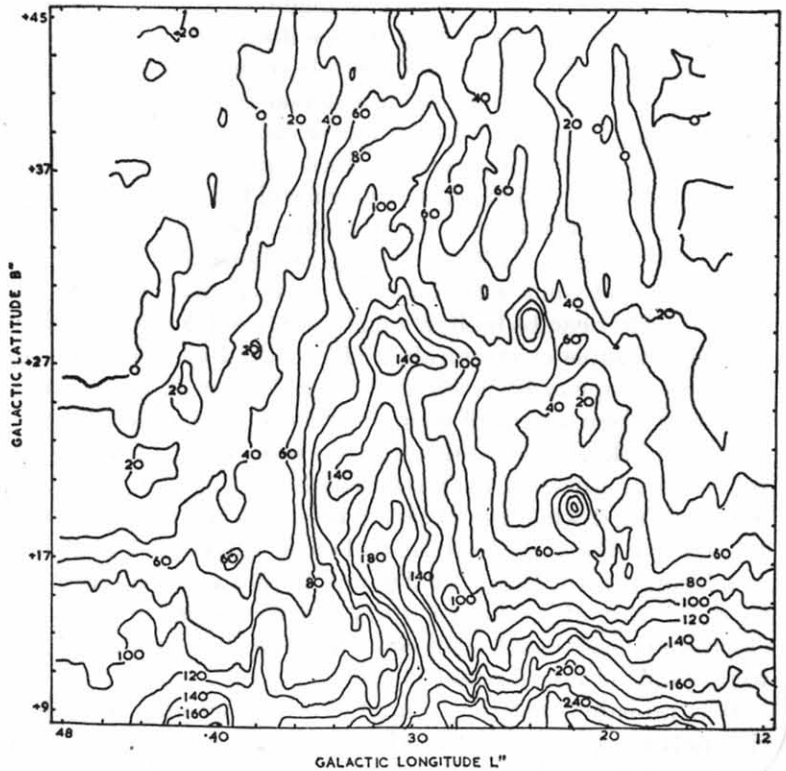


FIG. 6/4.—A map of the North Galactic Spur made at 240 Mc/s with a 1° beam, and reduced almost entirely automatically. The total computing time was about 30 hours. (*Reproduced by kind permission of Dr. M. I. Large.*)

Most of what I have described up to now has been published recently (Haslam, Davies and Large, 1962).

A DIGITALLY-CONTROLLED TELESCOPE

I would like now to say a few words about another use for computers in radio astronomy, to show the way in which our thoughts are going for the future. The foundations have already been laid at Jodrell Bank for a new telescope with an elliptical

bowl of about 120′ × 80′. We are proposing to install in the control system a small special purpose computer, with a store of about 4000 words. This would have as its primary function the conversion of the co-ordinate system to celestial co-ordinates, galactic, or any other system one might choose. It will have inputs from digitizers on the telescope itself, that is it will read directly the actual position of the telescope at any interval of time. It will compare this with the demanded position which it has just computed and will produce voltages at the outputs of the computer which can be amplified directly and fed to the driving system. Obviously, safeguards will have to be built into the computer to prevent undue torque, velocity or acceleration being demanded. The response of the servo system may be more rapid than with an analogue computer, because the optimum filter can be built in a mathematical form which might not be realizable electronically. The method of control of the telescope, too, can be much more flexible than in an analogue computer: for example, if we want to do a series of scans across a region of sky, the data defining the scans can be fed into the computer in advance. Each scan can be executed individually, being initiated when the observer presses a switch on the console; he can then select the time of each scan according to how well his apparatus is functioning and to how much interference there is, without having to worry about setting up rates and co-ordinate systems. A scan could very easily be repeated if it were spoilt by interference.

It is probable that the computer will spend only half of its time actually controlling the telescope, and the remainder could be used for doing simple real-time data processing. For instance, if the data are extremely redundant, the computer can be used to reduce the volume of punched tape which we have to store. If positional and intensity information could be combined on a single tape it might remove the difficulty of combining two tapes in the subsequent data processing.

The computer will have a number of input switches which can be used for control of rates of scanning, etc. However, the function of each switch could be determined by the programme in the machine, to make it completely flexible. It will also have eight output displays, of up to six decimal digits. These again will be general outputs, and what they display can be determined by the programme.

DISCUSSION

Dr. H. P. Palmer: 'How does this computer compare with the Ferranti Mercury?'

Dr. J. G. Davies: 'This is a special purpose computer, originally designed for industrial control. It will have a storage of 4000 words, of which maybe half will be used for the programme. Mercury has a fast store of 1000 words, and a magnetic-drum backing store of 16,000 words, of which a proportion is normally taken up by programmes and translation programmes. So 4000 for this special-purpose computer is fairly generous. When Atlas comes into operation in Manchester the total storage becomes 96,000 for the drum and fast store combined, backed by a magnetic tape store of 2½ million. We shall have a private line to the Atlas building which we hope will ultimately be wired directly into Atlas.'

(*Unidentified.*) 'Does your digitizing system integrate the receiver voltage?'

Dr. J. G. Davies: 'When we have used the data logger for real-time recording, we have employed an integrating condenser which is short-circuited directly after reading, every 10 seconds. The short-circuiting ensures that a burst of interference, which may ruin one reading, will be "forgotten" before the next reading. However when the chart-reading machine is used, it is possible to edit out short bursts of interference from the records, by drawing a pencil line underneath them, and also to integrate the noise by eye.

Dr. H. P. Palmer: 'You told us that the map which you showed was produced in a much shorter time than by hand. Does this allow for the time taken to write the programmes?'

Dr. J. G. Davies: 'The programming effort was perhaps comparable with the work required to produce only one of these maps by hand. The first map therefore takes as long as by hand, but the second is much quicker.'

REFERENCES

Haslam, C. G. T., Davies, J. G., and Large, M. I., 1962, *M.N.*, **124**, 169.
Mathewson, D. S., Large, M. I., and Haslam, C. G. T., 1960, *M.N.*, **120**, 242.
Westerhout, G., 1958, *B.A.N.*, **14**, 215.

7

RECENT DEVELOPMENTS IN RADIO TELESCOPE DESIGN

A Discussion

Contributors: M. Ryle, E. J. Blum, J. P. Wild and F. T. Haddock

PROFESSOR SIR BERNARD LOVELL was chairman of the discussion. In welcoming the speakers he said that the development of radio astronomy depended on the improvement of both sensitivity and resolving power of radio telescopes. In thinking about the requirements in radio astronomy over the next generation of instruments, one is made acutely aware of these factors. The degree to which the need for sensitivity or resolving power determines the design depends largely on the purpose for which the instrument is required. For such problems as source counting for cosmological investigations, the limitation is primarily one of confusion, and a cross or aperture synthesis type of telescope is appropriate. For many other problems, including long baseline diameter measurement, observations of time-varying phenomena and radar echo work, a fully-filled steerable aperture is the only way of obtaining the desired result.

APERTURE SYNTHESIS BY M. RYLE

As Professor Lovell has said, the two principal problems of investigating the radio emission from the sky are those of sensitivity and resolving power. If we consider first the sensitivity problem, we know that what matters is the ratio of power received from the source to the total system noise. The system noise that can be achieved is of the order of 100°K for wavelengths of about 50 cm down to about 3 cm. At wavelengths longer than about 1 m the system noise rises as about $\lambda^{2\cdot5}$, since most of the noise power comes from the sky. If you are interested in looking at the background component, the signal to noise ratio is substantially constant at $\lambda > 1$ m, and falls rapidly at

shorter wavelengths. This indicates an optimum wavelength of about 1 m for background observations with a fully-filled instrument. In observing radio sources, for which the flux density varies as about $\lambda^{0\cdot7}$, we find that for a given 'collecting area' the available signal to noise has a shallow maximum in the vicinity of 75 cm.

The requirements of frequency and sensitivity determine the total collecting area that we need to build. The question now arises whether the resolving power will be sufficient if this area is achieved in a fully-filled aperture. This is where aperture synthesis comes in, for if we use a partly filled aperture we can increase the resolving power of a telescope without sacrificing its sensitivity. We can then choose the wavelength and physical area to give the necessary signal to noise ratio, and make the resolving power what we need by partial filling.

That, then, briefly explains the need for partly filled apertures. Let us now see how they work. A large (square) aerial of side D may be considered as made up of a large number of smaller square aerials of side d (Fig. 7/1). In each of these small

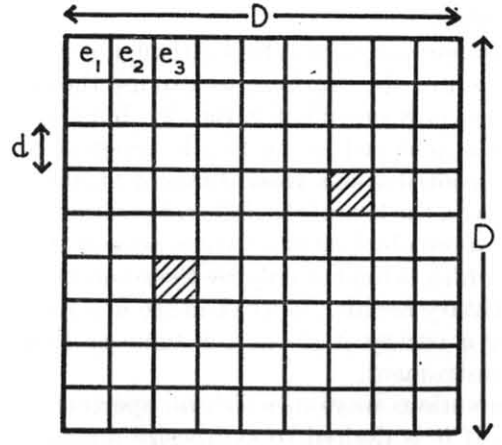


FIG. 7/1.—Diagram to illustrate the two-dimensional synthesis of an aperture.

aerials there will be an e.m.f. which is made up of the vector addition of all the contributions from various parts of the sky. These will be different in each of the small aerials. In a conventional aperture all the components are simply added together in phase (as in a parabolic reflector with dipole feed)

and only waves arriving normal to the aperture plane produce a large signal. The vector components from waves in other directions cancel, and the system has a very small response. Thus the power, P, received will be related to the small (vector) e.m.f.s, e_1, e_2, in the individual aerials, thus:

$$P \propto \Sigma e_n^2 + \Sigma e_m e_n \cos(\phi_m - \phi_n)$$

Suppose we want to reproduce the polar diagram of the large aerial by aperture synthesis. The first term represents the average sky brightness over the reception pattern of the elementary aerial, whilst the second provides the high resolution. We can now take two small aerials and connect them to a correlation type receiver which measures the quantity $e_1 e_2 \cos(\phi_1 - \phi_2)$. By moving the two aerials about and repeating the measurement we can measure all the cross-correlation terms and by adding the results (vectorially) obtain the resolving power that would have been achieved by the large aerial. Of course in the original aperture some spacings of the small aerials occur more commonly than others, so that to reproduce a uniformly illuminated aperture, the observations must be combined with suitable weights.

Now the important point is that although it takes a long time to make one set of observations, we can use the observations to investigate not only one point in the sky, but a large number of points. This is achieved by putting in progressive phase gradients in the contributions from different spacings and in this way we can scan the beam into different positions. Thus without making any more observations we can scan in the computer an area of sky which is limited only by the resolving power of each of the elementary aerials. As we shall see in a moment the time required for a survey is about the same as that for the full equivalent instrument.

There are various ways in which an aperture synthesis may be realized. If it is desired to synthesize a square aperture of size D, this may be done by fixing one small element and moving the other over an area $2D \times D$. Another way is to have the two elements mounted on railway tracks in the form of a letter T, which is effectively three arms (the fourth is unnecessary) of a cross of dimensions $2D$. We can if we like go to a one-dimensional synthesis, in which a long element (length $2D$) is built, and the synthesis is made by moving a small element along a

line at right angles to the long element. This is particularly convenient in practice when the long aerial is built E–W, since it is only necessary to provide for rotation about this long axis, for complete sky coverage to be obtained.

If we take this one stage further and fill in the other (N–S) arm, we are back to the Mills cross in which all the relative spacings are present simultaneously. Another method, proposed by Wild (1961), which also has all the spacings present simultaneously, consists of a circular ring of diameter D.

So far we have discussed pencil beam instruments. Interferometers may be synthesized equally well, by having the small elements of one arm displaced relative to the other.

It is now important to consider the signal to noise ratios which may be achieved in aperture synthesis. To make a map of a given area of sky with an aperture of sides a and b you must observe points in the sky separated in angle by amounts not greater than $\lambda/2a$ and $\lambda/2b$ to obtain full information. Let us consider the time it takes to observe an area of sky Ω using a time constant τ. For a conventional aperture (square, of size D) then the total number of observations is given by $4\Omega\frac{D^2}{\lambda^2}$. The signal to noise is proportional to $D^2\sqrt{\tau}$. If we now carry out the same survey using a two-dimensional synthesis, in which you have a pair of aerials of size d, then in each relative position of these you will have to make a total of $4\Omega\frac{d^2}{\lambda^2}$ observations, and the two aerials have to be placed at $2\frac{D^2}{d^2}$ independent spacings to complete the synthesis. The total observing time is thus $8\Omega\frac{D^2}{\lambda^2}\tau$, or just twice the time required for the filled aperture survey. The signal to noise ratio of each observation will be $2d^2\sqrt{\tau}$, and since there are $2\frac{D^2}{d^2}$ observations, and the noise powers add incoherently, the final signal to noise on the resulting map will be

$$2d^2\sqrt{\tau}\sqrt{\frac{2D^2}{d^2}}$$

If we cut down the integration time to $\frac{1}{2}\tau$ to make the observing

time equal to that for the filled aperture survey, then the final signal to noise ratio is $2dD\sqrt{\tau}$. There is a small numerical factor missing because we have not weighted the observations, but you can see that we now have a system which will survey a region with the same resolving power and a signal to noise which is down by a factor of d/D (not d^2/D^2 as might be thought). That is to say that the effective collecting area from the point of view of signal to noise is the geometric mean of the synthesized area and the small element. The same result also applies to a one-dimensional synthesis and to a Mills cross.

At Cambridge we are just starting construction of an instrument for two-dimensional synthesis. Instead of moving the second aerial over a two-dimensional field of points, the diurnal-rotation of the Earth is used to rotate the baseline through 180 degrees in twelve hours. Thus by placing the two elements at a set of points on an E–W baseline we have a complete set of spacings and orientations for the synthesis of a circular aperture of diameter equal to the maximum spacing used.

REFERENCES

Ryle, M., and Hewish, A., 1960, *M.N.R.A.S.*, **120**, 220.
Wild, J. P., 1961, *Proc. Roy. Soc. A.*, **262**, 84.

Partly Filled Apertures by E. J. Blum

In a normal Mills cross radio telescope an array of dipoles is built, the dipoles being linked by cables. Some tapering of the feed system is introduced in order to produce a satisfactory beam shape with low sidelobe levels. For very large cross-type aerials this design is not completely satisfactory and two alternatives have been put forward for the design of the large Benelux cross.

In the first design (Christiansen and Högbom, 1961) a long cylindrical paraboloid forms one of the arms, but gaps of length 'l' are left at regular intervals. Each section of this arm has its own individual line feed system and preamplifier. The elements of the orthogonal arm are then sections of cylindrical paraboloid of length l parallel to the first arm. Thus all spatial frequencies in the response of the aerial are present, up to the

maximum determined by the length of the arms. This scheme was also proposed by Bracewell (1961) and has many attractive features, but it is not much use in hydrogen line work where continuous following may be necessary in order to achieve long integration times. Therefore in the latest design for the Benelux cross, there is a double row of parabolae in each arm, arranged in a staggered pattern so that there are no gaps left in the spatial frequencies. There will be a total of 123 33-metre dishes in arms 1·5 km long, and many combinations of connections between the elements will be available (about 4000 in number).

Another possibility is to combine the principles of the cross aerial with that of the Kraus fixed reflector aerial to make a cross of two narrow sections of parabolic surface. The sections would have very long focal length, and would be arranged so that their foci were at ground level. Such an instrument would be limited to observing quite close to the zenith, but could well have such sensitivity that it could be used for cosmological investigations. Von Hoerner (1961) has described such a scheme. By combining this with a movable reflector it would be possible to explore the whole sky.

Discussion

In reply to questions Professor Blum said that the final design of the Benelux cross had not been decided upon. He agreed that the ultimate limit in the possibility of realizing very high resolutions (by whatever kind of instrument) would be set by differential variations in the ionosphere, or troposphere. Dr. Jennison emphasized that if one used interferometers between each of three aerials phase variations introduced by the ionosphere could be cancelled in the analysis. (Jennison, 1958). Professor Blum also pointed out that a filled aperture may be arranged to provide many independent beams in the sky simultaneously. If this is done then the method of aperture synthesis shows up less well in a comparison of the sensitivities of various types of radio telescope.

REFERENCES

Bracewell, R. N., 1961, *Stanford Radio Astronomy Inst. Publ.*, No. 14.
Christiansen, W. N., and Högbom, J. A., 1961, *Nature*, **191**, 215.
Jennison, R. C., 1955, *M.N.R.A.S.*, **118**, 276.
Von Hoerner, S., 1961, Publ. *N.R.A.O.*, **1**, 19 and 63.

A New Telescope for Solar Radio Observations

by J. P. Wild

I shall describe a new instrument for obtaining radio pictures of the Sun. The problem of solar radio astronomy differs from that of general sky radio astronomy in that the Sun is a variable radio source. This means that we must take pictures of the Sun in a short time, and cannot reasonably use the method of aperture synthesis. On the other hand, the Sun is such a strong source that the radio emission from the rest of the sky can be ignored, enabling one to make considerable economies in the design of an instrument.

The requirement of our new instrument to be built in Sydney is to form radio pictures of the Sun at 80 Mc/s with a resolution of 3·5 min of arc over a 2° field. The instrument is to be able to form a picture of the high intensity (burst) emission in one or two seconds, and of the weaker quiet Sun emission in a few minutes.

This could be done with a steerable paraboloid 3 km in diameter. A more reasonable proposal is to use a cross aerial of 6 km arms, and in looking for economies in the design, it seemed that a circular array offered several advantages. That is to say we simply use the outside rim of the 3 km paraboloid. The polar diagram of such an annulus is similar to that of a fully-filled aperture, but with large side lobes. Mathematically it is given by $(J_0(x))^2$ compared with the $\left(\frac{J_1(x)}{x}\right)^2$ for the filled circular aperture. However since the ring aperture contains every spacing and orientation present in the fully-filled aperture it is possible to recover the image that would have been produced by the filled aperture. This may be done by a convolution using the appropriate circularly symmetrical function.

There is no need to build the whole of the ring aerial. If the ring is broken up into a number of small, spaced elements, the polar diagram will be identical to that of a complete ring out to a certain angular radius determined by the spacing of the individual elements, while beyond this radius a complex side-lobe structure appears. Since the Sun is very much more intense than the rest of the sky, this side-lobe structure is unimportant provided it lies completely outside the desired field of view containing the Sun. This requirement can be met by making the gaps in the ring suitably close, and in the Sydney instrument it is proposed to build 100 parabolic dishes each 42 ft in diameter and steered towards the Sun. At each element there will be a preamplifier, and the open wire feeders will run along radii to the laboratory at the centre of the circle.

At the centre there will be 60 detectors with different phase connections from the mixers so that 60 beams will traverse the Sun simultaneously. The beams will sweep the Sun about once per second, a complete 60-line picture of the bursty Sun being formed in a second.

We have looked into a number of analogue means of restoring the image of the Sun from the records while observations are in progress, but the latest idea is to record the picture in the fast-store of a small special-purpose digital computer, convolve it with the necessary beam restoring function and display the output on a cathode-ray tube where it will be photographed by a cine camera.

The circle arrangement of elements is preferable to a cross for the following reasons:

(i) The total extent of flat land required by the instrument is smaller (although the area may be larger or more awkward in shape!).
(ii) The connections to the elements are twice as short as in the cross (with consequent reduced attenuation), and more symmetrically disposed.
(iii) Fewer elements are required. To do the same job with the same redundancy the circle needs $\pi/4$ times the number of elements as the cross.
(iv) Calibration of the phase connections can possibly be made by radiating a test signal from the centre of the circle.

Discussion

In reply to questions Dr. Wild said that some of the elements of the ring aerial could be omitted without loss of basic information. However the full range of possibilities and comparison with the cross type of aerial had not been worked out. The Sydney instrument was designed to be able to follow the Sun for two hours on each side of transit. There had been some talk of building the aerial elliptical rather than circular, in order to

avoid the foreshortening effect. This leads to the elegant result that the length of feeders is then exactly right if they are mixed at a focus of the ellipse, but for practical reasons it may be better to use the circular shape.

Large Radio Telescopes in America
by F. T. Haddock

The 1000-ft spherical reflector is being built in Arecibo, Puerto Rico, under the direction of Professor William Gordon of Cornell University. The reflecting membrane will be formed by suspending a wire mesh from the periphery of a large hole in the ground. At the focus will be a long travelling-wave antenna structure having a phase structure such that the spherical aberration of the reflector is eliminated. This instrument will be used with high-power radar to study back-scatter from electrons in the region from the ionosphere to the solar corona and to study the planets. It will also be used for passive radio astronomy observations at frequencies, hopefully up to 1420 Mc/s.

The 600-ft completely steerable paraboloid reflector being built at Sugar Grove, West Virginia, by the Navy is designed to operate up to 2400 Mc/s. In May about 75 per cent of the engineering design and 25 per cent of the construction had been completed. The surface will be formed out of 250 shaped panels about 70 ft square and covered with a $\frac{3}{4}$-in expanded metal mesh. The corners of the panels will be individually servo-controlled to the correct position as determined by a 'silo', a vast drum of invar metal, which will direct 250 pairs of light beams to the corners of each panel for position reference. When the instrument is not required for operational purposes it will be available for radio astronomy as a national facility. [On July 19, 1962, the United States Government announced the cancellation of this project. The reasons given were that the structure grew to unmanageable weight and complexity completely beyond that expected and this resulted in increasing costs. Furthermore, the instrument was of decreasing potential usefulness because of unforeseen advances in science and technology and was being bypassed by electronic and satellite technology.]

The 300-ft parabolic reflector at Green Bank, West Virginia,

is being constructed by the National Radio Astronomy Observatory. It is a meridian instrument designed primarily for the 21-cm hydrogen-line observations. It is nearing completion and will be in operation in early autumn.

The 140-ft precision reflector, partially constructed at the same observatory, is designed to operate at a wavelength of 3 cm. The design specifies a spherical oil-pad bearing which has led to nearly insurmountable fabrication problems. At present the construction of the project is at a reduced pace while a restudy of design and fabrication is under way.

The 400-ft by 600-ft north–south cylindrical parabolic reflector being built at the University of Illinois is nearing completion. This reflector was formed by shaping a suitable valley and lining it with roofing material and a reflecting wire mesh. Difficulties were experienced with rain water washing out the underlying ground. These problems have now been solved. It is planned to start in the near future transit observation of radio sources for cosmological purposes.

Discussion

In reply to questions Professor Haddock said that he knew of no American plan to establish large radio telescopes in space other than about 30-ft paraboloids for telemetry. At millimetre wavelengths there are plans to scan Venus at 13·5 millimetres and 19 millimetres from a range of about 20,000 kilometres on a fly-by trajectory in order to obtain a radiometric map of the planet and to study the solid surface.

8

INTERSTELLAR MATTER

F. D. KAHN

BASIC QUANTITIES

THE total intensity of starlight gives us a fair impression of the radiant energy density available in interstellar space. The radiation field there is usually likened to that of a black body at 10^4 °K, diluted by a factor 10^{-14}. This makes its energy density equal to that found in thermodynamic equilibrium at 3°K.

However, it is a poor approximation to compare the intensity distribution to that of the black body spectrum, since the interstellar field is due to the superposition of light from many stars, none of which radiates like a black body. The curve most commonly used for the energy distribution was calculated by Lambrecht (1955); his work was done before measurements of stellar ultra-violet radiation had been made from rockets and satellites, and will probably have to be revised in the all-important region $\lambda < 3000$ Å.

The density of matter in interstellar space can be estimated —given certain assumptions—from that of interstellar atomic hydrogen. Observation and interpretation of 21-cm line profiles give an average density of the order of one H atom per cm³ in our part of the Galaxy, and within a slab of some 500 parsec thickness around the galactic plane (Westerhout, 1957). With the commonly assumed composition for interstellar matter (see, for example, Kahn, 1960), this corresponds to a mass density of some $2{\cdot}8 \times 10^{-24}$ gm/cm³. For comparison, the smoothed-out mass density of the stars in our neighbourhood is estimated to be about $0{\cdot}08\ \odot$ per parsec³, or 6×10^{-24} gm/cm³. In other words near the Sun the mass contributed by the stars is about twice that contributed by interstellar matter.

COMPOSITION OF INTERSTELLAR MATTER

Interstellar spectral lines have been observed in absorption for H, Na, K, Ca, Ti, Fe, CH, CN, and NaH, while H, O and N

have been observed in emission. The relative intensities of the various lines agree with the assumption that the interstellar gas has a composition like that of a population I star. This means that the elements heavier than helium are expected to contribute only some two per cent by mass; the remainder would be hydrogen and helium, present in the ratio of about 3 : 2, by mass.

But a large uncertainty remains, partly because of the existence of interstellar dust, which incidentally obscures our view of stars lying in or near the galactic plane. The mean free path for a photon through the dust is about one kiloparsec, but this length varies with colour since blue light is scattered very much more than red. The variation of scattering efficiency with wavelength can be found by observation of the reddening of distant stars whose colour is known; a subsequent comparison of this observed variation of scattering efficiency can then be made with the variation predicted by theory for different models of dust grains. From such considerations and from data concerning the forward directivity of the scattering process, van de Hulst (1955) concludes that the best model for a dust grain has a diameter of 10^{-4} cm, and is composed of frozen H_2O, CH_4 and NH_3, together with various impurities. The mass of a grain of this size and consisting of such a mixture would be about 5×10^{-13} gm and its extinction area about $2{\cdot}7 \times 10^{-8}$ cm² for radiation with $\lambda = 5000$ Å. It follows that on the average there must be about $1{\cdot}1 \times 10^{-14}$ grains per cm³ in order to produce the required extinction; the grains therefore contribute some $5{\cdot}5 \times 10^{-27}$ gm/cm³ to the interstellar mass density. It is thus quite possible that they contain an appreciable fraction of the heavier atoms in interstellar space. The fraction may be larger than our figures indicate, for our estimate of the mass of the interstellar dust has probably been rather economical. We have taken rather large values for the photon mean free path and for the scattering efficiency of the typical grain, both of which would tend to reduce the mass of dust required.

IONIZED HYDROGEN (H II) REGIONS

Radiation from very early-type O and B stars contains an appreciable amount of ultra-violet light beyond the Lyman limit, that is with $\lambda < 912$ Å. This radiation can, and does,

ionize the hydrogen nearby to form H II regions. The size of an H II region can be simply related to L_*, the rate of production by its exciting star(s) of photons beyond the Lyman limit. These photons are mainly required to balance the recombinations which occur in the interior of the H II region, at a rate Bn_pn_e per unit volume and unit time.

Here B = recombination coefficient to all states
n_p, n_e = proton and electron densities, respectively

If β is the recombination coefficient to the ground state, then a Lyman-continuum photon is recovered in a fraction β/B of the recombinations. Supply and demand are balanced when

$$L_* = \int (B - \beta)n_pn_e dV \quad . \quad . \quad . \quad (1)$$

the integration being carried out over the volume V of the H II region.

The expression on the right-hand side of (1) bears a simple relation to some quantities which can be directly observed. Optical observation of H II regions is usually made by means of the Hα line, seen in emission. The rate at which such photons are produced is directly proportional to the recombination rate to the third and higher quantum levels of the H atom, and so to the product n_pn_e. Again, radio-astronomical observation registers the free–free emission of H II regions, and this also occurs at a rate proportional to the product of the ion and electron densities.

A major uncertainty in the use of relation (1) lies in the fact that a value for L_* can only be found from reliable models for the atmospheres of the stars concerned. These are not available yet. Observational determination of L_* is clearly out of the question. As we have seen Lyman continuum radiation is weakened by absorption as it travels away from the star where it was emitted. The boundary of the corresponding H II region occurs at the place where this radiation is all used up, and as far as we know the Sun lies in a non-ionized (H I) region. Thus no Lyman continuum radiation reaches the vicinity of the Earth from any star.

From measurements of angular sizes and from estimates of their distance one infers that H II regions have diameters of the order of one to ten parsecs. By comparison with these linear

dimensions the transition region, or ionization front, separating an H II from an H I region is rather thin.

To illustrate this, let us make the reasonable approximation that equal numbers of protons and electrons are present, per unit volume, in any one place, so that

nx = proton or electron density
$n(1 - x)$ = density of H atoms.

If J is the intensity of Lyman continuum radiation, the rate of production of new protons is proportional to $Jn(1 - x)$ and the rate of loss of protons by recombination to n^2x^2. In a state of ionization balance and at a given hydrogen density n the ratio $x^2/(1 - x)$ must then be proportional to J. Now let J_{75} and J_{25} be the values of J at the places where the proportional ionization is 75 per cent and 25 per cent, respectively. We find that $J_{75} : J_{25} = 27 : 1$. This reduction in intensity occurs on passage through an optical depth $\tau = 3{\cdot}3$, or a physical distance $d = 3{\cdot}3/\frac{1}{2}n\sigma$. Here we have taken $\frac{1}{2}n$ to be the typical density of atomic hydrogen within the ionization front (a suitable numerical value is $n = 10$ cm^{-3}), while for σ we take the absorption cross-section of the H atom at the Lyman limit, which is 6×10^{-18} cm^2. This makes d equal to about 0·04 parsec, which is very small in comparison with the size of the H II region.

Two processes essentially control the temperature of an H II region. They are:

(*a*) The gain of energy from the photo-electric effect. When an H atom is ionized, the electron carries off the excess energy of the ionizing photon above the Lyman limit, and this adds to the store of heat.

(*b*) The loss of energy due to the excitation by electron impact of a forbidden level of the O^+ ion, with $\chi = 3{\cdot}31$ eV.

Our diagram (Fig. 8/1) is taken from Spitzer's review of the subject (Spitzer, 1954), and illustrates how the balance is achieved. The steep gradient of the L_{ei} curve means that even if the oxygen abundance differs from that assumed in the calculation, the point of balance will not be much displaced horizontally. In general the temperature in an H II region is therefore near 10^4 °K, unless the O^+ ions are themselves ionized to O^{++}, in which case some less efficient cooling process takes over.

All this treatment has been given as if the typical H II region were a static configuration. This is not so. Stars which are hot enough to form an appreciable H II region use up their nuclear fuel within a few million years; this is a short period on the astronomical time scale. The dynamics of the process by which H II regions are set up must therefore be considered.

Now the formation of an H II region disturbs violently the equilibrium in the H I gas nearby, for, as we shall see, the typical

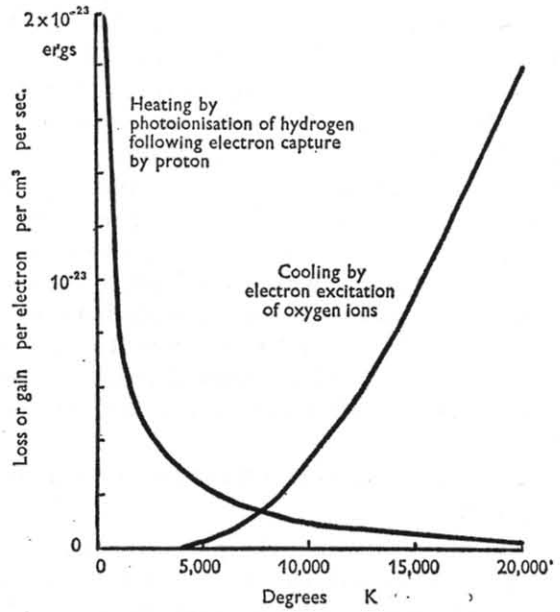


FIG. 8/1.—The temperature variation of the heating and cooling processes in regions of ionized hydrogen.

H I temperature is only of the order of 100°K. At a given density the typical H II pressure is therefore some 200 times larger than the typical H I pressure. (The extra factor 2 occurs because each H atom contributes two independent particles when ionized.) Thus when a bright star lights up in an H I region it produces an explosive increase in the pressure nearby. The shock waves from this event can propagate into the neutral gas surrounding the site, and will help to compress it and stir it up (Kahn, 1954; Goldsworthy, 1961).

Two interesting dynamical effects are associated with H II

regions. The first concerns the so-called elephants' trunks and globules, which are formed when a rather denser than usual mass of non-ionized gas is engulfed by an advancing H II region. The non-ionized gas is then steadily eaten away, the ionized gas so formed streams into the H II region, and initially has a higher density than the H II background. This means that the region near the ionization front will be abnormally bright in Hα, and such bright edges are in fact seen to occur. At the same time, recoil effects help to compress the trunk or globule still

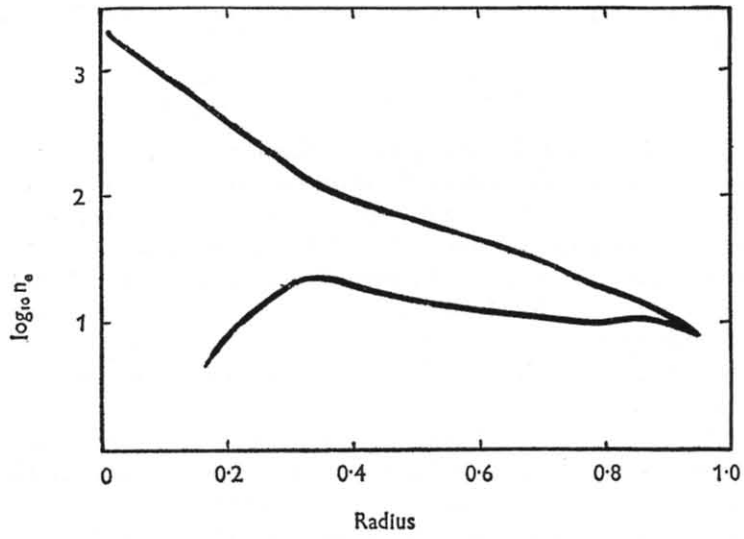


FIG. 8/2.—Variation of electron density with radial distance in the Orion and Rosette nebulae.

further. This theoretical explanation is due to Pottasch (1958) and recent observations by Courtès, Cruvellier and Pottasch (1962) have confirmed that it is correct. It has been suggested that globules may ultimately evolve into stars.

An interesting radio-astronomical observation leads to a determination of the run with radial distance of the ion and electron densities in some H II regions. Results for the Orion and Rosette nebulae are illustrated in Fig. 8/2; a fuller discussion of the measurements and their interpretation has been given by Menon (1962). The pronounced density gradients imply corresponding pressure gradients which cannot be supported

statically. The interpretation of the profiles is that in its orignial state the nebula was cool and non-ionized, with a density distribution which decreased outwards. The associated relatively small pressure gradient was balanced by self-gravitation. Sudden ionization then occurred, leaving the nebula much hotter and destroying the equilibrium. In the resulting outward motion the shape now shown by the Rosette nebula would be formed after some 50,000 years. A detectable dip should be observed at the centre of the Orion nebula after some 10,000 years; since none is seen there, this sets an upper limit to the period which has elapsed since the object was first ionized (Kahn and Menon, 1961).

EXCHANGE OF MATERIAL WITH STARS

Thus star formation is a process which is very much in progress at present. The material drawn into the stars comes from interstellar space; the rate of consumption is estimated to be of the order of 1 $\odot$ /annum in the Galaxy. For comparison the present galactic content of interstellar matter is some 10^9 $\odot$; since the age of the Galaxy is 10^{10} years, one infers that considerable quantities of matter must be returned into space from stars—otherwise the interstellar supply would by now be exhausted.

These numbers show that most probably much the greater part of the present interstellar matter has been processed in the stars at some time. One interesting conclusion to be drawn from this is that the abundance of interstellar deuterium is likely to be low, for it is known that the greater part of the typical stellar interior is too hot for this substance.

The return of matter into space is thought to be due principally to stellar winds, and to nova and supernova explosions. These latter often leave interesting remnants in space; for example the Crab nebula is the debris from a supernova which occurred some 900 years ago, and the Great Loop in Cygnus, a much larger structure, still in expansion, is believed to have been set off by a supernova which occurred perhaps 50,000 years ago.

NEUTRAL HYDROGEN (H I) REGIONS

The H I regions contain an estimated 90 per cent of the interstellar hydrogen. No radiation can reach them if its frequency

exceeds the Lyman limit; hence only atoms with an ionization potential below that of hydrogen (13·53 eV) can be ionized. The most important ionized constituent is probably carbon, and the estimated electron density is 10^{-4} of the density of atoms.

It is inferred from observation that the H I gas is not uniformly distributed in space; a rough model which is often used represents the gas as being present in discrete clouds, whose linear dimensions are of the order of five parsecs, and which occupy perhaps one-tenth of the space available to them. This deduction is based on the observed patchiness of the interstellar extinction, and on the frequency of occurrence of reflection nebulae; these latter are H I clouds containing dust which scatter light towards the observer from stars sufficiently close to them (Strömgren, 1948). The chopped-up character of the Ca^+ and Na absorption lines as well as an analysis of their curves of growth suggests that the clouds have random velocities of around 10 km/s relative to the local standard of rest (Blaauw, 1952, Münch, 1958). Finally it is inferred from the maximum measurable intensity of the 21 cm line that the (harmonic mean) temperature of the H I gas is 125°K (van de Hulst *et al.*, 1954).

The irregular and supersonic motion of the neutral gas draws its energy from expanding H II regions and supernova explosions. The temperature is determined principally by heating processes due to the degradation of this turbulent energy (Kahn, 1955*a*) and collisions with cosmic ray particles (Hayakawa, 1960). The important cooling processes seem to be due to excitation of low-lying rotational states of H_2 molecules (Kahn, 1955*b*) and of low-lying forbidden states of C^+, Fe^+ and Si^+ (Seaton, 1955).

An outstanding unsolved problem concerns the abundance of H_2 molecules in H I regions. Suitable detectable lines in the H_2 spectrum probably exist only in the far ultra-violet, at $\lambda = 1049$ Å, and in the far infra-red, at $\lambda = 28$, 42 and 84 μ. The attempt to observe these will soon be made from a high-altitude balloon and/or satellite.

MAGNETIC FIELDS

If there is an interstellar magnetic field the motion of the interstellar plasma even in the H I regions will be closely

coupled to it. The motion of the neutral gas cannot differ much from that of the plasma over a scale large compared with one mean free path for a neutral atom with respect to the ions. According to Osterbrock (1961) this length is surprisingly small, about 10^{17} cm or 0·03 pc, in a cloud with 10^{-3} ions/cm^3. The coupling of gas, plasma and magnetic field is therefore good on most length scales likely to be important.

Attempts to find an interstellar magnetic field by observation of the Zeeman effect at 21 cm have now yielded a positive result, and suggest a field strength of about 5×10^{-6} Γ (Davies *et al.*, 1962). Indirect evidence that a field exists is also quite strong, and is based on these arguments:

(i) The dependence of the amount of interstellar extinction on the direction of polarization (interstellar polarization) implies that some large-scale field lines up interstellar dust grains, which are themselves not isotropic in structure.

(ii) Cosmic ray particles must be confined within some reasonably small neighbourhood of the Galaxy. This can best be done by a large-scale magnetic field. If the particles were not so confined, they would fill the whole Universe, and their energy density measured at the Earth would be representative for all space. An intolerable problem concerning the supply of their energy is then raised.

(iii) The continuum radio noise from the Galaxy is generally ascribed to synchrotron emission in the interstellar magnetic field (see, for example, Biermann and Davis, 1960).

(iv) Spiral arms of galaxies like ours consist mainly of interstellar gas and of stars genetically associated with such material. The spiral shape would be rapidly destroyed by differential rotation in the gravitational field of the Galaxy as a whole, unless the gas were subject to some other suitable large-scale field of force. A magnetic field might serve, but would have to have a strength of at least several times 10^{-5} Γ.

(v) The interstellar gas lies quite accurately in the galactic plane, except for a twist in the outer parts of the Galaxy. The cause of this may be traced back to the uneven pressure distribution exerted by the intergalactic matter on the galactic magnetic field as it flows past the Galaxy.

The resulting stress leads to a deformation of the distribution of interstellar matter (Kahn and Woltjer, 1959).

AN IMPORTANT UNSOLVED PROBLEM

Van Woerden, Rougoor and Oort (1957) have discovered by means of 21 cm observations that there is a steady outflow of matter from the region of the galactic centre.

The mass involved is several solar masses a year, and the typical velocity about 50 km/s. The driving force responsible is not known.

REFERENCES

Biermann, L., and Davis, L., 1960, *Zs. f. Ap.*, **51**, 19.
Blaauw, A., 1952, *B.A.N.*, **11**, 459.
Courtès, G., Cruvellier, P., and Pottasch, S. R., 1962, *Ann. d'Ap.*, **25**, 214.
Davies, R. D., Verschuur, G. L., and Wild, P. A. T., 1962, *Nature*, **196**, 563.
Goldsworthy, F. A., 1961, *Phil. Trans.* (A), **253**, 277.
Hayakawa, S., 1906, *Publ. A.S.*, Japan, **12**, 110.
van de Hulst, H. C., 1955, *Mem. Soc. Roy.*, Liége, **15**, 393.
van de Hulst, H. C., Muller, C. A., and Oort, J. H., 1954, *B.A.N.*, **12**, 117.
Kahn, F. D., 1954, *B.A.N.*, **12**, 187.
1955*a*, in *Gas Dynamics of Cosmic Clouds* (North-Holland Publishing Co.), p. 60.
1955*b*, *Mem. Soc. Roy.*, Liége, **15**, 578.
1960, in *Die Entstehung von Sternen* (Springer), p. 131.
Kahn, F. D., and Menon, T. K., 1961, *Proc. Nat. Acad. Sci.*, **47**, 1712.
Kahn, F. D., and Woltjer, L., 1959, *Ap. J.*, **130**, 705.
Lambrecht, H., 1955, *Mem. Soc. Roy.*, Liége, **15**, 562.
Menon, T. K., 1962, *Ap. J.*, **135**, 394 and **136**, 95.
Münch, G., 1958, *Rev. Mod. Phys.*, **30**, 998.
Osterbrock, D. E., 1961, *Ap. J.*, **134**, 270.
Pottasch, S. R., 1958, *B.A.N.*, **14**, 29.
Seaton, M. J., 1955, *Ann. d'Astrophys.*, **18**, 188.
Spitzer, L., 1954, *Ap. J.*, **120**, 1.
Strömgren, B., 1948, *Ap. J.*, **108**, 242.
Westerhout, G., 1957, *B.A.N.*, **13**, 201.
van Woerden, H., Rougoor, W., and Oort, J. H., 1957, *C.R.* (Paris), **244**, 1691.

Note: Several references are made to *Mem. Soc. Roy.*, Liége, **15**. This volume is usually called 'Les particules solides dans les astres'.

9

GALACTIC HYDROGEN LINE STUDIES

D. S. HEESCHEN

THE 21 cm line studies, largely by astronomers at Leiden and at Sydney, have brought about a remarkable increase in our knowledge of galactic structure. We can here give only a brief summary of what has been done thus far. Emphasis will be placed on observational results and techniques of analysis, rather than on instrumentation, which is in itself a large and complex field.

BASIC EQUATIONS

We summarize here some of the basic equations used in the analysis of 21 cm line observations, without derivation. For a more detailed discussion of the analytical tools and techniques, the reader is referred to Heeschen (1955) or to several of the Leiden papers in *B.A.N.*, Vols. 12 and 13.

The brightness temperature in the hydrogen line, $\Delta T(\nu)$, is given by

$$\Delta T(\nu) = T_K\,(1 - e^{-\tau_\nu}) \quad . \quad . \quad . \quad (1)$$

where T_K is the kinetic temperature of interstellar hydrogen and τ_ν is the optical depth in the line at frequency ν. This equation is strictly correct only if the excitation of the hydrogen line is by collisional processes, if T_K is constant along the line of sight, and if continuous absorption and emission processes can be neglected. If the line excitation is not by collisional processes, T_K must be replaced by a temperature—the so-called spin temperature—which better expresses the relative population of the two hyperfine levels of the ground state of neutral hydrogen. In most cases encountered in galactic 21 cm line work, equation (1) is appropriate, the principal exception being in the case of absorption line studies.

The optical depth in the line is given by

$$\tau_\nu = 5{\cdot}44 \times 10^{-14} \int_0^\infty \frac{n}{T_K} f(V)\mathrm{ds} \quad . \quad . \quad (2)$$

where n is the total number of hydrogen atoms per cm^3 and $f(V)$, with units cm^{-1} seconds, is a line shape function. This equation is generally valid as long as most of the hydrogen atoms are in the ground state. If $f(V)$ and T_K are constant along the line of sight, equation (2) becomes

$$\tau_\nu = 5{\cdot}44 \times 10^{-14} \frac{N_H}{T_K} f(V) \quad . \quad . \quad (3)$$

where $N_H = \int_0^\infty$ nds. The broadening of the hydrogen line, represented by $f(V)$, is in most cases due to random motions of the hydrogen clouds. Two types of function have been most commonly used for $f(V)$, a gaussian and an exponential of the form $e^{-\frac{|v|}{\eta}}$.

For the case $\tau \ll 1$ equations (1) and (2) reduce to

$$\Delta T(\nu) = \tau_\nu T_K = 5{\cdot}44 \times 10^{-14} N_H f(V) \quad . \quad (4)$$

We also have in this case

$$\int_0^\infty \Delta T(\nu)d\nu = 5{\cdot}44 \times 10^{-14} N_H \int_0 f(V)dV$$
$$= 5{\cdot}44 \times 10^{-14} N_H \quad . \quad . \quad . \quad (5)$$

that is, the area under a line profile is proportional to N_H, the surface density of neutral hydrogen.

GALACTIC ROTATION

(i) A knowledge of galactic rotation is essential for almost all galactic hydrogen line work, because it provides a means of determining the distance of an H I cloud or complex emitting 21 cm line radiation. Galactic rotation at distances from the centre greater than the Sun's distance cannot be determined directly from 21 cm line observations. In most 21 cm line work, a function $\omega(R)$ determined by Schmidt (1956) has been used. This law of galactic rotation was obtained from a model of mass distribution in the Galaxy, which in turn is based in large part on values of $\omega(R)$ for $R < R_0$ determined from 21 cm line observations.

(ii) Galactic rotation in the inner region of the Galaxy—at distances from the centre less than the Sun's distance—can be studied from 21 cm line observations. The radial velocity, relative to the local standard of rest, of a point distance R from

the galactic centre in a direction making an angle λ with the direction to the centre, is, assuming circular galactic rotation:

$$V_r = R_0[\omega(R) - \omega_0] \sin \lambda \quad . \quad . \quad (6)$$

where R_0 is the Sun's distance from the centre and ω is the angular velocity of rotation. For a given value of λ less than 90°, as one moves away from the Sun, V_r goes through a maximum (positive or negative depending on which side of the Galaxy we look) and then changes signs for $R > R_0$. The maximum radial velocity occurs at the point where the radius vector from the galactic centre is perpendicular to the line of sight. The distance from the centre of the position of maximum velocity is

$$R_m = R_0 \sin \lambda \quad . \quad . \quad . \quad (7)$$

(iii) $\omega(R)$ ($R < R_0$) has been determined from hydrogen line observations with the aid of equations (6) and (7), by Kwee, Muller and Westerhout (1954) and by Kerr (1962). The two sets of measurements, made in opposite halves of the Galaxy, agree rather well, and in general the behaviour of the circular galactic rotation at distances of from about three to eight kpc from the centre is reasonably well established. At small distances from the centre sin λ becomes small and $\omega(R)$ cannot be determined as reliably.

(iv) To determine $\omega(R)$ from 21 cm line observations, it is necessary that (*a*) R_0 and ω_0 are known. Errors in these parameters will produce both scale and zero point errors in the derived $\omega(R)$, but will not change its general form. (*b*) There is H I at, or very near, the point of maximum radial velocity along a given line of sight. (*c*) The motion of the gas is circular, with no radial motion and with no significant systematic peculiar motions in any particular region. (*d*) The random velocity of H I clouds is known, and can be allowed for in the analysis.

Probably none of these conditions is completely fulfilled, the result being to introduce some uncertainty in the derived $\omega(R)$. The effects of these uncertainties, and the possibility of large-scale deviations from circular motion, will be discussed in later sections.

LARGE-SCALE GALACTIC STRUCTURE

(i) A major objective of 21 cm line work is to provide a picture of the density distribution of neutral hydrogen in the plane of the Galaxy. The starting point for all 21 cm line studies of galactic structure is the galactic rotation curve, $\omega(R)$. An observed line profile gives the radial velocities of major H I emitting regions in a particular direction. Distances to the H I regions are then obtained from the observed velocities and the adopted galactic rotation curve. The integrated area of a line profile, or of a feature within a line profile, is a measure of N_H. N_H can then be converted to H I density by estimating the extent of the H I complex from the shape of the line profile and the galactic rotation curve or by adopting from other considerations an extent along the line of sight of, say, a spiral arm.

In practice, the procedure is complicated by a number of factors. Uncertainties in $\omega(R)$ obviously carry over directly into uncertainties in any derived neutral hydrogen distribution. Local systematic deviations from the general galactic rotation may cause emitting clouds to be placed at incorrect distances and thus somewhat distort the derived distribution. Any general, large-scale deviation from circular motion again distorts the distribution. Finally, the random motions within H I clouds, or cloud complexes, broadens the line profile and makes the true velocity distribution more difficult to determine. Unfortunately, however, the magnitude of random cloud motions may differ from one region of sky to another, or along the line of sight in a given region. The average magnitude of these random motions is known reasonably well and can be allowed for in the analysis, but the variations mentioned above may again distort the derived gas distribution.

The inner region of the Galaxy ($R < R_0$) presents an additional difficulty in analysis because of the ambiguity in radial velocity. Along any line of sight from the Sun there are two regions, equidistant from the centre, at which the gas will have the same radial velocity due to galactic rotation. Sorting out this ambiguity is difficult and uncertain.

(ii) Very extensive studies of the neutral hydrogen distribution in the plane of the Galaxy have been made at Leiden (van de Hulst, Muller and Oort, 1954; Westerhout, 1957) and at Sydney (Kerr, Hindman and Gum, 1959). The Leiden observations cover roughly galactic longitudes 320° to 240° (old coordinates), while the Sydney workers concentrated on those longitudes unavailable to Leiden. Most of the analysis in the above reference is limited to the outer region of the Galaxy,

$R > R_0$. Schmidt (1957) has made a detailed and complicated analysis of the gas distribution in the inner region, $R < R_0$.

From these studies there has come an excellent first picture of the large-scale distribution of neutral hydrogen in the Galaxy. The hydrogen is clearly distributed in lanes, or spiral arms, and a number of such arms can be recognized and followed for considerable distances around the centre. The fact that such a picture of the distribution, with well-defined general features, is obtained indicates that the problems of interpretation mentioned in 3(i) are not enough to swamp the gross features of hydrogen distribution. On the other hand, the detailed distribution appears quite complex. Some of this apparent complexity is undoubtedly due to the factors mentioned above—systematic deviations from general galactic rotation, variations in random motion, etc.

The current picture of neutral hydrogen distribution shows several features which suggest that not all of the assumptions which have been made in the analysis—and which in the present state of our knowledge must be made—are completely valid. The best defined 'spiral arms' appear very nearly circular about the galactic centre, rather than spiral. In the vicinity of the Sun, where analysis is especially difficult because of the small radial velocities and resultant larger relative effect of random velocities, there is a suspicious symmetry with respect to the Sun. Finally, the Leiden and Sydney portions of the distribution do not fit together as well as might be hoped.

In spite of these unresolved problems, the general picture of the neutral hydrogen distribution which has come from the work at Sydney and Leiden is a most important and remarkable contribution to the study of galactic structure. Further refinements should lead to better knowledge of random motions and of systematic deviations from circular galactic rotation.

(iii) Kerr (1962) has examined in some detail the fit between the Leiden and Sydney models. He believes that use of the Leiden circular velocity model leads to an implausible spiral structure in the southern Milky Way. He finds that one way to produce a better fit on the two sides of the Galaxy is to postulate a radial expansion, in addition to the rotation, amounting to seven km/s at the Sun's distance. An alternative possibility, leading to similar results, is a local outward systematic motion of the Sun and gas in the solar neighbourhood of 7 km/s. There

is at present no direct evidence of such motion, except the improvement it makes in the appearance of the derived neutral hydrogen distribution. It is an interesting possibility which needs further investigation.

Some local deviations from circular motion must certainly exist. Westerhout (1957) and others have pointed out regions in which systematic motions appear to be present. However, about the only way at present to distinguish systematic motions is by postulating that the distribution of neutral hydrogen is smooth in the Galaxy and then looking for regions which, on the basis of the assumption of no systematic motions other than galactic rotation, seem to violate this postulate. There are unfortunately no reliable distance criteria for 21 cm line emitting regions, independent of the galactic rotation curve, which is itself derived from 21 cm line observations with the aid of the assumption we wish to test.

(iv) The 21 cm studies referred to above have shown that the bulk of the interstellar gas is confined to a very thin disk. The thickness of the disk, between points where the hydrogen density falls to half its maximum value, is of the order of 200 parsecs. The disk is also remarkably flat, except in the outermost regions. There, distortions of the disk have been noted, amounting in some cases to several hundred parsecs (Burke, 1957; Kerr, 1957; Westerhout, 1957). The distortion is of the form of a general bending down of the hydrogen layer in the outer region of the Galaxy, in the quadrant centred around longitude 240°, and a similar upward bend in the opposite direction. Several explanations of this distortion have been proposed (above references and Kahn and Woltjer, 1959) but no thoroughly convincing explanation has yet been presented.

The flatness of the hydrogen layer has been used, along with other radio and optical data, to make a new determination of the principal plane of the Galaxy (Blaauw, Gum, Pawsey and Westerhout, 1960). A new system of galactic coordinates based in large part on this determination of the H I plane, has been adopted by the I.A.U.

THE CENTRAL REGION OF THE GALAXY

(i) The central part of the Galaxy, within about 3 kpc of the galactic centre, exhibits some interesting and complex features.

In 1957 it was found (van Woerden, Rougoor and Oort, 1957) that a spiral arm about 3 kpc from the centre was expanding with a velocity of some 50 km/s, as well as taking part in the general galactic rotation. Rougoor and Oort (1960) have subsequently made an extensive investigation of the distribution and motion of neutral hydrogen in the central region, using emission and absorption line profiles obtained with the 82-ft telescope at Dwingeloo. They have developed the following picture. In the centre of the Galaxy is a disk of neutral hydrogen. The hydrogen density at the centre must be very high, perhaps 1,000 atoms/cm^3 at a distance of 10 pc., and drops off rapidly with increasing distance from the centre until there is practically no gas at a distance of 300 to 350 pc. From this distance to about 500 pc there appears to be very little neutral hydrogen. At 500 pc a ring of neutral hydrogen about 100 pc wide appears. The ring and central disk have a thickness of only about 80 pc, and both are rotating with high velocity—the ring at some 265 km/s, the disk at velocities up to about 220 km/s near its edge. Neither the disk nor the ring appears to be expanding.

Between the Sun and the centre of the Galaxy is a spiral arm, the so-called 3 kpc arm, at about that distance from the galactic centre. It has been traced over some 90° of galactocentric longitude. This arm participates in galactic rotation, with a circular velocity of about 200 km/s. It is also expanding outward from the centre with a velocity of 50 km/s. The thickness of the arm is 120 pc—only one half the thickness of the neutral hydrogen layer further from the centre. On the far side of the galactic centre gas appears to be streaming outward with velocities of 100 to 200 km/s.

Rougoor and Oort compute that, from the observed rate of expansion, all gas in the central region should be removed in the order of 10^7 to 10^8 years if there is no replenishment. Since this time is much shorter than the age of the Galaxy there must be replenishment of gas in the central region, probably by gas streaming into the nuclear region from the galactic halo. Pariiski (1961) has discussed a possible exchange of gas between the halo and the nuclear region of the Galaxy.

(ii) Some other phenomena associated with this central region of the Galaxy may have a bearing on the problem of neutral hydrogen distribution and motion. The intense continuum radio source Sagittarius A is generally considered to be at the centre of the Galaxy, primarily because its position coincides so closely with that of the centre, and absorption studies place its distance at about that of the galactic centre. Drake (1959) investigated this source at 3·75 cm wavelength, with relatively high resolution. He found that it consists of two small diameter sources within 15 pc of the galactic centre. In addition he found two sources in the galactic plane symmetrically placed about 80 pc on either side of the centre. Drake interprets these latter two sources as being the tangential points of a ring of continuum emission. All four of the continuum sources lie within the central disk of neutral hydrogen.

Westerhout (1958) has analysed observations of the galactic continuum radiation to obtain a model of the distribution of ionized hydrogen in the Galaxy. He finds practically no ionized hydrogen within 3 kpc of the galactic centre, and a strong, sharp peak in the ionized hydrogen density about 3·5 to 4 kpc from the centre, just outside the rapidly expanding H I arm. The relationship, if any, between this ionized hydrogen concentration, the complex source Sagittarius A, and the complex distribution and motion of neutral hydrogen in the central hydrogen in the central region is not fully understood. Further investigations can be expected to lead to exciting results relating to the dynamics and driving forces of the Galaxy.

LOCAL STRUCTURE OF NEUTRAL HYDROGEN

(i) Studies of local structure in the solar neighbourhood are difficult at low galactic latitudes because the radial velocity of galactic rotation is small and therefore does not provide much assistance in separating hydrogen emission regions at different distances. Most investigations of local structure have therefore utilized 21 cm line observations at intermediate and high galactic latitudes.

(ii) Heeschen and Lilley (1954) studied the neutral hydrogen distribution as a function of galactic latitude at the longitudes of the galactic centre and anti-centre. They found secondary maxima in the hydrogen distribution at an angle of about 20° to the plane of the Galaxy which might be associated with the well-known Gould's belt phenomena. Davies (1960) concluded from a more extensive study of 21 cm line observations out of

the plane of the Galaxy, that the Sun is probably situated in a local cloud complex of neutral hydrogen, dust and young stars. The so-called Gould's belt is a manifestation of this local system. The strong secondary concentrations of neutral hydrogen observed at latitude $+20°$ in the longitude of the galactic centre and $-20°$ in the longitude of the anti-centre can be interpreted in other ways, however—for example, as being associated with the well-known cloud complexes of Ophiuchus and Taurus, which might be completely unrelated—and other groups have found no strong evidence for a local system of hydrogen. This question therefore must remain open.

(iii) The radial velocity of local hydrogen has been studied by various investigators. McGee and Murray (1961) find that in the galactic plane hydrogen is streaming away from the Sun, in the directions of the centre and the anti-centre, with a velocity of about $+6$ km/s, while perpendicular to the plane hydrogen is streaming into the solar region, from above and below, with about the same velocity. Erickson, Helfer and Tatel (1959) find evidence that the local motion is in general circular, about the galactic centre, with some complex non-circular components.

There is as yet, however, no clear picture of either the distribution or the motion of neutral hydrogen in the solar neighbourhood. Nor has any full-scale attempt yet been made to obtain the solar motion relative to local hydrogen. Lilley and Brouer are about to start a long-term observing programme aimed at this latter problem. A great deal more work on local structure is obviously needed.

(iv) Almost all workers agree on the thickness of the hydrogen layer in the solar neighbourhood, about 200 pc, and on the density of neutral hydrogen in the solar neighbourhood, about 0·5 atom/cm^3. There is no agreement, however, on the position of the Sun—some investigators placing it 50 pc above the plane, others in, or slightly below, the plane.

GAS AND DUST

(i) The relationship between interstellar dust and neutral hydrogen has been studied by a number of investigators (Lilley, 1955; Heeschen, 1955; Bok, Lawrence and Menon, 1955; Davies, 1956; and some of the Leiden papers previously referred to). Lilley (1955) was the first to show that in general a positive

correlation exists between the amounts of neutral hydrogen and dust present in a given direction. He found an average gas to dust ratio of about 100 : 1. The correlation between gas and dust is only rough, however, and the gas to dust ratio appears to vary considerably from one region to another. Some of this may be due to observational difficulties, and to the difficulty of determining the gas and dust content along identical lengths of path in a given region. In the case of several heavily obscured regions, however, there definitely seems to be a lower 21 cm line intensity than one would predict from the gas to dust ratio, and in some cases the intensity is even lower than that in neighbouring unobscured regions. Attempts have been made to explain this by postulating that in heavily obscured regions atomic hydrogen is more easily converted to moleculer hydrogen with the aid of the high concentration of dust, thereby reducing the atomic hydrogen content.

(ii) Davies (1956) made a detailed study of a region of Auriga in which the neutral hydrogen intensity was significantly lower than that in surrounding regions. He found that the low intensity H I 'cloud' coincided in position with an obscuring dust cloud. From an analysis of the difference in hydrogen line profiles in neighbouring regions and in the region of the cloud, Davies found that the kinetic temperature in the cloud was 60°K—significantly lower than the average kinetic temperature of interstellar neutral hydrogen. He also deduced a very high neutral hydrogen density, of about 250 atoms/cm^3, and a gas to dust ratio of 300. Davies has found several other such 'clouds' which yielded similar results. He postulates that the presence of molecular hydrogen in these clouds has cooled down the neutral hydrogen to well below the average temperature of the interstellar medium. This mechanism may help explain the apparent break-down of correlation between gas and dust in heavily obscured regions.

CORRELATION BETWEEN NEUTRAL HYDROGEN AND STELLAR ASSOCIATIONS AND CLUSTERS

(i) Heeschen and Drake (1956) have made observations which suggest that there is a cloud, or cloud complex, of neutral hydrogen associated with the Pleiades star cluster. Drake (1958) observed other clusters as well, and found evidence for neutral

hydrogen associated with several of them. Helfer and Tatel (1959), on the other hand, interpreted their observations in the region of the Pleiades as indicating no association between neutral hydrogen and the star cluster. This question remains to be resolved.

(ii) A number of investigations have been made of the relation between neutral hydrogen and stellar OB associations (Howard, 1958; Kassim, 1961; Menon, 1956, 1958; Raimond, 1957; Dieter, 1960; Wade, 1957; Matthews, 1956). In some cases there clearly seems to be interstellar hydrogen associated with the stellar association. In other cases, there does not appear to be hydrogen connected with the stellar association, or observations by different observers are not in agreement. Thus no clear-cut general picture has yet emerged regarding the relationship between interstellar hydrogen and stellar associations.

(iii) The general procedure which has been used in attempts to relate interstellar hydrogen with stellar clusters or associations has been to look for an excess of hydrogen line emission at the position, and at the radial velocity of the cluster or association. This procedure is valid in principle, but runs into a number of difficulties in practice. Most clusters and associations are in or near the galactic plane, where there is generally a long path length in neutral hydrogen. The 21 cm line profiles are therefore usually rather complex, and only a small portion of the total hydrogen line emission observed from the region containing a stellar cluster or association may be expected to arise from hydrogen associated with the cluster or association. Separation of this component, if it exists, from the general interstellar field may be quite difficult.

CONCLUSION

It is clear that a great deal more work is required in the field of galactic 21 cm line studies. Refinement, and a general filling in of details, of the excellent picture of gross neutral hydrogen distribution developed by the Leiden and Sydney groups is required. Problems of the local structure and motion of neutral hydrogen, and of the relationships between H I and dust, stellar clusters and associations, and H II regions are essentially barely touched thus far. Much remains to be learned also of the small-scale structure of neutral hydrogen, that is, studies of the interstellar cloud structure.

The physical conditions of interstellar gas, especially temperature and random motions, need much additional study. It has generally been assumed in hydrogen line work that the kinetic temperature of hydrogen is, on the average, constant, with a value of about 125°K. The principal evidence in favour of this assumption is the fact that brightness temperatures in the hydrogen line up to about this value, but never above it, have been observed. Davies has shown, however, that, in some regions at least, considerably lower temperatures may prevail. Further studies of random motions in clouds, or cloud complexes, and of the small-scale cloud structure of the interstellar medium, are important. Knowledge of these parameters is important to almost all other galactic hydrogen line work, because they effectively smear out, or distort, the true systematic velocity distribution of interstellar hydrogen.

Most of the problems require high resolution in both frequency and direction, and very stable receivers. The question of whether there is a detectable amount of neutral hydrogen in the galactic halo remains open also. Here a small telescope might be useful, but the receiver must be extremely sensitive and stable.

REFERENCES

Blaauw, A., Gum, C. S., Pawsey, J. L., and Westerhout, G., 1960, *M.N.R.A.S.*, **121**, 123.
Bok, B. J., Lawrence, R. S., and Menon, T. K., 1955, *P.A.S.P.*, **67**, 108.
Burke, B. F., 1957, *A.J.*, **62**, 90.
Davies, R. D., 1956, *M.N.R.A.S.*, **116**, 443.
Davies, R. D., 1960, *M.N.R.A.S.*, **120**, 483.
Dieter, N. A., 1960, *Ap. J.*, **132**, 49.
Drake, F. D., 1958, Thesis, Harvard Univer.
Drake, F. D., 1959, *Nat. Rad. Ast. Obs. Ann. Rpt.*
Erickson, W. C., Helfer, H. L., and Tatel, H. E., 1959, *I.A.U.*, Symp. No. 9, pp. 390.
Gum, C. S., Kerr, F. J., and Westerhout, G., 1960, *M.N.R.A.S.*, **121**, 132.
Heeschen, D. S., 1955, *Ap. J.*, **121**, 569.
Heeschen, D. S., and Drake, F. D., 1956, *A. J.*, **61**, 5.
Heeschen, D. S., and Lilley, A. E., 1954, *Pub. Nat. Acad. Sci.*, **40**, 1095.
Helfer, H. L., and Tatel, H. E., 1959, *Ap. J.*, **129**, 565.
Howard, W. E., III, 1958, *A.J.*, **63**, 50.
Hulst, H. C., van de, Muller, C. A., and Oort, J. H., 1954, *B.A.N.*, **12**, 117, No. 452.
Kahn, F. D., and Woltjer, L., 1959, *Ap. J.*, **130**, 705.
Kassim, M. A., 1961, *Ap. J.*, **133**, 821.
Kerr, F. J., 1957, *A.J.*, **62**, 93.
Kerr, F. J., 1962, *M.N.R.A.S.*, **123**, 327.

Kerr, F. J., Hindman, J. V., and Gum, C. S., 1959, *Aust. J. Phys.*, **12**, 270.
Kwee, K. K., Muller, C. A., and Westerhout, G., 1954, *B.A.N.*, **12**, 211, No. 458.
Lilley, A. E., 1955, *Ap. J.*, **121**, 559.
Matthews, R. T., 1956, Thesis, Harvard Univ.
Menon, T. K., 1956, *A.J.*, **61**, 9.
Menon, T. K., 1956, *Ap. J.*, **127**, 28.
McGee, R. X., and Murray, J. D., 1961, *Aust. J. Phys.*, **14**, 260.
Pariiski, Yu. N., *Sov. Ast. A. J.*, **5**, 280, 1961.
Raimond, E., 1957, *B.A.N.*, **13**, 269, No. 475.
Rougoor, G. W., and Oort, J. H., 1960, *Proc. Nat. Acad. Sci.*, **46**, 1.
Schmidt, M., 1956, *B.A.N.*, **13**, 15, No. 468.
Schmidt, M., 1957, *B.A.N.*, **13**, 247, No. 475.
Wade, C. M., 1957, *A.J.*, **62**, 148.
Westerhout, G., 1957, *B.A.N.*, **13**, 201, No. 475.
Westerhout, G., 1958, *B.A.N.*, **14**, 215, No. 488.
van Woerden, H., Rougoor, W., and Oort, J. H., 1957, *C.R.*, **244**, 1691.

10

EXTRAGALACTIC HYDROGEN LINE STUDIES

V. C. REDDISH

Theory

The neutral hydrogen atom has two hyperfine energy levels differing in energy by $9{\cdot}403 \times 10^{-18}$ ergs. In the lower level the magnetic moments of proton and electron are antiparallel, in the upper level they are parallel. The natural lifetime in the upper state is 10^7 years, but intervals between collisions in the interstellar gas are much less than this. Consequently most transitions, up and down, occur in collisions; but a few downward transitions occur by radiating a quantum $h\nu_0 = 9{\cdot}403 \times 10^{-18}$ ergs, corresponding to radiation of frequency $\nu_0 = 1420{\cdot}40$ mc/s, $\lambda = 21$ cm. The natural line width is very small, but thermal and cloud velocities of a few km/s produce doppler broadening; at this frequency a radial velocity of 1 km/s produces a doppler shift of 4·73 kc/s. Thus the number of atoms radiating at a given *observed* frequency is the number moving with the corresponding radial velocity, Fig. 10/1.

Since equilibrium between the energy states is maintained by collisions the Boltzmann equation can be applied.

$$\frac{n_2}{n_1} = \frac{g_2}{g_1} e^{-\frac{h\nu_0}{kT}} \quad . \quad . \quad . \quad . \quad (1)$$

where n_2 and n_1 are the numbers of atoms in the upper and lower states respectively, g_2 and g_1 are the corresponding statistical weights (the number of pigeon-holes on each energy shelf!) and T the excitation temperature; in the conditions dealt with here this is given closely by the kinetic temperature. It is useful to note here that

$$\frac{h\nu_0}{k} = 0{\cdot}07^\circ\mathrm{K} \quad . \quad . \quad . \quad . \quad (2)$$

and thus for temperatures not less than a few degrees Kelvin

$$\frac{n_2}{n_1} \approx \frac{g_2}{g_1} = 3 \quad . \quad . \quad . \quad . \quad (3)$$

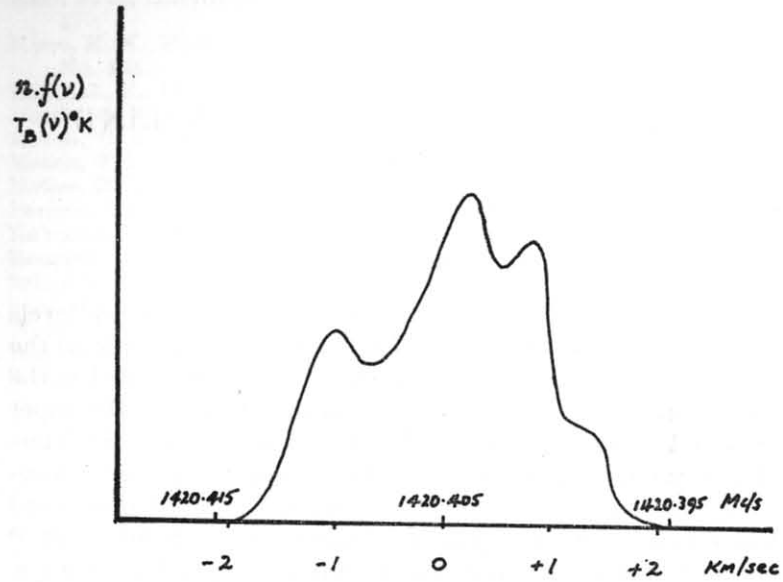


FIG. 10/1.—The distribution of neutral hydrogen atoms as a function of radial velocity. n is the total number of hydrogen atoms, $f(v)$ the fraction with radial velocities necessary to doppler shift 1420·405 mc/s into unit frequency range at v. $T_B(v)$ is the observed brightness temperature.

Let us now consider the radiation from a cloud of neutral hydrogen. Let

κ_ν = absorptivity/unit volume

ε_ν = emissive power/unit volume/unit frequency range

τ_ν = optical depth

Then considering a column of the gas of unit cross-sectional area, along the line of sight (Fig. 10/2) we have

$\kappa_\nu\, ds$ = absorptivity of length $ds = d\tau_\nu$

$\varepsilon_\nu\, ds$ = emissive power of length

$$\text{ds/unit frequency range} = \frac{\varepsilon_\nu}{\kappa_\nu} d\tau_\nu$$

$$= \varepsilon_{B_\nu}\, d\tau_\nu$$

by Kirchoff's law, where ε_{B_ν} is the emissive power/unit frequency range of a black body at the same temperature.

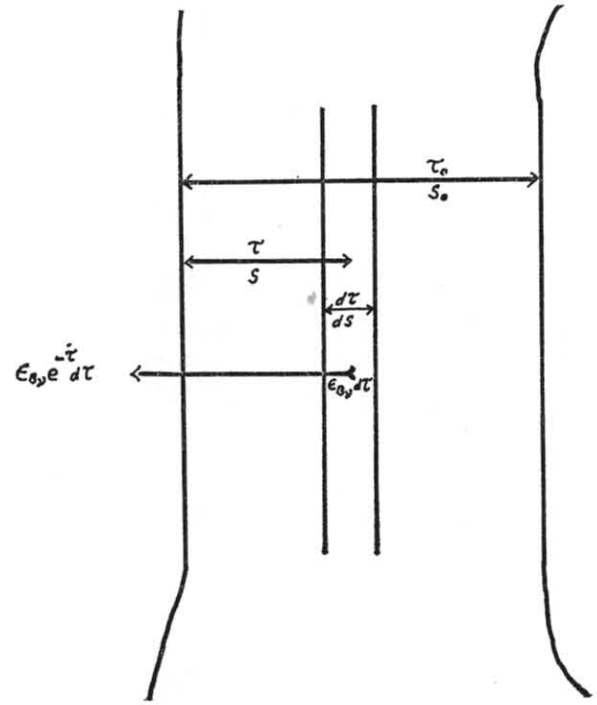


FIG. 10/2.—The emission from an elementary cross-section of neutral hydrogen gas.

Now $\varepsilon_{B_\nu}\, d\tau_\nu$ emitted at depth τ_ν is reduced to $\varepsilon_{B_\nu}\, d\tau_\nu\, e^{-\tau_\nu}$ on reaching the surface; hence the emission at the surface from the whole column is given by

Observed emission per unit frequency range

$$= \int_0^{\tau_{0\nu}} \varepsilon_{B_\nu}\, e^{-\tau_\nu}\, d\tau_\nu \quad . \quad . \quad . \quad (4)$$

If the temperature and density are uniform along the column then ε_{B_ν} is constant, then the observed emission per unit frequency range

$$= \varepsilon_{B_\nu}(1 - e^{-\tau_{0\nu}}) \quad . \quad . \quad . \quad (5)$$

But for wavelengths as long as 21 cm, ε_{B_ν} is proportional to kT; thus the brightness temperature (i.e. the temperature at which a black body would have to be to give the same emission) is

$$T_B(\nu) = T(1 - e^{-\tau_{0\nu}}) \quad . \quad . \quad . \quad (6)$$

For small optical depths τ_{0_ν} this becomes

$$T_B(\nu) = T\tau_{0\nu}. \quad . \quad . \quad . \quad (7)$$

Now

$$\kappa_\nu = \left(n_1 \alpha_{\nu_0} - n_2 \frac{g_1}{g_2} \alpha_{\nu_0}\right) f(\nu) \quad . \quad . \quad (8)$$

where α_{ν_0} is the atomic absorption coefficient, $\frac{g_1}{g_2}\alpha_{\nu_0}$ is the coefficient for stimulated emissions (negative absorption), and $f(\nu)$ is the fraction of the neutral hydrogen atoms which have the radial velocities necessary to doppler shift 1420·405 mc/s into unit frequency range at ν, Fig. 10/1.

Thus using (1)

$$\kappa_\nu = n_1 \alpha_{\nu_0} (1 - e^{-\frac{h\nu_0}{kT}}) f(\nu)$$

$$= n_1 \alpha_{\nu_0} \frac{h\nu_0}{kT} f(\nu) \quad . \quad . \quad . \quad . \quad (9)$$

Hence

$$T_B(\nu) = \frac{\alpha_{\nu_0} h\nu_0}{k} f(\nu) \int_0^{S_0} n_1 \, ds \quad . \quad . \quad (10)$$

and using (3)

$$T_B(\nu) = \frac{\alpha_{\nu_0} h\nu_0}{4k} f(\nu) \, \sigma_H \quad . \quad . \quad . \quad (11)$$

where $\sigma_H = \int_0^{S_0} n ds$ is the surface density of neutral hydrogen in atoms per cm²; that is to say, the total number of neutral hydrogen atoms along the column of unit cross sectional area. Thus $nf(\nu)$ on Fig. 10/1 can be replaced by a scale of $T_B(\nu)$. Over the width of the broadened line $f(\nu)$ varies rapidly with frequency and the quantity $\left(\frac{\alpha_{\nu_0} h\nu_0}{4k}\right)$ can be regarded as a constant; since $f(\nu)$ is zero outside the line we have

$$\int f(\nu) \, d\nu = \int_0^\infty f(\nu) \, d\nu = 1 \quad . \quad . \quad (12)$$

and hence

$$T_B(\nu) \, d\nu = \text{constant} \times \sigma_H \quad . \quad . \quad (13)$$

Inserting the relevant quantities gives

$$T_B(\nu) \, d\nu = 2{\cdot}58 \times 10^{-15} \sigma_H \quad . \quad . \quad (14)$$

where σ_H is again in atoms/cm².

CALCULATION OF NEUTRAL HYDROGEN CONTENT OF A GALAXY

Let us now turn to the problem of observing a whole galaxy. Here we are dealing with a thin circular disk of hydrogen in non-uniform rotation about its axis; the stars in the galaxy concern us only in so far as their mass affects the motions of the gas and their radiations affect its temperature. The motion of the gas will depend on distance from the centre. On the evidence available it is unlikely to depend significantly on position angle. The circular velocity as a function of distance from the centre in M31 is given in Fig. 10/3. There may also be a small component

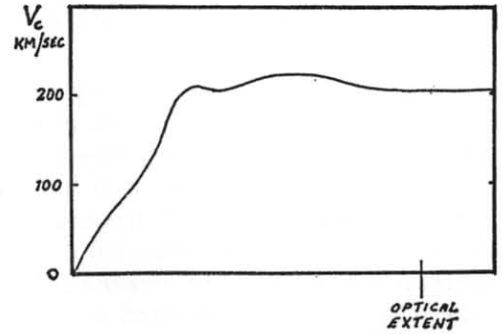


FIG. 10/3.—Circular velocity plotted against distance from the centre of M31.

of radial velocity—a few km/s—if the suspicions of some observers of our own Galaxy are correct. The surface density distribution of the gas depends on distance from the centre and also to a considerable extent on position angle: for example, the observations made both at Jodrell Bank and at Leiden show considerable differences between the NE and SW halves of M31,

For our immediate purpose, however, we shall consider the relative distribution along the SW major axis, Fig. 10/4, to be typical.

Let us then suppose for the moment that Figs. 10/3 and 10/4 represent the motions and density distribution of the hydrogen, relative to the marked optical extent, in any other galaxy we observe, and consider the observation of a galaxy with inclination i (angle between the line of sight and the normal to the

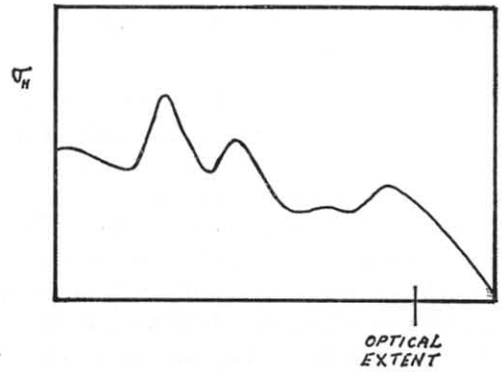


FIG. 10/4.—The surface density of neutral hydrogen plotted against distance along the SW major axis of M31.

plane of the galaxy), radial (redshift) velocity V_0, and optical extent θ min. of arc, using a radio telescope of known beamshape and bandshape. What brightness temperature would we expect to observe for a given total mass of hydrogen in the galaxy, and how would it vary with the frequency of observation? Suppose our radio telescope is directed at the galaxy in question; a certain deflection of the recording instrument is produced. If the telescope is now directed at a black body of larger angular extent than the telescope beam, what must be the temperature of the black body to produce the same deflection of the recording instrument? This is the temperature which will subsequently be referred to as 'observed brightness temperature'; it is in principle the method by which we calibrate our instrument by observations on our own Galaxy. The instrumental parameters involved are the bandshape and beamshape, Figs. 10/5 and 10/6. (If the beam is not circularly symmetric Fig. 10/6 will have to be replaced by a contour map giving

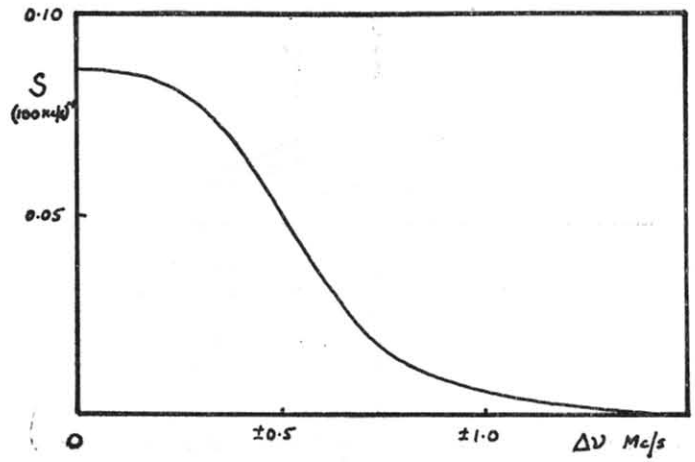


FIG. 10/5.—The instrumental bandshape. S is the fraction of the energy in the receiver band which falls in an element $d\nu$ wide (here 100 kc/s) at $\Delta\nu$ mc/s from the centre of the band, when observing a source which is equally bright at all frequencies.

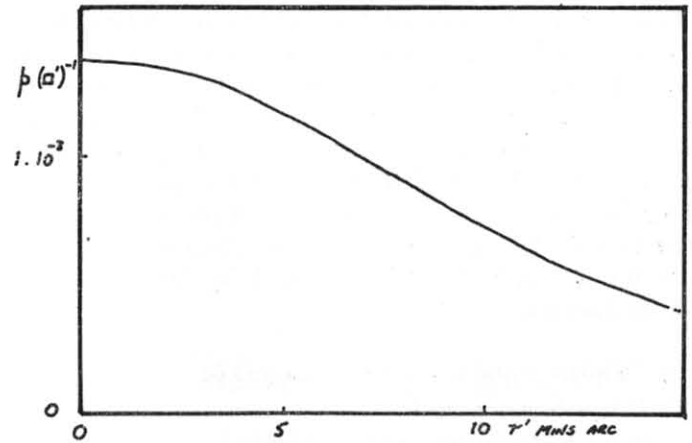


FIG. 10/6.—The beamshape. p gives the fraction of the energy received by the telescope, from a sky at uniform temperature, which is received in unit solid angle (here 1 sq. min. arc) at an angle r' from the centre of the beam.

$p(x'\ z')\ (\square')^{-1}$ as a function of two rectangular co-ordinates x' and z', $(x'^2 + z'^2 = r'^2)$.) We shall take the major axis of the galaxy to be the x axis (Fig. 10/7) and the beam centre at $x_0\ z_0$.

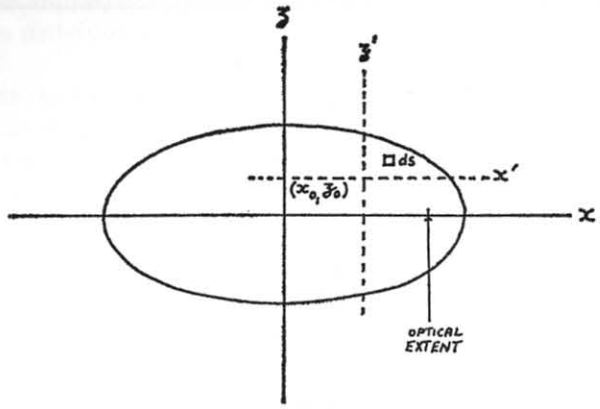


FIG. 10/7.—The coordinate systems relative to the aerial beam, and to the galaxy.

DETAILS OF THE COMPUTATION

The calculation then proceeds in the following manner. Choose a size of elementary area on the galaxy $\delta x \delta z = \delta s$, small enough to ensure that the rotational velocity does not change over it by more than a small fraction of the receiver bandwidth, i.e. so that S can be regarded as constant for the area, and small enough to ensure that $p(r')$ does not change much over it. Choose a scale for σ_H (Fig. 10/4). Note that the abscissae of Figs. 10/3 and 10/4 now refer to the optical extent of the galaxy being observed. Let y be the co-ordinate perpendicular to x *in the plane of the galaxy*. Then $y = z \sec i$. The results for a particular galaxy would be tabulated in the following manner.

(i) Tabulate x and z at intervals δx and δz

(ii) Calculate $y = z \sec i$

(iii) Tabulate δs

(iv) Calculate $r = (x^2 + y^2)^{\frac{1}{2}}$

(v) Calculate $\cos\theta = \dfrac{x}{r}$; θ = p.a. from major axis, in the plane of the galaxy

(vi) Tabulate $V_c\,(r)$ from Fig. 10/3

(vii) Tabulate $\sigma_H(r)$ from Fig. 10/4

(viii) Calculate $V_0 + V_c \cos\theta \sin i$ = line of sight velocity of element ds

(ix) Calculate Tdv from $\sigma_H \sec i$; ($Tdv = 2{\cdot}58 \times 10^{-15}$. $\sigma_H \sec i$ for dv in c/s. Here we must use the adopted dv of Fig. 10/4, 100 kc/s, so that $Tdv = 2{\cdot}58 \times 10^{-20}$ °K(100 kc/s)). This is then the brightness temperature of the element ds diluted over a rectangular band 100 kc/s wide

(x) Calculate the frequency ν for the doppler shift corresponding to (viii) from

$$\nu = 1420{\cdot}405 - 4{\cdot}73\,(V_0 + V_c \cos\sin i) \times 10^{-3}\ \text{mc/s}$$

(xi) Difference ν from the centre frequency of the observing band ν_1 to get $\Delta\nu = \nu - \nu_1$

(xii) Tabulate $S(\Delta\nu)$ $\left\{\begin{matrix}\text{units} \\ (\delta\nu)^{-1} \\ \text{here } (100\ \text{kc/s})^{-1}\end{matrix}\right\}$ from Fig. 10/5

(xiii) Tabulate $p(r')$ {units (area)$^{-1}$, here $(\square')^{-1}$} from Fig. 10/6

$$\left\{\begin{matrix}\text{n.b.} \\ \text{in Fig. 10/7,}\end{matrix}\quad\begin{matrix}x' = x - x_0 \\ z' = z - z_0\end{matrix}\right\} r' = (x'^2 + z'^2)^{\frac{1}{2}}$$

(xiv) Calculate $T \times S \times p \times \delta s = T_{ds}$ (product of (iii), (ix), (xi), (xii)). This is the contribution to the observed brightness temperature by the elementary area ds

(xv) Sum T_{ds} over the whole beam to get the observed brightness temperature $T_{\text{obs}} = \sum_{\text{beam}} T_{ds}$ at the frequency of observation ν_1; i.e. sum of column (xiv)

(xvi) If desired, repeat at various ν_1 to obtain the spectrum of H line emission by the galaxy,

$$T_{\text{obs}}\,(\nu_1)$$

(xvii) In the case of a galaxy larger than the beam, repeat at various $x_0 z_0$.

The analytical problem is essentially to adjust Figs. 10/3 and 10/4 until the calculated $T_{\text{obs}}\,(\nu_1 x_0 z_0)$ *agree with those actually observed.*

The total amount of hydrogen corresponding to $T_{\text{obs}}(\nu_1)$ is determined from

$$\mathcal{M}_H = \int_0^R 2\pi r \sigma_H dr$$

MEASURED NEUTRAL HYDROGEN MASS OF EXTERNAL GALAXIES

A brief summary of published results is given in Table 10/I. They show that the fraction of the total mass of a galaxy which is in the form of neutral hydrogen is larger in irregular than in spiral galaxies. In addition to the data for which references are given further substantial observations made at Harvard and at Jodrell Bank are nearing publication.

TABLE 10/I

Galaxy	Type	d kpc	M_H 10^9 suns	M 10^9 suns	M_H/M	Reference
IC 1613	Irr	630	0·05	0·33	0·15	(1)
SMC	Irr	631	0·75	6	0·125	(2), (3), (4)
NGC 6822	Irr	500	0·15	1·5	0·10	(1)
LMC	Irr	631	1·1	14	0·08	(2), (4)
M 33	Sc+	630	1·0	18	0·06	(5)
M 82	Irr	2600	1·1	20	0·055	(1)
M 101	Sc−	2600	3·0	140	0·02	(5)
Galaxy	Sb?		1·5	70	0·02	(6)
M 31	Sb−	630	4·5	380	0·01	(6)
M 81	Sb−	2600	1·5	150	0·01	(1)

REFERENCES

1. Volders and Högbom, 1961, *B.A.N.*, **15**, 311.
2. Kerr, Hindmann, and Robinson, 1954, *Aust. J. Phys.*, **7**, 297.
3. Kerr and de Vaucouleurs, 1956, *Aust. J. Phys.*, **9**, 90.
4. de Vaucouleurs, 1960, *Ap. J.*, **131**, 265.
5. Volders, 1959, *B.A.N.*, **14**, 323.
6. van de Hulst, Raimond, and van Woerden, 1957, *B.A.N.*, **14**, 1.

11

THE MAGNETIC FIELD OF THE GALAXY

R. D. DAVIES

MAGNETIC fields play an important part in many galactic phenomena such as the polarization of starlight, the isotropy of cosmic rays, synchrotron radio emission and the retention of galactic spiral structure. That a systematic magnetic field should exist is evident from the regular distribution of the polarization of starlight in our own Galaxy (Hiltner, 1956) and in external galaxies, such as M31 (Hiltner, 1958) and M82 (Elvius and Hall, 1962) which suggests that the magnetic field is closely parallel to the spiral structure. However, the magnitude of the field cannot be measured by optical means and has hitherto been inferred from the observed intensity of radio emission and theories of the polarization of starlight. The value derived was 3×10^{-5} gauss (Woltjer, 1961).

A direct determination of the magnetic field can be made using the Zeeman splitting of the 21 cm line of neutral hydrogen (Bolton and Wild, 1957). In the presence of a longitudinal field the line is split into left- and right-hand circular components separated by 2·8 Mc/s per gauss. If the field is directed away from the observer, which we will call a positive field, the left-hand component is displaced to the higher frequency in an emission profile; the displacement is reversed for an absorption profile. The most sensitive determination of the magnetic field is obtained from the absorption spectra of the intense radio sources which have deep and narrow features due to individual clouds lying between the source and the observer.

The Zeeman programme at Jodrell Bank has included investigations of the deepest features in the absorption spectra of Cassiopeia A, Cygnus A and Taurus A and the 50 km/s feature in Sagittarius A due to the expanding arm 3 kpc from the galactic centre. A bright narrow emission feature at $\delta = +10°$ and R.A. $= 4^h20^m$ was also studied. The most significant results come from Cassiopeia A which lies in a direction which makes about 30° with the spiral arms. It has

two blended features in the Orion arm and two major features which are themselves blends of lines in the Perseus arm. The narrow line in the spectrum of Taurus A which lies in the anti-centre direction gives a useful limit to the radial component of the galactic magnetic field.

The observing technique consisted of switching between the two senses of circular polarization and measuring the frequency shift. This was achieved by subtracting the two polarized signals observed in a narrow frequency band which was swept across the absorption feature. The receiver was calibrated by radiating a circularly polarized signal into the primary feed of the aerial. Each feature was scanned many times and the results were then compared with the expected polarization difference profile to obtain the best fitting value of the magnetic field. The results are presented in Table 11/1 and show the best fitting value and the range of values contained within two standard deviations and therefore having a 95 per cent probability of including the actual value.

TABLE 11/I

Source	Velocity of feature (km/s)	l^{II}	b^{II}	Longitudinal component of magnetic field (units of 10^{-6} gauss)
Cassiopeia A	Orion arm −0·8	111·5	−0·2	−4·3 ± 3·8
	Perseus arm −38·1			+4·3 ± 9·2
	Perseus arm −48·2			−6·7 ± 3·0
Taurus A	10·3	184·5	−5·8	+6·4 ± 4·5
Cygnus A	−84·5	76·5	5·7	−7 ± 54
Sagittarius A	−53·4	0	0	+28 ± 110
Emission at $\delta = +10°$ R.A. $= 4^h20^m$		184·5	−26·9	−88 ± 130

The table shows that no magnetic fields greater than 7×10^{-6} gauss have been detected in any of the clouds in the Cassiopeia A and Taurus A absorption spectra. The field measured in the high frequency component of the Cassiopeia A Perseus arm absorption appears to have a significant value of $-6{\cdot}7 \times 10^{-6}$ gauss. The results of all the features when taken together are consistent with a general magnetic field parallel to the spiral

arms of -7×10^{-6} gauss; this field is directed in an anti-clockwise direction in the Galaxy. The significance of such positive results requires confirmation but it is in any case evident that the general galactic magnetic field must be less than about 7×10^{-6} gauss.[1]

These limits to the galactic magnetic field are considerably less than the fields mentioned previously and it may be asked if the field in some way avoids the particular clouds studied. Estimates based on the observed absorption spectra show the temperature of the absorbing clouds to be 60 to 100°K and the density to be about 20 hydrogen atoms per cm^3. The temperature is rather lower and the density is higher than that found in emission by van Woerden, Takakubo and Braes (1962); however the difference is not so great as to suggest that the magnetic field would interact differently with the two types of clouds and we thus conclude that the field if it is higher than the limits of the Zeeman experiment must avoid all clouds. This situation is also at variance with the observation that the magnetic fields penetrate dust clouds and produce the alignment of grains which give the polarization of starlight. It also does not agree with a current picture of a primordial magnetic field in which the gas was mixed in a tenuous state, and from which the clouds later condensed (Hoyle, 1953). Indeed on this view, because of the freezing into the magnetic field of the neutral hydrogen by the agency of a small concentration of electrons and positive ions the process of gravitational contraction would produce an *increase* of magnetic field in the clouds relative to the surrounding medium. .

The other radio astronomical observation which leads to information about galactic magnetic fields is the study of the synchrotron emission from the disk of the Galaxy. The information comes from the observed intensity of emission which gives the magnitude of the field under certain assumptions and from the plane of the associated linear polarization which at the point of emission is perpendicular to the magnetic field. Linear polarization has been observed in the emission from Taurus A at frequencies above 2000 Mc/s and in several areas in the Galaxy at a frequency of 400 Mc/s. Observations at a range of frequencies can be used to remove the effects of Faraday rotation in the interstellar medium and in the ionosphere. A

[1] But see postscript 2.

single-frequency observation of the emission near the galactic plane does not give a unique value of the field direction.

The flux of synchrotron emission can be related to the magnetic field by

$$J(\nu) = K\, N(>E)\, \nu^{-s}\, B^{1+s}$$

where K = constant, s is the radio spectral index of the emission and $N(>E)dE = C\,E^{-2s}$ is the energy spectrum of the cosmic ray electrons with energy E which gives a peak radio emission intensity at a frequency ν. Using a spectral index of 0·5 at a frequency of 80 Mc/s and the observed values of $J(\nu)$ and $N(>E)$ (Bierman and Davis, 1960; Meyer and Vogt, 1961; Earl, 1961) the magnetic field required is 3×10^{-5} gauss, which is a factor of five greater than the upper limit required by the Zeeman experiment.

A further clue to the inadequacy of the picture that synchrotron emission comes from an isotropic flux of electrons spiralling in a uniform magnetic field in the galactic disk was derived from a survey of the anticentre region of the Galaxy at 237 Mc/s by Davies and Hazard (1962). They found that the emission was largely broken up into irregular features some of which were associated with visible nebulosities supposed to be supernova remnants. If this nearby region which can be resolved with a 1° beam is typical of the disk then we must consider the disk composed of localized regions of emission probably the remnants of supernovae which exploded several tens of thousands of years ago and are now objects similar to the Cygnus loop and IC 443. It is unlikely that the emission regions contain excess cosmic ray electrons 10^4 years after their initial release and the enhanced emission comes from the normal flux of galactic cosmic ray electrons passing through the high magnetic fields of 10^{-3} to 10^{-4} gauss expected at the expanding envelopes of supernova-type explosions (Hoyle, 1960). The high general galactic field is no longer necessary to produce the observed level of emission from the disk. Since the emission depends on the magnetic field to the power $(1 + s)$ it can be described in terms of a hierarchy of regions varying in age and energy output from objects like Cassiopeia A and Taurus A through the Cygnus loop to extended areas with magnetic fields only several times that of the galactic background. Fields of 3×10^{-4} gauss would need to occupy ~2 per cent of space to explain the measured flux of synchrotron emission.

The theory of Davis and Greenstein (1951) required a magnetic field of 10^{-5} to 10^{-4} gauss to line up spheroidal dust particles and produce the polarization of starlight. They supposed that the spinning particles were aligned with their axes of rotation parallel to the magnetic field by the process of paramagnetic relaxation. However Henry (1958) has shown that by increasing the ferromagnetic content of the particles fields of as little as 10^{-7} gauss may be sufficient. Also Platt (1956) suggests that small particles 10Å in diameter can produce the observed optical effects; these can be aligned by fields of $\sim 10^{-6}$ gauss. Thus there are theories capable of explaining the optical polarization of starlight with fields less than those set by the limit of the Zeeman experiment.

It is of interest to consider the relevance of the present results to the problems of galactic cosmic rays. Cosmic rays can be held in the Galaxy by the general galactic magnetic field if the magnetic energy density is greater than the cosmic ray energy density. This requires that the magnetic field should be 6×10^{-6} gauss or greater. Accordingly it seems difficult to hold cosmic rays in the Galaxy with a general magnetic field of the size measured although the localized regions of higher field may help resolve the difficulty. This problem can be approached in a different way. High energy cosmic rays will be held in the Galaxy so long as the radii of curvature of their orbit is less than $\frac{1}{10}$ of the size of the halo (say 1 kpc) otherwise they will diffuse out. For an energy of 10^{18} eV a magnetic field of at least 10^{-6} gauss is required in the halo to give a radius of gyration of 1 kpc or less so that the cosmic ray will be held in the Galaxy. Linsley, Scarsi and Rossi (1962) have registered airshowers of energies up to 2×10^{19} eV. The incident protons responsible for the most energetic showers cannot be contained within the Galaxy by the magnetic fields suggested. Cosmic ray protons of energy greater than about 10^{18} eV will spend most of their lifetime in intergalactic space.

Another phenomenon in which interstellar fields play an important part is star formation from interstellar clouds. The condensation of a gas cloud in the presence of a magnetic field is retarded by the coupling between the magnetic field and the electrons and the consequent increase in magnetic pressure. Mestel and Spitzer (1956) have shown that with a field of 10^{-6} gauss condensation within a cloud as large as 200 solar masses

will only take place if sufficient dust is present to shield the gas from ionizing radiation and allow recombination of the charged particles. The results of the Zeeman experiment indicate that fields of this low value may exist in the interstellar neutral hydrogen clouds.

A general galactic magnetic field is often invoked to explain the stability of spiral arms of gas and dust in galaxies. Under the influence of differential rotation alone galaxies would show numerous tightly wound arms. Theoretical work on this problem is in an early state. Hoyle and Ireland (1961) considered a given arm to be a temporary phenomenon having an age of $\sim 3 \times 10^8$ years. In their picture the spiral arm field required to maintain stability was $\sim 5 \times 10^{-6}$ gauss.

It is concluded that if the results of the Zeeman experiment show that the general galactic field is less than 7×10^{-6} gauss then the current theories of radio emission from the disk of the Galaxy and the retention of cosmic rays appear to need modification. Theories have already been proposed in the literature which appear to explain the polarization of starlight, star formation and the stability of spiral arms with fields of the order of the limit set by the experiment. It is necessary to push the limit further down and sample more clouds seen both in emission and absorption.

Postscript added in proof (*January 1963*)

(1) More recent observations using a sensitive parametric amplifier and a narrower bandwidth of 3 Kc/s have shown the existence of magnetic fields in a number of neutral hydrogen clouds seen both in absorption and in emission. The strongest longitudinal field measured so far is 25×10^{-6} gauss; it was found in one cloud of the narrow Taurus A absorption feature which was discovered to be double. Fields in other components range from 10×10^{-6} gauss to less than 2×10^{-6} gauss. The average field appears to be about 5×10^{-6} gauss and is the same size as previously measured. The new results show the variability of the field in individual clouds; this is thought to be due to the varying states of gravitational contraction of the clouds. The conclusions reached in the paper above remain valid.

(2) New evidence for the existence of large-scale magnetic fields in the bright extragalactic radio sources has recently been found. Observations at Cal. Tech., Jodrell Bank and Sydney have revealed linear polarization amounting to 10 per cent in some cases. The polarization is thought to arise from synchrotron emission from relativistic electrons moving in ordered magnetic fields with a scale comparable with the dimensions of the sources themselves, namely 10 to 100 kpc.

REFERENCES

Bierman, L., and Davis, L., 1960, *Z. Ap.*, **51**, 19.
Bolton, J. G., and Wild, J. P., 1957, *Ap. J.*, **125**, 296.
Davies, R. D., and Hazard, C., 1962, *M.N.R.A.S.*, **124**, 147.
Davis, L., and Greenstein, J. L., 1951, *Ap. J.*, **114**, 206.
Earl, J. A., 1961, *Phys. Rev.*, Letters, **6**, 125.
Elvius, A., and Hall, J. S., 1962, *Sky and Telescope*, **23**, 254.
Henry, J., 1958, *Ap. J.*, **128**, 497.
Hiltner, W. A., 1956, *Vistas in Astronomy*, Vol. 2, Pergamon Press, p. 1080.
Hiltner, W. A., 1958, *Ap. J.*, **128**, 9.
Hoyle, F., 1953, *Ap. J.*, **118**, 513.
Hoyle, F., 1960, *M.N.R.A.S.*, **120**, 338.
Hoyle, F., and Ireland, J. G., 1961, *M.N.R.A.S.*, **122**, 35.
Linsley, J., Scarsi, L., and Rossi, B., 1962, Kyoto Conference, *J. Phys. Soc. Jap.*, **17**, A-II.
Mestel, L., and Spitzer, L., 1956, *M.N.R.A.S.*, **116**, 503.
Meyer, P., and Vogt, R., 1961, *Phys. Rev.*, Letters, **6**, 193.
van Woerden, H., Takakubo, K., and Braes, L., 1962, *B.A.N.*, No. 524.
Woltjer, L., 1961, *Ap. J.*, **133**, 352.

12

THE MECHANISMS OF RADIO EMISSION

M. I. LARGE

Two physical processes are generally thought to operate at the source of the continuum radio emission from the Milky Way. These are the *thermal* process in which radio emission arises from the free–free transitions of thermal electrons in an ionized gas, and the *synchrotron* process, in which high energy electrons emit 'braking radiation' as they are accelerated in the galactic magnetic field. In this lecture I give an outline of the theory of these processes, and discuss some of the applications to galactic radio astronomy. No mention is made of gyro emission, as this has been described by Dr. Wild in his paper on solar radio astronomy.

THE THERMAL RADIO EMISSION FROM AN IONIZED GAS

(i) *Black body radiation*

If a true black body at a kinetic temperature T subtends a solid angle Ω in the sky, then the energy flux density per unit frequency interval reaching the observer is given by

$$S = \frac{2kT}{\lambda^2}\Omega \quad . \quad . \quad . \quad . \quad (1)$$

in the radio frequency band. This relationship follows directly from the Rayleigh-Jeans law. In general, astronomical objects are not black bodies at radio frequencies, but equation (1) is often used in defining a convenient measure of radio brightness, namely the brightness temperature. If the measured flux density of a source subtending a solid angle Ω is S, then the brightness temperature is

$$T_b = \frac{\lambda^2}{2k\Omega}S \quad . \quad . \quad . \quad (2)$$

T_b is therefore not normally the same as the kinetic temperature, and is a function of the radio frequency.

(ii) *Thermal emission from material of finite optical depth*

If a gas is partly transparent, the brightness temperature is less than the electron kinetic temperature by an amount which depends on the optical depth. Consider a plane lamina of gas of thickness ds in which the absorption coefficient per unit length is κ at a particular radio frequency. The electron kinetic temperature is T_e; that is to say that if the lamina were placed in a black enclosure at temperature T_e, it would be in thermal equilibrium. Then the apparent brightness temperature of the lamina would be $T_e\kappa ds$ as follows from the thermodynamic principle of detailed balancing. The quantity κds is called the element of optical depth τ. Suppose now that the lamina of gas lies at a distance s from the observer. Then it is seen by the observer through a total optical depth τ, where

$$\tau = \int_0^s \kappa ds$$

and it is easy to show that the energy flux from the lamina will be reduced by a factor $e^{-\tau}$ in passage through this optical depth. Thus the contribution dT_b which the lamina makes to the brightness temperature seen by the observer is given by

$$dT_b = e^{-\tau}\, T_e d\tau \quad . \quad . \quad . \quad (3)$$

Integrating this we find

$$T_b = \int_0^\infty e^{-\tau} T_e(\tau) d\tau \quad . \quad . \quad . \quad (4)$$

If a cloud of gas of total optical depth τ_1 is seen in front of a background brightness temperature T_0, then (4) becomes

$$T_b = \int_0^{\tau_1} e^{-\tau} T_e(\tau) d\tau + T_0 e^{-\tau_1} \quad . \quad . \quad (5)$$

The significance of this is easily seen if T_e is taken to be constant, for then

$$T_b = T_e(1 - e^{-\tau_1}) + T_0 e^{-\tau_1} \, . \quad . \quad . \quad (6)$$

the brightness temperature is a weighted mean of the electron temperature of the cloud of gas and the background temperature.

(iii) *Calculation of the optical depth*

The value of the absorption coefficient (and hence the optical depth) depends on detailed processes in the gas. The principal

mechanism is the free–free transitions of electrons in the coulomb potential gradients of the positive ions. However in the case of radio emission the calculation can be carried out by the methods of classical electromagnetism, since $h\nu \ll kT$ and $\lambda \gg N_e^{-\frac{1}{3}}$ (i.e. the medium is continuous).

The first step is to set up the equation of motion of the free electrons,

$$m\ddot{x} + mf_{\text{coll}}\dot{x} = \varepsilon_0 E_0 e^{i\omega t} \quad . \quad . \quad . \quad (7)$$

where mx is the momentum and f_{coll} the collision rate. The second term is thus the rate of loss of momentum by collisions. Using equation (7) in the normal way as a 'constitutive relation' in Maxwell's equations we can calculate the absorption coefficient to be (in m.k.s. units)

$$\kappa = \frac{e^2 N f_{\text{coll}}}{\varepsilon_0 mc\omega^2 \left\{1 - \dfrac{e^2 N}{\varepsilon_0 m\omega^2}\right\}^{\frac{1}{2}}} \quad . \quad . \quad (8)$$

putting in numerical values and converting to c.g.s. units this becomes

$$\kappa = \frac{0{\cdot}106\, N f_{\text{coll}}}{\omega^2} \quad . \quad . \quad . \quad (9)$$

N is the electron density and ω is the radio (angular) frequency. The expression

$$\left\{1 - \frac{e^2 N}{\varepsilon_0 m\omega^2}\right\}^{\frac{1}{2}}$$

represents the effective refractive index and is very nearly unity.

The problem now reduces to a determination of f_{coll} the collision frequency. The only important collisions are those of an electron with a heavy nucleus. The calculation of the effective collision rate is a complicated matter, the result depending on the assumed distribution of electron velocities, the degree of screening of the nucleus by the bound electrons, etc. Various workers have considered the problem, and all are in substantial agreement. Shklovsky (1956) for example referring to the results of detailed calculations by Ginzburg uses the result

$$f_{\text{coll}} = 3{\cdot}65 \frac{N}{T_e^{3/2}} \left(19{\cdot}8 + \ln\left(\frac{T^{3/2}}{f}\right)\right) \quad . \quad . \quad (10)$$

It may seem strange that the expression for the collision rate

contains a frequency dependent term. It arises from the fact that any one electron-ion interaction may be 'fast' enough to behave as a collision in the sense that it produces low radio frequencies, but too 'slow' to produce the higher frequencies. Thus the collision rate appears to fall slowly with increasing radio frequency.

This collision rate yields as the absorption coefficient

$$\kappa = 9{\cdot}8 \times 10^{-3} \frac{N_i N_e}{T^{3/2} f^2} \left(19{\cdot}8 + \ln \frac{T^{3/2}}{f}\right) \quad . \quad (11)$$

which to a reasonable approximation is

$$\kappa = \frac{N_i N_e}{T^{3/2} f^{2{\cdot}1}} \quad . \quad . \quad . \quad . \quad (12)$$

the exponent 2·1 including the slow logarithmic variation.

In convenient units, then, the optical depth τ may be written

$$\tau = 3{\cdot}1 \times 10^{-4}\, T_e^{-3/2}\, \lambda^{2{\cdot}1} E$$

where E is the emission measure, $E = \int N_i N_e ds$, measured in cm^{-6} parsec, and λ is the wavelength in cms.

(iv) *The spectrum of thermal radio emission*

We are now in a position to comment on the spectrum of thermal radio emission. We have seen that the contribution to the radio brightness of a thermal source is

$$T_b = T_e(1 - e^{-\tau})$$

For optically thin regions, $\tau \ll 1$ and we have

$$T_b \approx \tau\, T_e$$

so that

$$T_b \propto \lambda^{2{\cdot}1}$$

whereas for optically thick regions $T_b = T_e$.

These results are summarized in Fig. 12/1 which is based on a diagram from Pawsey and Bracewell's book (1955).

If a thermal radio source is seen against an appreciable background temperature, the situation is more complicated. For taking equation (6) we have

$$T_b = T_e(1 - e^{-\tau_1}) + T_0 e^{-\tau_1} \quad . \quad . \quad (6)$$

The brightness temperature of the source would normally be obtained by comparison with a nearby region, when the

background temperature T_0 is observed. Thus the apparent temperature of the source above its immediate surroundings is

$$T = (T_e - T_0)(1 - e^{-\tau}) \quad . \quad . \quad . \quad (13)$$

showing that the thermal source can appear in absorption if seen

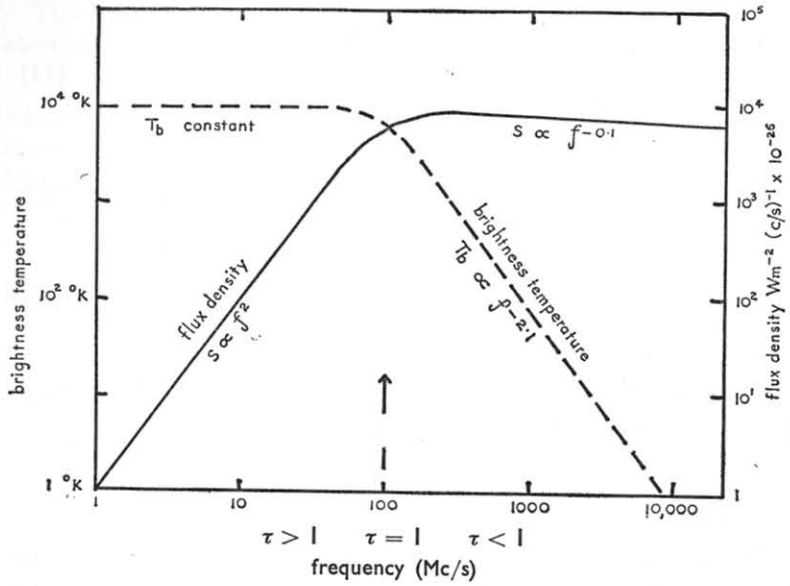


FIG. 12/1.—The radio brightness temperature and flux density of a typical thermal radio source are plotted as a function of frequency on a logarithmic scale, to illustrate the range of continuum spectra that may occur in thermal radio emission.

against a high background temperature. A very striking example of this is seen in the Australian 19·7 Mc/s survey (Shain, Komesaroff and Higgins, 1962) in which the galactic ionized hydrogen appears as a marked absorption against a very high non-thermal background.

(v) *Summary of properties of thermal emission*

(*a*) The spectral index is given by $\alpha \leqslant 2{\cdot}1$ in the expression $T_b \propto \lambda^{\alpha}$.

(*b*) The emission is isotropic.

(*c*) Thermal regions appear in absorption if the background brightness temperature is greater than the electron temperature.

(*d*) The amount of thermal radio emission is calculable from the kinetic temperature and emission measures where these are optically determinable.

It is clear from observations of the Galaxy and extragalactic sources that the spectral index is steeper than 2·1, and that therefore the emission from these sources is not entirely thermal. In this conclusion, radio astronomers were anticipated by Lucretius (*c.* 55 B.C.). He wrote, referring no doubt to the properties of thermal emission that we have discussed: 'For all these reasons, those who have imagined that the raw material of things is fire and that the universe consists of fire alone have evidently wandered far from the truth.' And, as if that were not clear enough, he continues to explain that no combination of thermally emitting regions can give rise to the observed non-thermal spectrum of the Galaxy: 'no possibility that the immense variety of these things could result from variations in the density or rarity of the fire. Its heat would simply be fiercer as its parts were more concentrated, milder as they were dispersed and dissipated.'

SYNCHROTRON EMISSION

The distinguishing feature of thermal emission is that the electrons in the gas are taken to have a Maxwell distribution of energies, that is to say the gas may be said to have a kinetic temperature. In the synchrotron emission process the bulk of the emission arises from the acceleration of a few very high energy electrons in interstellar magnetic fields. An electron moving at small speeds is constrained by a magnetic field to move in circles about the magnetic field at the Larmor frequency

$$\omega_H = \frac{eH}{mc} \quad \left(\frac{e}{m} B \text{ in m.k.s. units}\right)$$

The electron may have a uniform component of velocity along the magnetic field, so that the motion is spiral. Since the Larmor frequency is 2·8 Mc/s per gauss, in the Galaxy slow electrons would radiate a dipole field at a few cycles per second.

The picture is very different if the electron is moving at relativistic speeds. For then the radiated energy is strongly concentrated by the doppler effect into a cone of vertex angle

$\theta \approx 2\dfrac{m_0c^2}{E}$ about the direction of instantaneous velocity. Here m_0 is the rest mass of the electron and E its energy.

Since a typical value of E for a cosmic ray electron may be 10^9 eV and $m_0c^2 = 0{\cdot}5 \times 10^6$ eV, the vertex angle of the cone of radiation is measured in minutes of arc, and an observer sees a short pulse of radiation as the spiralling electron momentarily moves tangentially to the line of sight. The pulse lasts for a time Δt given by

$$\Delta t \approx \tfrac{1}{2}\left(\frac{\theta}{\omega_H}\right)\left(\frac{m_0c^2}{E}\right)^2$$

the term $\left(\dfrac{\theta}{\omega_H}\right)$ being the time for which the instantaneous velocity is within θ of the line-of-sight, and the term $\left(\dfrac{m_0c^2}{E}\right)^2$ appearing because the energy is concentrated into a short pulse. Using the relativistic expression for the Larmor frequency we find that Δt can be written

$$\Delta t = \left(\frac{mc}{eH}\right)\left(\frac{m_0c^2}{E}\right)^2 \quad . \quad . \quad . \quad (14)$$

Thus the observer sees a sequence of pulses, with repetition frequency ω_H and duration Δt. Since ω_H is very small, the spectrum is virtually continuous. A full theory requires a Fourier analysis of the train of pulses. The spectrum has the form

$$Pdf \propto H\, p(\omega/\omega_m)\, df \quad . \quad . \quad . \quad (15)$$

where P is the power radiated per unit bandwidth and $p(\omega/\omega_m)$ is a complicated function with a maximum at $\omega = \omega_m = \dfrac{1}{\Delta t}$.

In order to proceed further we need to know something of the distribution of the relativistic electron energies. If they are taken to follow a power law of the form

$$N(E)dE = kE^{-\gamma}dE$$

where $N(E)dE$ is the number of particles in the energy range E to $E + dE$ and γ is a constant, then the integration may be performed to determine the expected radio spectrum of the synchrotron process.

It may be shown that the radio emissivity ε is given by an expression containing the terms

$$\varepsilon(f) \propto H^{(\gamma+1)/2}\,(\sin\theta)^{(\gamma+1)/2} f^{-(\gamma-1)/2} \quad . \quad . \quad (16)$$

with other terms that depend on the assumed cut-off energies of the electron energy distribution. Here θ is the angle which the electron velocity makes with the magnetic field.

We see that this emissivity has a spectrum $f^{-(\gamma-1)/2}$ which readily leads to spectra in which the observed flux density falls with frequency in marked contrast to the thermal emission process. The observed radio non-thermal brightness temperatures are thus directly related to the galactic magnetic field and the spectrum of high energy electrons. It remains to be seen whether the limits placed on the magnetic field by radio Zeeman effect measurements and observations of cosmic ray electrons will substantiate the theory that the galactic continuum radio emission is synchrotron radiation.

INTERPRETATION OF THE RADIO CONTINUUM EMISSION OF THE GALAXY

(i) *Separation of thermal and non-thermal components of galactic emission*

We have already seen how thermal sources may be recognized if they appear in absorption or with a characteristic emission spectrum. In the more general case of the Galaxy, where there are both thermal and non-thermal sources, it is possible in principle to deduce the fraction of the emission arising in thermal and non-thermal processes by an examination of the radio spectrum. For example it has been shown (Westerhout, 1958) that if thermal and non-thermal sources are well mixed in the line-of-sight the apparent brightness temperature will be given by

$$T = (T_e + T_n/\tau)(1 - e^{-\tau}) \quad . \quad . \quad . \quad (17)$$

By assuming $\tau \propto f^{-2{\cdot}1}$ and T_n (the non-thermal component) $\propto f^{-\alpha}$ it is possible to use a number of surveys at different frequencies to deduce the distribution of thermal and non-thermal emission, and also the variation of α. Fig. 12/2 shows the result of an attempt to do this (Large, Mathewson and Haslam, 1961) using surveys at 1390, 408 and 85 Mc/s. A much

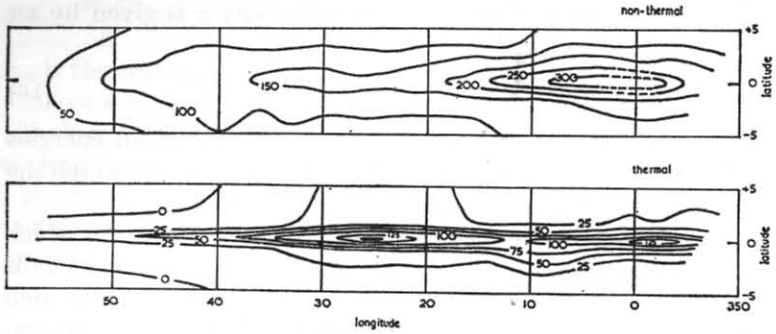


FIG. 12/2.—The radio emission from part of the Galaxy at 408 Mc/s resolved into thermal and non-thermal components on the basis of continuum spectrum measurements. The isophotes are marked in °K of brightness temperature.

more detailed analysis should be possible now that high resolution surveys ranging from 19·7 Mc/s to 3000 Mc/s are available.

(ii) *Polarization*

Thermal radiation is unpolarized, but the synchrotron process leads to emission partially polarized with the electric vector parallel to the acceleration (Westfold, 1959). Many factors arise which will subsequently serve to reduce the polarization, of which the most serious is differential Faraday rotation across the bandwidth in the interstellar material. Until recently attempts to detect polarization have failed for technical reasons, but in the last year or so positive measurements have been made at several observatories.

(iii) *The spiral structure of the Galaxy*

It is well known, from the 21 cm line observations, that the neutral hydrogen in the Galaxy is distributed in an approximately spiral structure. This being so, one might expect to find evidence of this structure in the radio continuum, particularly in the distribution in longitude. Mills (1959) was the first to point out that the 85 Mc/s radiation from the Galaxy showed marked steps in the longitude distribution, and he was able to fit a spiral structure which explained the steps; the argument being that you would expect to see enhanced emission where the line-of-sight was tangential to a spiral arm. This would indeed

be true of thermal emission, which is radiated isotropically, but at 85 Mc/s the radiation is largely non-thermal, and the structure to be observed depends on the polar diagram of the synchrotron emission and needs careful examination. The polar diagram is shown in Fig. 12/3. It is a polar plot of the function $(\sin\theta)^n$ for $n = 1$, 1·5 and 2, corresponding to values of $(\gamma + 1)/2$ in equation (16), and radio spectral indices of -2, $-2{\cdot}5$ and -3 respectively. Hanbury Brown and Hazard (1961) calculated the longitude distribution you might expect for a spiral model of the Galaxy and concluded that 'steps' would not be visible unless the magnetic field deviated somewhat from the exact line of the spiral arm, thus introducing a small component of H perpendicular to the line-of-sight. Furthermore examination of the distribution of thermal emission does not apparently reveal the pronounced 'steps' expected at the tangential points of spiral arms for an isotropic emission process.

I think these conclusions should be regarded as tentative, an example of the kind of investigation that is possible.

(iv) *Non-isotropic distribution of electron velocities*

In drawing the radiation polar diagram for synchrotron emission (Fig. 12/3) I implicitly assumed that the relativistic electrons were isotropically distributed. If they are not the combination of the anisotropy with the directional properties of synchrotron emission can give rise to 'rainbows' of enhanced radio brightness. An interesting example of the kind of effect that might occur is provided by recent work of Sciama (1962). He has shown that if the galactic magnetic field varies only slowly in space and time compared with the Larmor radius and Larmor frequency, then the electron will move in such a way that the magnetic moment of the spiralling motion is conserved. From this it follows that the electron trajectories are restricted to values of θ less than $\sin^{-1}(H/H_0)^{\frac{1}{2}}$ where θ is the angle the electron velocity makes with the magnetic field, H, and H_0 is the (greater) magnetic field in which the electron velocities were isotropic.

Since an individual relativistic electron emits synchrotron radiation only in the direction of its instantaneous velocity, the result of restricting the pitch angles of the spiral motion will be to confine the synchrotron emission to cones of vertex angle 2θ. For example suppose the observer to lie in a region of uniform

magnetic field $H = \frac{1}{2}H_0$. Then θ will be 45°, and the observer will see the synchrotron emission only at angles less than 45° to the magnetic field. Divergence and curvature of the magnetic

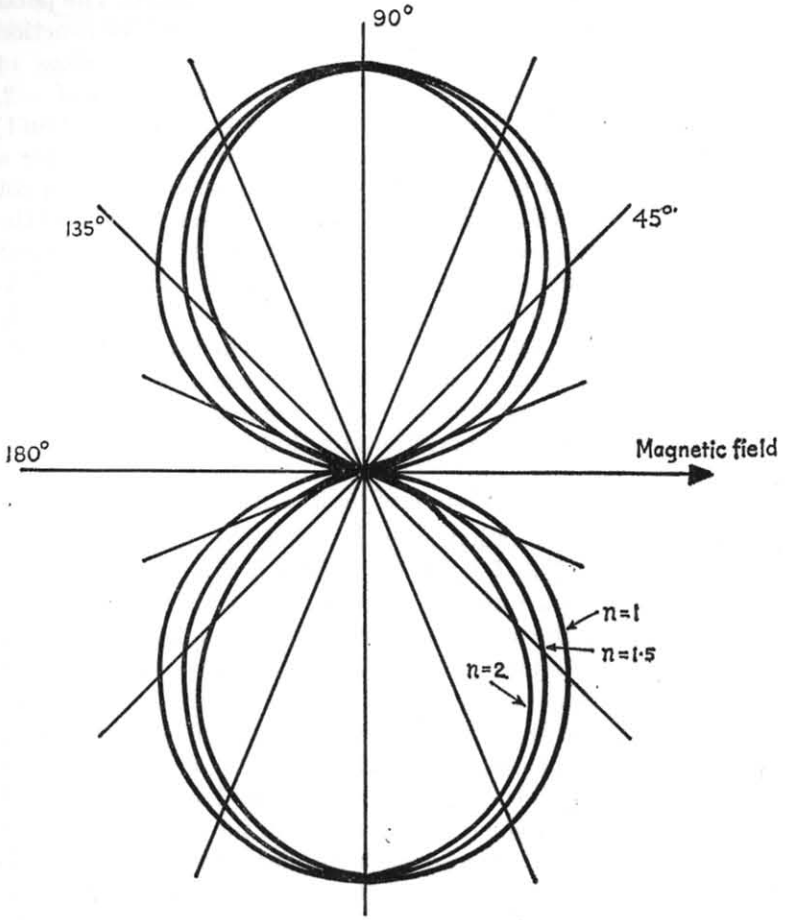


FIG. 12/3.—A polar plot of the function $r = (\sin\theta)^n$ showing the directional properties of synchrotron emission. The diagram has been plotted for values of n of 1, 1·5 and 2 corresponding to radio spectral indices of -2, $-2{\cdot}5$ and $-3{\cdot}0$.

field would modify this simple picture, and it has been suggested (Large, Quigley and Haslam, 1962) that a mechanism of this kind might provide an explanation of the large-scale features observed in the radio continuum at high galactic latitudes.

REFERENCES

Hanbury Brown, R., and Hazard, C., 1960, *The Observatory*, **80**, 137.
Haslam, C. G. T., Davies, J. G., and Large, M. I., 1962, *M.N.R.A.S.*, **124**, 169.
Hill, E. R., Slee, O. B., and Mills, B. Y., 1958, *Aust. J. Phys.*, **11**, 530.
Large, M. I., Mathewson, D. S., and Haslam, C. G. T., 1961, *M.N.R.A.S.*, **123**, 113.
Large, M. I., Quigley, M. J. S., and Haslam, C. G. T., 1962, *M.N.R.A.S.*, **124**, 405.
Lucretius, c. 55 B.C., *The Nature of the Universe*, trans. R. E. Latham, Penguin Books, London, 1961.
Mills, B. Y., 1959, *Paris Symposium on Radio Astronomy*, Stanford University Press.
Pawsey, J. L., and Bracewell, R. N., 1955, *Radio Astronomy*, Oxford University Press.
Sciama, D. W., 1962, *M.N.R.A.S.*, **123**, 317.
Shain, C. A., Komesaroff, M. M., and Higgins, C. S., 1962, *Aust. J. Phys.*, **14**, 508.
Shklovsky, I. S., 1956, *Cosmic Radio Waves*, trans. Harvard University Press, 1960.
Westerhout, G., 1958, *B.A.N.*, **14**, 215.
Westfold, K. C., 1959, *Ap. J.*, **130**, 241.

13

THE SPECTRA OF RADIO SOURCES

R. G. CONWAY

No spectral lines appear in the spectrum of the discrete radio sources, so that the methods of classification and of distance measurement, which are so important in spectroscopy, are not available to the radio astronomer. The emission occurs as a continuum, over a wide frequency range, and the flux density, S, may be compared with a power law of the form $S \propto f^x$. By the spectrum of a radio source we shall mean a plot of log S versus log f. It is known that thermally emitting sources (including the Sun, and the Moon) exhibit a spectral index x which is positive, or near zero. By contrast, the majority of discrete sources are non-thermal, and have a negative spectral index.

The determination of the spectrum of a source requires the measurement of its flux density in absolute terms at several frequencies. Such absolute measurements consist of comparing the power received (a) from the source with an aerial of known gain, and (b) from a heated resistor. They have only been done at a few frequencies, namely 38 Mc/s, 400 Mc/s, 1400 Mc/s and 3200 Mc/s. Intensity measurements at other frequencies, which consist of the relative intensities of sources at those frequencies, are usually calibrated from one source (e.g. Cassiopeia A) by assuming the form of its spectrum (Whitfield, 1957).

The results presented in this paper consist of measurements at nine frequencies, made at three observatories, Cambridge, Jodrell Bank, and the observatory of the California Institute of Technology. They will be published shortly in the Monthly Notices of the R.A.S. (Conway, Kellerman and Long, in preparation). A total of 161 sources were measured, 28 of them over the full frequency range, and a further 46 with virtually complete coverage between 38 and 1420 Mc/s. Most of the remaining 87 sources were observed at four or more frequencies.

Both interferometers and pencil-beam aerial systems were used in this work. A careful examination was made of the

TABLE 13/I

Observing systems

Frequency (Mc/s)	Observatory	System	Beamwidth (degrees)	Interferometer spacing
38	Cambridge	Aperture Synthesis	0·8 × 1·5	—
178	Cambridge	Pencil Beam and Interferometer	0·4 × 4	470λ
240	Jodrell	Pencil Beam	1·1	—
408	Cambridge	4-Aerial Interferometer	0·5 × 3	70 & 787λ
412	Jodrell	Pencil Beam	0·8	—
710	Cal. Tech.	Interferometer	1·1	150λ
960	Cal. Tech.	Pencil Beam and Interferometer	0·8	200λ
1420	Cal. Tech.	Interferometer	0·5	150λ
3200	Cal. Tech.	Pencil Beam	0·25	—

sources of error, to ensure that the data were homogeneous. Errors for most sources were estimated at 10–20 per cent, being due principally to the effects of confusion with nearby sources, and of resolution of the source by the aerial pattern. A valuable check was obtained by comparing the 408 Mc/s observations, made with an interferometer at Cambridge, with the 412 Mc/s observations, made with a pencil-beam system at Jodrell Bank. Very satisfactory agreement was obtained.

As explained above, absolute flux density measurements have only been made at 38 Mc/s (Adgie and Smith, 1956; Long, unpublished), 400 Mc/s (Seeger, 1956), 1400 Mc/s (Heeschen, 1960) and 3200 Mc/s (Broten and Medd, 1961). We have calibrated the remaining frequencies by assuming that the mean of seven calibration sources had a spectrum of constant spectral index. These calibration sources were 3C 123, 161, 196, 348, 353, 380 and 409. We have not followed precedent in using Cassiopeia A since:

(i) it has a comparatively large angular diameter
(ii) its intensity is much greater than that of the majority of sources whose spectra we wish to determine.

The spectra of the 161 sources were inspected in graphical form, and it was clear that while many were consistent with a straight line, corresponding to a simple power law:

$$\log S = x \log f + \text{constant}$$

other examples showed significant deviations from such a law. The spectra were divided into the following classes:

Class T (5 sources)	Thermal sources (H II regions) which will not be considered further.
Class S (112 sources)	The spectrum is straight to within the experimental error over the observed frequency range.
Class S_1 (14 sources)	The spectrum is straight from 38 to 3200 Mc/s.
Class S_2 (12 sources)	The spectral index is constant at low frequencies but becomes more negative above 1400 Mc/s.
Class C (18 sources)	The spectrum is curved, i.e. the spectral index is a function of frequency.

This method inevitably contains a systematic bias in that the weaker sources, which have larger errors, and where the data are not as complete, tend to fall in Class S. In particular, sources in Class S have not been measured at 3200 Mc/s; if they had been, they would be placed either in S_1 or S_2. With this in mind, we estimate that between 30 per cent and 50 per cent of all sources should have spectra that are straight over the frequency range 38—3200 Mc/s, the remainder showing deviations from a straight spectrum. The deviations which have been measured are invariably in such a sense as to make the spectrum convex, i.e. the spectral index becomes more negative with increasing frequency.

The data were also analysed using the Mercury Computer of Manchester University. The best fitting straight line of the form

$$\log S = x \log f + \text{constant}$$

was computed, using a least squares fit with each flux density measurement weighted inversely as the square of its uncertainty. The spectral index 'x' and the computed flux density at 400 Mc/s, designated S_{400}, were computed for each source.

To improve the homogeneity of the data, the relatively few measurements at 3200 Mc/s were not included in this analysis. Instead, each measured value was compared with the extrapolation from lower frequencies in the form of the radio ρ, where

$$\rho = \frac{\text{Extrapolated flux density at 3200 Mc/s}}{\text{Measured flux density at 3200 Mc/s}}$$

It was found that the values of ρ clustered around the two values 1·0 and 1·25, with a deficit of intermediate values. This result supports the separation of classes S_1 and S_2 above.

For those sources where the data were sufficient, a further calculation was made of the variation of spectral index with frequency. For this purpose two lines were fitted to the measurements, one using 38, 178, 240, 408 and 412 Mc/s, and the other using 408, 412, 710, 960 and 1420 Mc/s. The two spectral indices so obtained correspond approximately to the slopes of the tangents to a curved spectrum at 100 Mc/s and 1000 Mc/s, and have been designated x_{100} and x_{1000}.

The difference $\Delta x = (x_{100} - x_{1000})$ provides a quantitative measure of the curvature of a spectrum, and is free of the statistical bias affecting the classification of spectra mentioned above. In general the two methods agree satisfactorily; Class C sources have values of Δx which are always positive, mostly between +0·20 and +0·60, while the remainder have values of Δx which are not systematically positive, usually lying in the range $\pm$0·20.

A few of the sources measured have been identified with galactic or extragalactic objects. Four supernova remnants are all classed as S_1 (one of these, IC 443, has not been measured by us at high frequencies, but other measurements (Heeschen, 1961) are consistent with this classification). Three sources identified with stars (3C 48, 196 and 286, see below, G. R. Burbidge, Chapter 16) are of spectral Class C, S_2, C respectively.

Out of 35 galaxies, identified by Minkowski (1961), 5 are Class S_1, 7 Class S_2, 6 Class C and 17 Class S. These proportions are what one would expect in a random sample from the whole list. It is perhaps surprising that the largest known radio emitters (Cygnus A, 3C 295) and the smallest, 3C 48, all show similar spectra, namely Class C. This is evidence that the same emission mechanism is involved in each case (almost certainly it is the synchrotron mechanism), but at the same time it shows that a classification of radio sources by their spectral features will have only limited physical meaning.

The presence of spectral features should in principle enable the red shift, and hence the distance, of an extra-galactic source to be measured. The spectra of sources, whose distance is known, however, show such a wide variety of form, that we cannot, at

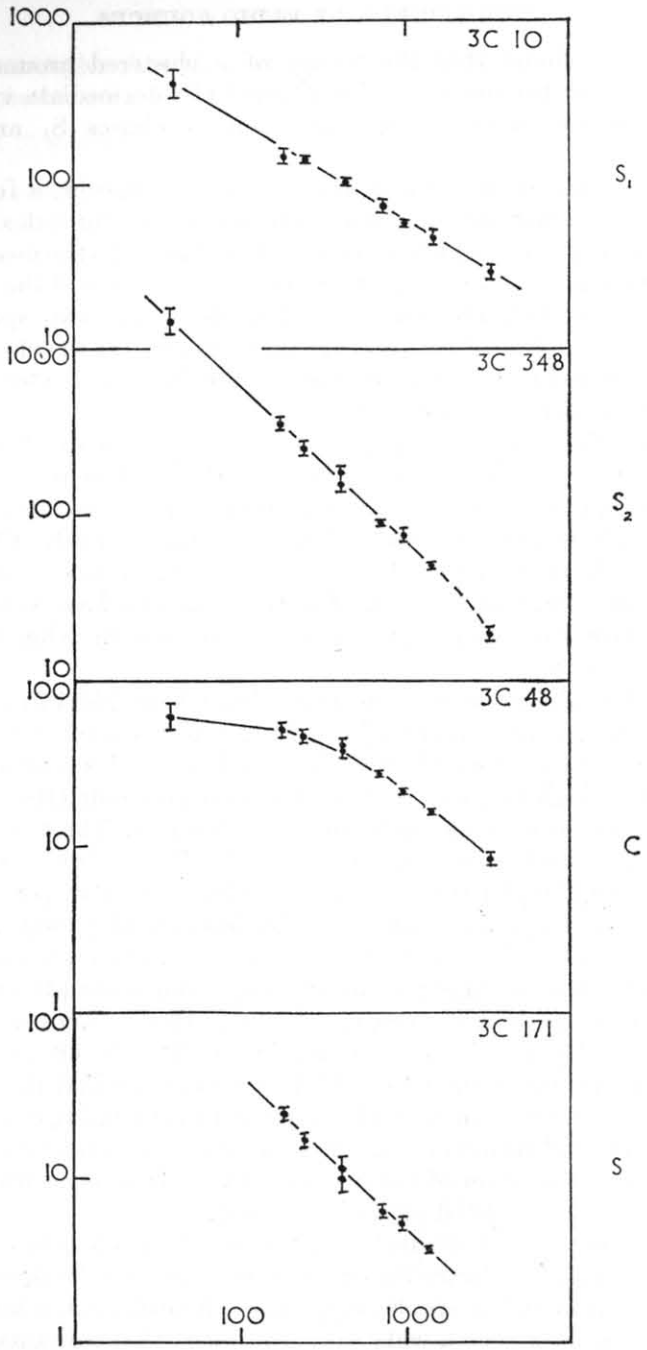


FIG. 13/1.—The spectrum of four non-thermal sources, illustrating the four spectral types.

least at present, determine the red shift of an individual radio source from a knowledge of its spectrum.

A histogram of the mean spectral index of all sources shows a marked concentration about the value −0·70, over 60 per cent of the sources having values between −0·60 and −0·80. We found none with spectral index less than −1·15, a result which cannot wholly be accounted for by instrumental selection. The histogram of sources at low galactic latitude ($b < 10°$) is not so peaked as the general histogram, indicating that galactic

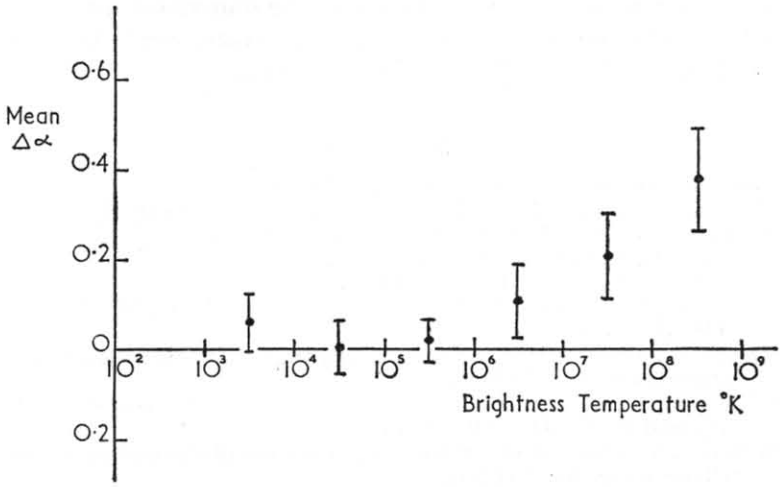


FIG. 13/2.—The average spectral curvature of sources in different ranges of brightness temperature. The vertical lines indicate the r.m.s. scatter of the index ($x_{100} - x_{1000}$) about the average value. Each point in the figure represents several sources.

sources may have a wider spread of spectral index than extra galactic sources.

We have attempted to correlate the spectral index against the distance and against the absolute luminosity of identified sources. In neither case could we establish correlation. We have also tested spectral features against the brightness temperature of the sources. For this purpose we have used the flux density computed at 400 Mc/s, and the angular diameter measure by Moffett (1961) and Maltby (1961) and by Allen, Palmer and Hanbury Brown (1962). There appears to be good correlation between brightness temperature and spectral curvature, as measured by the index ($x_{100} - x_{1000}$).

In qualitative terms, this correlation means that sources of high brightness temperature tend to be of spectral Class C. It is generally accepted that non-thermal sources emit by the synchrotron mechanism. A curved spectrum would then imply that a source is deficient in high energy electrons ($E > 3 \cdot 10^9$ eV) while still comparatively rich in electrons of lower energy. One may tentatively construct a model in which sources of Class C are at an early stage of evolution, having very quickly lost the highest energy electrons, but not being yet old enough for the remainder to have decayed. In such young sources the magnetic field, and therefore the brightness temperature, would be high (Kellermann, Long, Allen and Moran, 1962).

REFERENCES

Adgie, R., and Smith, F. G., *Obs.*, **76**, 181 (1956).
Allen, L. R., Palmer, H. P., and Brown, R. Hanbury, 1962 (in the press).
Broten, N. W., and Medd, W. J., *Ap. J.*, **132**, 279 (1960).
Heeschen, D. S., *P.A.S.P.*, **72**, 368 (1960).
Heeschen, D. S., *Ap. J.*, **133**, 322 (1961).
Kellerman, K. I., Long, R. J., Allen, L. R., and Moran, M., *Nature*, **195**, 692 (1962).
Maltby, P., Observations of California Institute of Technology Radio Observatory, No. 5 (1961).
Minkowski, R., Proc. IV Berkeley Symposium on Mathematical Statistics and Probability, 245 (1961).
Moffet, A. T., Observations of California Institute of Technology Radio Observatory, No. 4 (1961).
Seeger, C. L., *B.A.N.*, **13**, 100 (1956).
Whitfield, G. R., *M.N.R.A.S.*, **117**, 680 (1957).

14

THE ANGULAR SIZES OF RADIO SOURCES

H. P. PALMER

DISCRETE radio sources have frequently been called 'point' sources, because they have angular sizes far smaller than the angular resolving power of the radio telescopes with which they are observed. With the commissioning of ever larger telescopes, with fully filled, partially filled or synthesized apertures, giving resolving powers of a few minutes of arc, this description is used less, but it remains true that the majority of sources appear as 'points' to the instruments used to detect them in the first instance, and to measure their celestial positions and flux density. Measurements of their angular size require much higher resolving powers, and can at present be made only by interferometric techniques, with an appropriate aerial separation or baseline. An interferometer of baseline L metres, receiving radiation of wavelength λ metres, can be used to study angular structure rather smaller than its resolving power, λ/L radians, or 206,264 λ/L secs.

OBSERVING TECHNIQUES

For some of the earliest work on Cygnus A and Cassiopeia A Hanbury Brown, Jennison and Das Gupta (1952) used an instrument working on an entirely new principle, the correlation interferometer, which did not preserve the RF phase of the received signals. This instrument suffers from particularly stringent signal/noise ratio limitations when it is used to observe weaker sources, and all later observations have been made with interferometers in which the receivers at the two aerials remain phase coherent at least for the duration of the observations. These instruments are therefore radio equivalents to the Michelson Stellar Interferometer (1920).

For aerial separations < 1 km, phase coherence may be achieved by sending the RF signals down low loss cables to the

central interferometer equipment. Alternatively one may transmit a CW signal to the receiver at each aerial, use it there as a 'local' oscillation, and transfer the noise bands to the interferometer at the intermediate frequency. With such systems the phase information may be preserved from day to day, so that the signal/noise ratio may be improved by integrating successive records, and the accuracy of position measurements becomes comparable with the resolving power of the interferometer, rather than the resolving power of the individual aerials.

For aerial separations greater than about 1 km radio link systems are to be preferred, and phase coherence is achieved by the transmission of a signal at or near the frequency of the local oscillator. These systems remain phase stable for the duration of an observation (of order 30 minutes) but at the longest baselines no attempt has yet been made to preserve phase coherence from day to day.

Whether the interferometer multiplies the bands of noise power received at the two aerials in a square law detector or in a linear multiplier, the term which contains the angular information from the interferometer has the form, if the output deflection D is proportional to received power,

$$D \propto S\sqrt{A_1 A_2}\cos\phi$$

where S is the total flux density from the source falling per square metre of aerial, A_1 and A_2 are the areas of the two aerials, and ϕ is the phase difference between the two signals when they reach the interferometer (see for instance Pawsey and Bracewell, 1955). This phase difference may be considered as the sum of two terms:

$$\phi = \frac{2\pi}{\lambda}(\tau + L\cos\beta)$$

where β is the angle between the baseline and the direction of the radio source. The constant τ arises from any asymmetry in the instrument, which makes the 'optical paths' from the source to the interferometer unequal in the two arms. The term $L\cos\beta$ varies as the source moves across the sky, and the resulting change of phase produces the output fringe pattern at the frequency

$$f = \frac{1}{2\pi}\frac{d\phi}{dt} = \frac{L}{\lambda}\sin\beta\frac{d\beta}{dt}$$

f depends of course on the celestial coordinates of the source, and on the direction of the baseline, so that the expression for f becomes more complicated when it is expressed in variables such as Right Ascension and Declination (see Chapter 15). Only a limited range of values of f can be used, for if there is less than one fringe in the time of observation, the fringe amplitude may be indeterminate, while if the frequency is too high, the output time constant of the recording system attenuates the fringe amplitude. As the signal/noise performance of the instrument will be impaired if this time constant is made shorter, it is necessary to arrange that the output fringe frequency shall be within the band about 0·1 to 0·5 fringe/minute. This can be achieved, of course, by choice of a baseline of length about 300 wavelengths, but for longer baselines, when the natural fringe frequency f becomes inconveniently high, the output frequency may be reduced by introducing into the apparatus a continuous phase change, at an appropriate rate f_c and in a sense such that the output frequency $f_0 = |f - f_c|$.

If the received signals were 'monochromatic' continuous oscillations the constant term τ would be unimportant, but as we are concerned with random noise signals of finite bandwidth B Mc/s, this term must be brought to zero, by introducing an equal delay into the opposite arm of the instrument. The precision with which the delay must be equalized in the two arms depends on the signal frequency bandwidth of the instrument. For a rectangular bandpass, an inequality of 1/2B will reduce the fringe amplitude by 36 per cent.

THE REDUCTION OF THE OBSERVATIONS

Transit observations will enable one to measure the amplitude (and possibly the phase) of the fringe pattern produced by a source at a given aerial separation. In principle, the information about the angular dimensions of the source may be obtained by increasing the separation of the aerials until the amplitude of the fringe pattern decreases. One may assume that the performance of the instrument remains constant as the equipment is moved about, or, preferably, one may calibrate the observations at each baseline by observing a source which has previously been shown to be much smaller than the resolving power of that baseline. For the longest baseline such a procedure

cannot be used and then the problems of calibration are much more difficult. They may be approached by the use of steerable aerials to make observations away from transit, so that the effective length of the baseline is reduced by foreshortening, and the observations thus compared with previous ones at shorter baselines (Allen *et al.*, 1962).

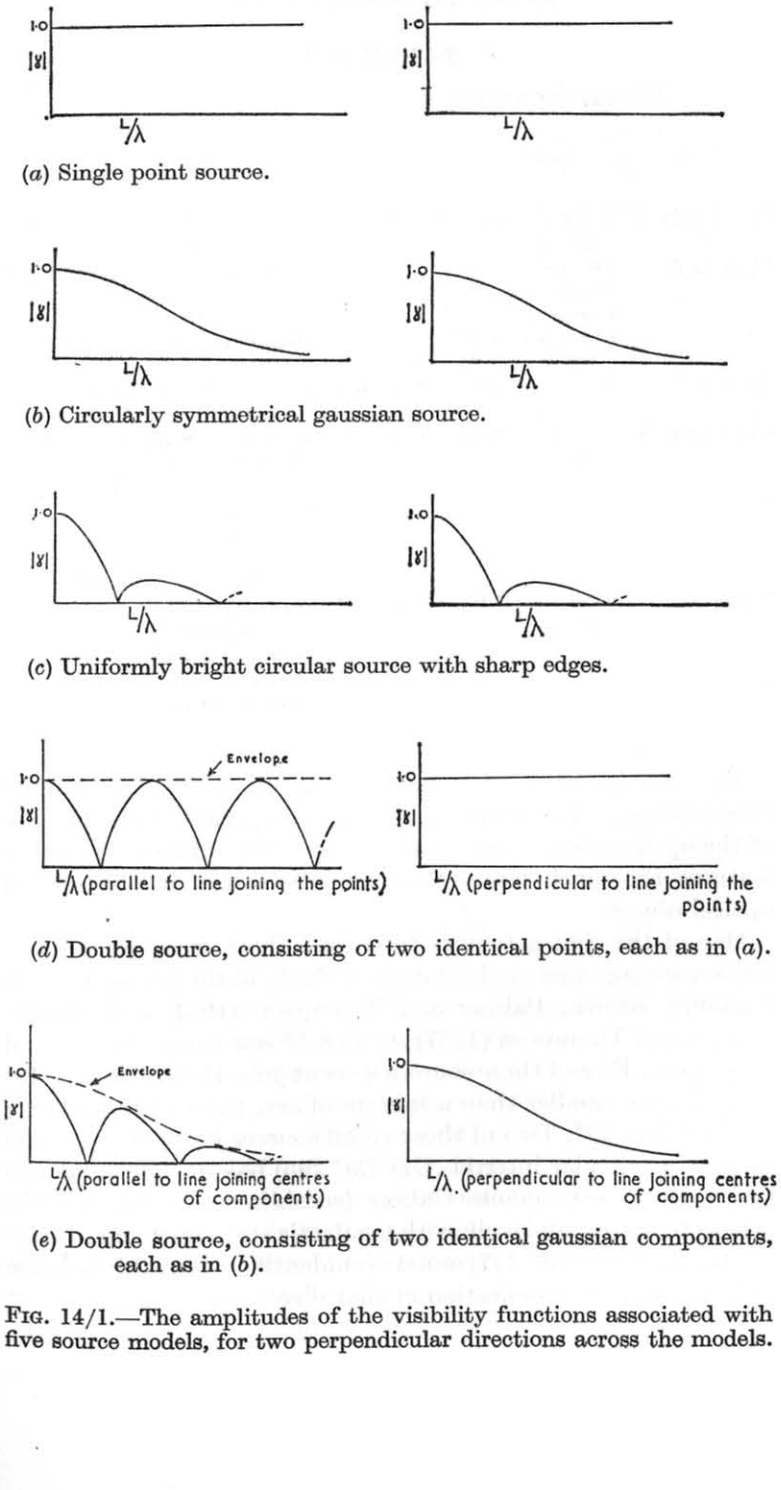
The angular information is revealed more clearly if one compares the values of D at different spacings. More accurately, one calculates values of the fringe visibility

$$\gamma = \frac{\text{fringe amplitude at baseline } L}{\text{fringe amplitude at a very short baseline}}$$

One can also define the phase of the fringe pattern for each value of γ and a plot of the resulting fringe visibility against baseline is referred to as the visibility function for that direction across the source, and it is in fact the Fourier transform of the brightness distribution across the source projected in that direction. If the visibility function is completely specified in amplitude and phase for a number of directions across the source, the two-dimensional brightness distribution may be calculated uniquely. In general the phase of the fringe pattern is more difficult to measure than its amplitude, and is sometimes ignored. The diagrams of Fig. 14/1 show only the theoretical amplitudes of the visibility functions for two baselines at right angles, for sources having different and increasingly complicated distributions of surface brightness. For only 40 or 50 sources are the observational points sufficiently well defined and numerous to justify the fitting of a visibility function, and for most sources there are available at the present time only the values of fringe amplitude for two or three baselines. In those cases the information about the detailed structure of the sources is not yet known, but the data can still give some statistical guidance about the angular scale of radio sources, and the design of future experiments.

EARLY RESULTS

The more intense sources were studied first, and the situation as it was known in 1954–5 is summarized in Table 14/I, together with notes on more recent work.



(*a*) Single point source.

(*b*) Circularly symmetrical gaussian source.

(*c*) Uniformly bright circular source with sharp edges.

(*d*) Double source, consisting of two identical points, each as in (*a*).

(*e*) Double source, consisting of two identical gaussian components, each as in (*b*).

FIG. 14/1.—The amplitudes of the visibility functions associated with five source models, for two perpendicular directions across the models.

TABLE 14/I

The angular dimensions of the most intense sources

Name	Data known in 1954–5	Recent work
Cassiopeia A	3′ by 4′ (Smith, 1952) (214 Mc/s)	3′ × 3·8′ (Moffet & Maltby, 1962) (960 Mc/s)
Cygnus A	Double source, Separation 88″ Components 30″ × 51″ (Jennison and Das Gupta, 1953) (125 Mc/s)	Small nucleus with 10 per cent of flux between main components (Lequeux, 1962) (1420 Mc/s)
Taurus A	3½′ × 5½′ (Mills, 1953) (101 Mc/s)	3·3′ × 3·7′ (Moffet and Maltby, 1962) (960 Mc/s)
Centaurus A	3′ × 6½′ (Mills, 1953) (101 Mc/s)	Halo 3° by 8° which probably has double structure (Twiss *et al.*, 1960) (1427 Mc/s) Core double Components 4′ × 2′ Separation 7′ (Maltby, 1961) (960 Mc/s)
Virgo A	2½′ × 5′ (Mills, 1953) (101 Mc/s)	Halo and double core Components 23″ Separation 33″ (Lequeux and Heidman, 1961) (1427 Mc/s)

The angular sizes of Cassiopeia A, Taurus A, the core of Centaurus and the core of Virgo A are comparable with the sizes of the optical objects with which they are identified. The other features are much larger than any dimension in the related optical object.

One of the first attempts to extend these measurements to weaker sources was made at Jodrell Bank, in the period 1953–6. Hanbury Brown, Palmer and Thompson (1954) and Morris, Palmer and Thompson (1957) studied 15 sources in the zenithal strip there. Five of these sources were at galactic latitudes $> 10°$ and all were smaller than a minute of arc, three of them being smaller than 12″. Two of these small sources have since proved to be of particular interest. One (3C 295) has since been identified with a very remote Galaxy (redshift $z = 0{\cdot}46$), and the other (3C 196) tentatively with a galactic star (Burbidge, Chapter 16). The third (3C 147) remains unidentified, probably because of heavy optical obscuration in that direction.

RECENT RESULTS

In the last few years several interferometers have come into use which have permitted the study of some hundreds of sources smaller than two or three minutes of arc. Some details of these instruments are given in Table 14/II.

TABLE 14/II

Some interferometers used for angular diameter measurements

Observatory	Frequency	Baselines	Results	Reference
Cambridge	159 Mc/s	320λ EW 27λ NS	Catalogue of 471 sources	Edge *et al.*, 1959
Cambridge	178 Mc/s	465λ EW	Catalogue of 64 sources	Elsmore *et al.*, 1959
Cambridge	178 Mc/s	465λ EW	Statistical analysis of diameters of 600 sources	Leslie, 1961
Sydney	85 Mc/s	30λ NS 3000λ EW	90 diameters in catalogues of Mills *et al.*	Mills *et al.*, 1960–61
Cal. Tech.	960 Mc/s	195–1557λ EW & NS	Catalogue of 195 sources	Moffet *et al.*, 1962
Nançay	1420 Mc/s	40–1000λ EW 40–1800λ NS	Catalogue of 25 sources	Lequeux thesis, 1962 Lequeux *et al.*, 1961
Jodrell Bank	158 Mc/s	2200–61,100λ EW	Catalogue of 384 sources and statistical analysis	Allen *et al.*, 1962

These instruments have produced a large amount of data, much of which is published, or in publication. At last the resolving powers and signal/noise ratios appear adequate to give some angular information about a statistically useful number of sources, as is clear from Fig. 14/2. This summarizes the measurements of Moffet (1961), Maltby (1961) and Allen *et al.* (1962) for a group of 55 sources having galactic latitudes $b^{\text{I}} > 12°$, and shows how the percentage of sources having values of fringe visibility $|\gamma| \geqslant 0{\cdot}7$ and $|\gamma| \geqslant 0{\cdot}5$ diminishes progressively as the baseline is lengthened.

These sources are almost all of them more intense than 20

flux units (at 159 Mc/s) and for this group about 50 per cent have an angular scale larger than 30″, and only 10 per cent are smaller than 4″.

A detailed study of the measurements on individual sources shows that there is very little correlation between the angular size of a source and the flux density received from it, and that

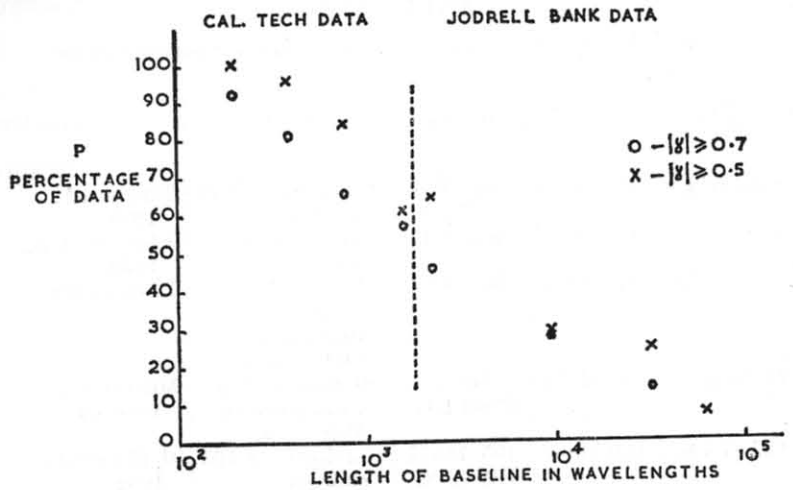


FIG. 14/2.—Data from Cal. Tech. and Jodrell Bank have been used to find the percentage (P) of sources with (a) fringe visibility $\gamma \geqslant 0{\cdot}7$ and (b) $\gamma \geqslant 0{\cdot}5$. The observations refer to a restricted set of 55 sources (described in the text) and were made with baselines of 195λ to $61{,}100\lambda$.

their structure is usually more complicated than the single circular gaussian brightness distribution conventionally assumed until recently. Maltby and Moffet (1961) classified the structures of the extragalactic sources they studied in detail as 'double', 'halo and core', 'simple' and 'unresolved', as shown in Table 14/III. Allen *et al.* (1962) have now resolved almost all the sources in a rather larger group, though without obtaining such detailed structures. They conclude from statistical arguments that at least two-thirds of the sources lying at galactic latitudes $\geqslant 12°$ probably have a double structure.

This complexity of angular structure leads to considerable difficulties in the analysis and reduction of the results. It is no longer permissible to assume that the sources are circularly symmetrical and the 'diameter' calculated from an isolated

TABLE 14/III

A classification of the structures determined by Maltby and Moffet for the more intense extragalactic sources

Structure class	No. of sources
Double sources	
1. Component intensity ratio $\leqslant 1{\cdot}5 : 1$	11
2. „ „ „ $> 1{\cdot}5 : 1$	24
Halo and Core	6
Simple	4
Unresolved	45

value of the fringe visibility at a particular spacing may not bear much relation to the angular dimensions of the source, when these are eventually determined. A realistic interpretation for any particular complex source requires observations at many values of aerial separation, with baselines in two or more directions. It would probably be extremely informative to have such data for all the sources having flux densities greater than some limiting value, provided that value were set low enough to admit a statistically useful number of sources and not, as in Table 14/I, at about 1000 flux units, so that the sample only includes 4 or 5 sources. Such data are not yet available from any of the instruments listed in Table 14/II. Only Moffet and Maltby at the Owens Valley Observatory in California, and Allen *et al.* at Jodrell Bank, have so far had adequate resolving power together with a signal/noise ratio which is good enough for this purpose. The observers in California have used several baselines $L \leqslant 1557\lambda$, in two azimuths, while at Jodrell Bank we have used four EW baselines with lengths in the range 2200λ–$61{,}100\lambda$. Although these observations were made at frequencies which differed by a factor of six, the results from the two observatories agree well on the average, as was seen in Fig. 14/2. This is usually found to be the case also when the sources are considered individually, showing that, on the whole, angular structure does not change with frequency. The exceptions are 'halo and core' objects, for which in some cases the haloes are less pronounced at high frequencies.

The most complete angular information available at the present time for a usefully large set of sources may be derived

by combining the results from both observatories. One can then begin to discern some general features in the structure of radio sources. From such considerations a 'general' model has been proposed (Palmer, 1961) which is reproduced in Fig. 14/3. It

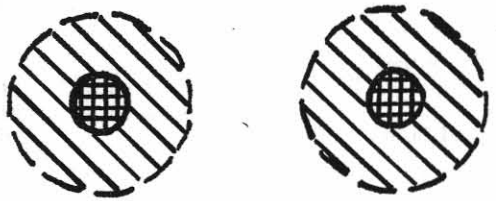


FIG. 14/3.—The proposed general model of radio source structure, which has two separated components, each of which has a bright 'core' and a broad 'halo'.

has two separated components, each of which has a bright core surrounded by a 'halo' of lower surface brightness. As the components may not have equal diameters, and the flux density received from the two components may also be unequal and may be variously divided between the core and halo, there are clearly many parameters associated with each source, and the existing data are not adequate to define all of them.

Allen *et al.* (1962*b*) have considered the set of 55 sources mentioned above, which were well observed at each observatory. In most cases the data are sufficient to justify the fitting of a visibility function, as in the examples of Fig. 14/4. Angular sizes can therefore be assigned to the components and, for some double sources, to their separations. A histogram of the component diameters is given in Fig. 14/5*a*, which shows that there is wide dispersion in the angular sizes which vary from less than 1″ to about 300″.

This information about component diameter is derived from the envelope of the visibility amplitude function, while the separation between the components is derived from the spacing between successive minima of the function. In general, the first baseline at which $|\gamma|$ is significantly less than 1·0 is likely to be highly sensitive to the orientation of the source major axis, while the longest baseline at which fringes were observed may merely reflect the existence of small hot regions in the radiating components, colloquially known as 'pips'. With these considerations in mind, the following rule has been adopted as one method

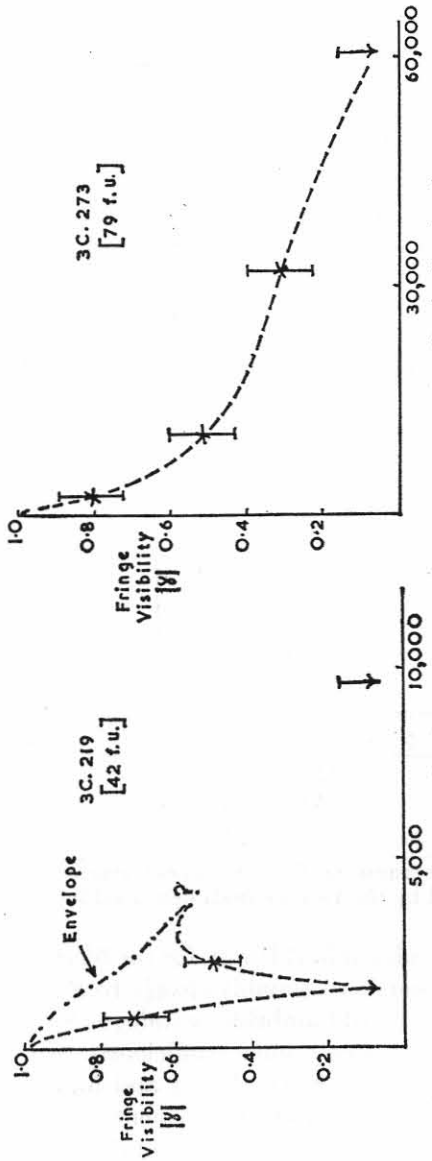


FIG. 14/4.—The observed values of fringe visibility for two sources, with approximate limits of error. The dotted lines show possible forms of the visibility functions.

of analysing the Jodrell Bank results. A value of angular diameter θ is calculated from the data at the longest baseline at which $|\gamma| \geqslant 0{\cdot}4$. This value of θ is, in general, taken to represent the angular sizes of the radiating components, and may be calculated even where the data are too sparse to permit the fitting of a visibility function. Fig. 14/5*b* shows the distribution

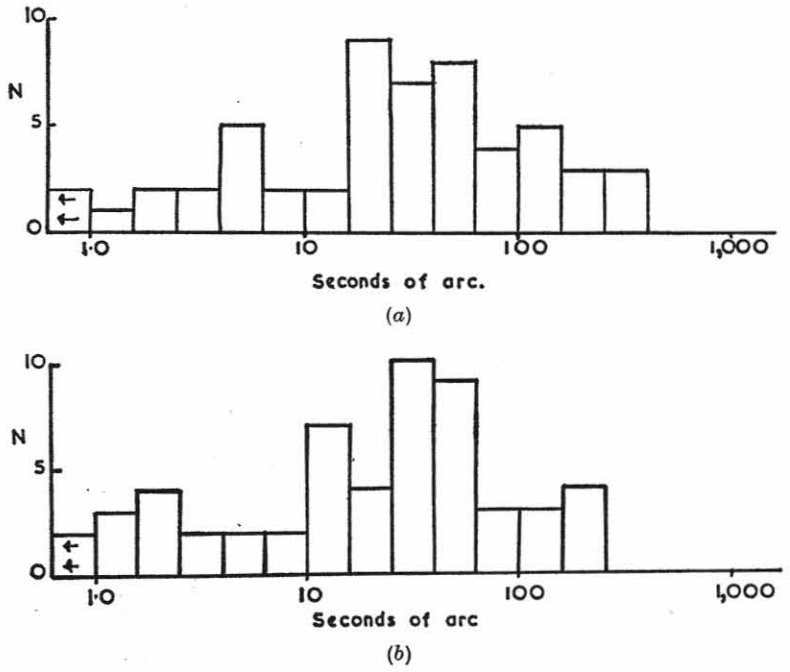


FIG. 14/5.—Histograms of the component diameters of 55 sources, derived by the two methods described in the text.

of diameters calculated in this way for the 55 sources mentioned above. It corresponds reasonably closely to Fig. 14/5*a*. Most of the sources in the third Cambridge catalogue having $S \geqslant 12$ f.u. and $b^{1} \geqslant 12°$ have been studied sufficiently to justify analysis by this method (Allen *et al.*, 1962*a* and *b*). Similarly, if one analyses, by this method, the component diameters for the 70 most intense sources in the third Cambridge catalogue, which Dewhirst (Chapter 17) discussed from the point of view of optical identifications, one finds a similarly wide distribution, as shown in Fig. 14/6. The median angular size for the sources

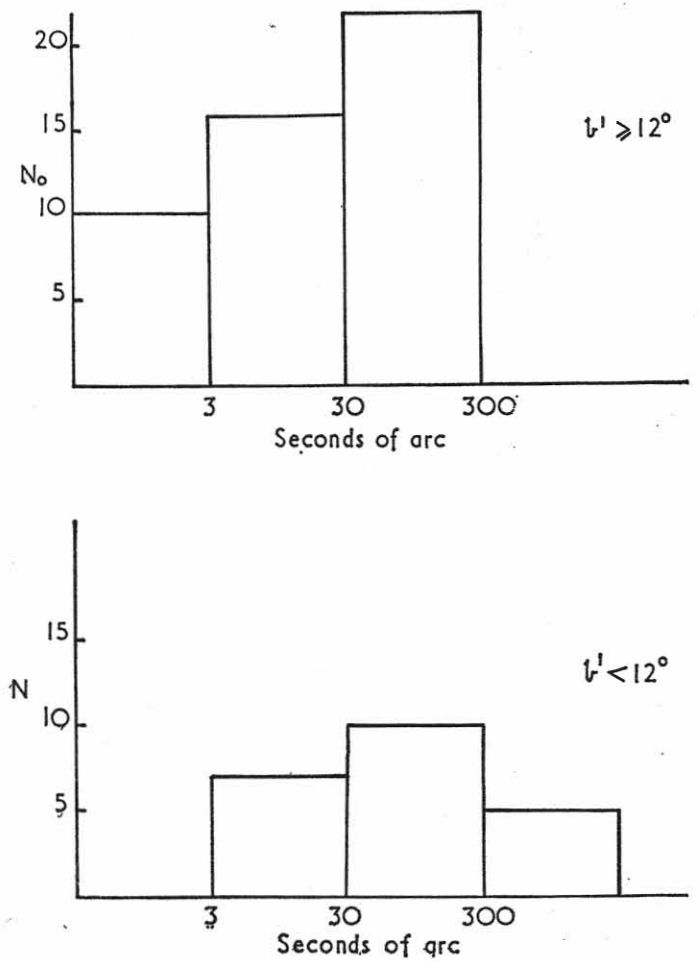


FIG. 14/6.—Histograms of the angular sizes of the 70 most intense sources in the 3C catalogue (flux density $\geqslant$ 25 f.u.) for two ranges of galactic latitude.

at low galactic latitudes is about three times larger than for the sources at high latitudes.

CONCLUSION

The available data suggest that the angular dimensions of sources of flux $S \geqslant 12$ flux units, and $b^{1} \geqslant 12°$ may be briefly described as follows:

More than two-thirds of the sources are probably double, with two separated components. The median angular separation between the components of double sources is approximately 120″, in the range 3″ to 1000″. The median component diameter is 30″ in the range 1″ to 300″.

REFERENCES

Allen, L. R., Anderson, B., Conway, R. G., Palmer, H. P., Reddish, V. C., and Rowson, B., 1962a, *M.N.R.A.S.*, **124**, 477.

Allen, L. R., Hanbury Brown, R., and Palmer, H. P., 1962b, *M.N.*, in publication.

Edge, D. O., Shakeshaft, J. R., Baldwin, J. E., McAdam, B., and Archer, S., 1959, *Mem. R.A.S.*, **68**, 37.

Elsmore, B., Ryle, M., and Leslie, P. R. R., 1959, *Mem. R.A.S.*, **68**, 61.

Hanbury Brown, R., Jennison, R. C., and Das Gupta, M. K., 1952, *Nature*, **170**, 1061.

Hanbury Brown, R., Palmer, H. P., and Thompson, A. R., 1955, *Phil. Mag.*, **46**, 857.

Jennison, R. C., and Das Gupta, M. K., 1953, *Nature*, **172**, 986.

Lequeux, J., 1962, Thesis, Observatoire de Paris.

Lequeux, J., and Heidman, J., 1961, *Comptes Rendus*, **253**, 804.

Leslie, P. R. R., 1961, *M.N.*, **122**, 51.

Maltby, P., 1961, *Nature*, **191**, 793.

Maltby, P., and Moffet, A. T., 1962, Observations of Cal. Tech. No. 1.

Michelson, A. A., 1920, *Ap. J.*, **51**, 257.

Mills, B. Y., 1953, *Aust. J. Phys.*, **6**, 452.

Mills, B. Y., Slee, O. B., and Hill, E. R., 1960, *Aust. J. Phys.*, **13**, 676.

Mills, B. Y., Slee, O. B., and Hill, E. R., 1961, *Aust. J. Phys.*, **14**, 497.

Morris, D., Palmer, H. P., and Thompson, A. R., 1955, *Phil. Mag.*, **46**, 857.

Palmer, H. P., 1961 I.A.U. Symposium 15, *Problems of Extra-galactic Research.*

Pawsey, J. L., and Bracewell, R. N., 1955, *Radio Astronomy*, O.U.P.

Rowson, B., 1962, *M.N.*, in publication.

Smith, F. G., 1952, *Nature*, **170**, 1061.

Twiss, R. G., Carter, A. W. L., and Little, A. G., 1960, *Observatory*, **80**, 153.

15

SOURCE BRIGHTNESS DISTRIBUTION

B. ROWSON

As described by H. P. Palmer in Chapter 14, transit observations made at Jodrell Bank with a few baselines covering a wide range of resolutions have shown that many discrete radio sources have a complex structure and must consist of at least two components. This conclusion was derived from a statistical analysis on a large number of sources because too little information was obtained from any one individual source. However individual sources have been studied in detail at lower resolution and similar conclusions reached (Moffet and Maltby, 1961). It is clearly desirable to investigate the structure of a few individual sources at the high resolution obtainable with the former equipment for comparison with the structures obtained at lower resolution.

From the theoretical standpoint the simplest way of obtaining the brightness distribution of a source is to scan it with an aerial which gives the required resolution and produce isophotes of the source. Since the observations described here were made with a resolution of a few seconds of arc at a wavelength of 1·89 m it is clear that some more indirect method of observation has to be used. If we were to use an aperture synthesis method similar to that described by M. Ryle in Chapter 18 many sites would have to be used for the remote aerial and the experiment would take a prohibitive length of time. There would also be the problem of preserving the phase of the fringes when using baselines of length comparable with the size of the irregularities of ionization density in the ionosphere. Let us consider how much information can be obtained by observing with a fixed baseline using only the changes of position angle and resolution that are caused by the apparent motion of the source across the sky, and how these observations can be made in practice.

In order to define the direction of the baseline in celestial co-ordinates, consider the baseline produced to intersect the

celestial sphere at a point B which will have an hour angle and declination (h, d). If the source S has the celestial co-ordinates (H, D) the cosine rule (e.g. Smart, 1947) can be used in the spherical triangle NSB (Fig. 15/1) to give

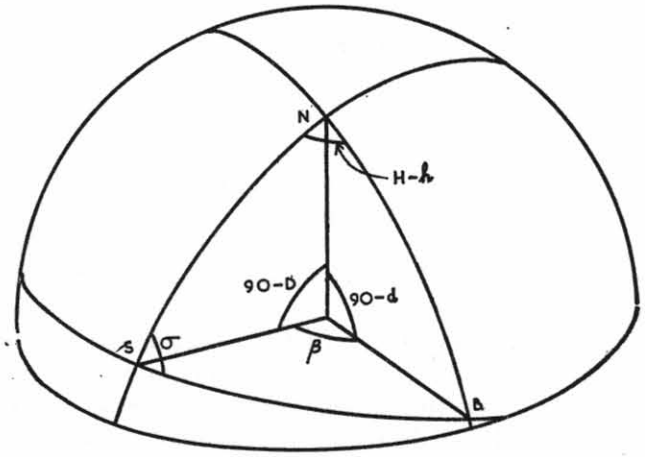


FIG. 15/1.—The spherical triangle used for calculating differential delays and resolution.

$$\cos\beta = \sin d \sin D + \cos d \cos D \cos (H-h) \quad . \quad (1)$$

where β is the angle between the direction of the baseline and the line of sight from the observer to the source.

The difference in time of arrival of a given wavefront at the two aerials is

$$T = \frac{L}{C}\cos\beta \quad . \quad . \quad . \quad (2)$$

where L is the length of the baseline and C is the velocity of light. The natural fringe frequency (i.e. that which would be observed if no phase rotation were introduced into the equipment) is given by the rate of change of this differential delay times the frequency of observation. Thus the fringe frequency F is given by

$$F = \frac{dH}{dt}\frac{Lf}{C}\cos d \cos D \sin (H-h) \quad . \quad . \quad (3)$$

where H is measured in radians and f is the observing frequency.

For the present baseline of 61,100 wavelengths (which corresponded to a length of 70 miles) continuous phase change and

tracking of the delay were required. The above equations were therefore used to design a simple analogue computer which drove a magslip phase rotator to control the frequency of the fringes recorded on the chart (Elgaroy, Morris and Rowson, 1962) and also varied a mercury delay line which was used to delay the home station signal at intermediate frequency. Having built electronic equipment which would follow a source over many hours it was necessary to use aerials which would do the same. The interferometer used the 250-ft radio telescope together with a distant aerial consisting of a small broadside array of full wave dipoles giving a collecting area of 36 m² supported on an alt-azimuth mounting.

In order to interpret the results it is convenient to consider a hypothetical experiment in which the source is considered fixed in the zenith while one interferometer aerial is fixed at the origin on a horizontal plane. The position of the other movable aerial on this 'fringe amplitude' plane can then be given in terms of the length and direction of a radius vector from the origin or in terms of rectangular cartesian co-ordinates (x, y) with the y co-ordinate increasing in a northerly direction. Then if this aerial is moved over the plane to simulate the results that are obtained with the real equipment as the source is tracked, the length of this radius vector in the hypothetical experiment must be equal to the foreshortened length (N') of the real baseline, while its direction is given by the position angle (σ) of the direction in which resolution is obtained on the celestial sphere with the real equipment. Thus when a source is observed at a given hour angle H the corresponding values of x and y are given by

$$x = N' \sin\sigma = N \sin\sigma \sin\beta \quad . \quad . \quad (4)$$

$$y = N' \cos\sigma = N \cos\sigma \sin\beta \quad . \quad . \quad (5)$$

where N the length of the actual baseline measured in wavelengths is given by

$$N = \frac{Lf}{C}$$

Further application of the sine and cosine rules respectively in the spherical triangle NSB gives

$$\sin\sigma \sin\beta = \cos d \sin (H-h) \quad . \quad . \quad . \quad (6)$$

$$\cos\sigma \sin\beta = \frac{\sin d - \cos\beta \sin D}{\cos D} \quad . \quad . \quad (7)$$

Using equations (1), (6) and (7) to eliminate σ and β from equations (4) and (5) we obtain

$$x = N \cos d \sin (H-h) \quad . \quad . \quad . \quad . \quad . \quad (8)$$

$$y = N[\sin d \cos D - \cos d \sin D \cos (H-h)] \quad . \quad (9)$$

Two deductions can be made from these equations. Firstly equation (8) shows that, for a given actual baseline, the east–west component of the resolution depends only on the hour angle of the source and not on its declination. Secondly it can be shown that the locus of the point of observation on the fringe amplitude plane is an ellipse, for if we have an ellipse with semi-major axis a and semi-minor axis b its equation can be given in terms of the parameter ϕ as

$$x = a \sin \phi$$

$$y = b \cos \phi$$

Comparing these equations with equations (8) and (9) it is seen that the observational ellipse must have its centre at (O, N sin d cos D) and its parameter $\phi = (H-h)$. Fig. 15/2 shows three ellipses of observation for three different source declinations when using the present baseline which is nearly horizontal but makes an angle 10° to the east–west direction. In the special case of a horizontal east–west baseline $d = 0$ and the observations follow an ellipse centred on the origin as noted by Ryle (1962).

Thus it is evident that insufficient information is obtained to derive the brightness distribution uniquely from observations made with a single baseline and so models of the source brightness distribution have to be assumed and deductions from these models compared with the results. In this way it is hoped to find the simplest model that is consistent with the experimental results. This procedure would be necessary in any case with this equipment because measurements of phase cannot be made.

Because this method of interpreting the results depends on being able to compare the fringe visibility distributions from physically reasonable source models with the experimental results, it is convenient to have a physical analogue of the experiment. This is provided by Fraunhofer diffraction which might be produced by the equipment shown schematically in

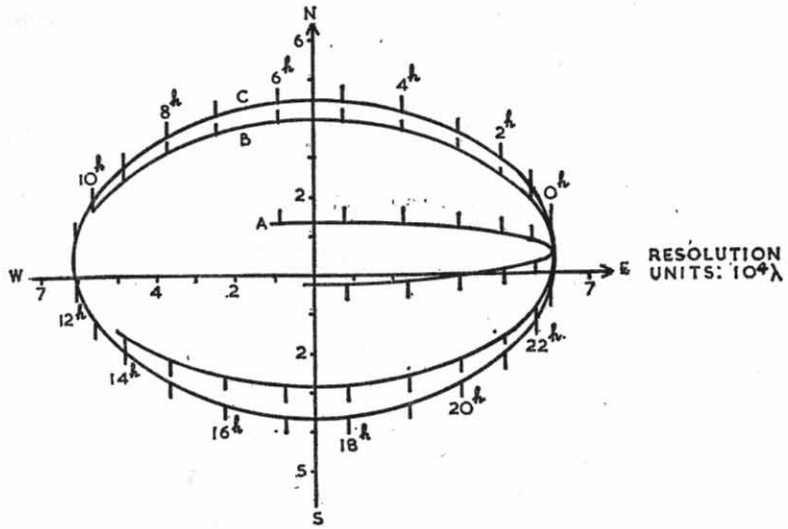


FIG. 15/2.—Some examples of loci of observations on the fringe amplitude plane. The three ellipses refer to three different source declinations observed with the present baseline and are as follows:

A: declination = 7°
B: declination = 33°
C: declination = 41°

Fig. 15/3. In this apparatus the origin is to be taken as the point where the optical axis intersects the image plane. Then if a model of a source in the form of a positive transparency is placed between the two lenses, it can be shown that the square root of the light intensity on the image plane is proportional to the fringe amplitude in the interferometer experiment. Of special interest is the source model consisting of two centres of emission of comparable intensity. Here the optical analogue clearly shows that the fringe amplitude distribution must consist of straight 'ridges' and 'valleys' of high and low fringe visibility with one ridge passing through the origin.

Because of the complexity of the equipment used in long baseline observations it is not possible to obtain a direct calibration from a knowledge of the equipment parameters. However a calibration is required in order to relate the amplitudes of the observed fringes to the amplitude that would be observed at low resolution. This calibration is obtained by observing a source of known intensity which is assumed to be unresolved at

the baseline used. The calibration source used in these observations was 3C 48, which was assumed to have a flux of

$$50 \times 10^{-26} \text{ w m}^{-2}(\text{c/s})^{-1}$$

Then, providing that the intensities of the other observed sources are known, the observations can be expressed in terms of fringe visibility which is defined as the ratio of the fringe

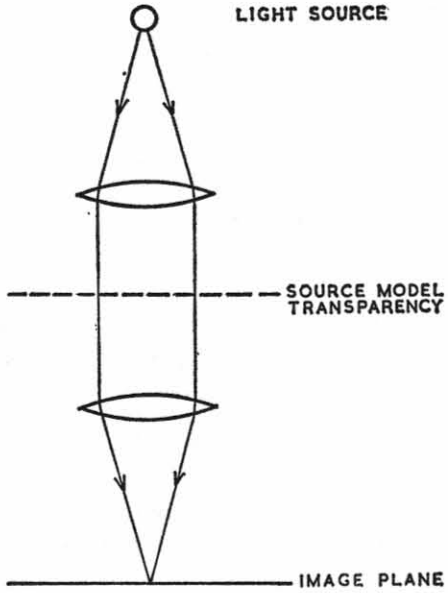
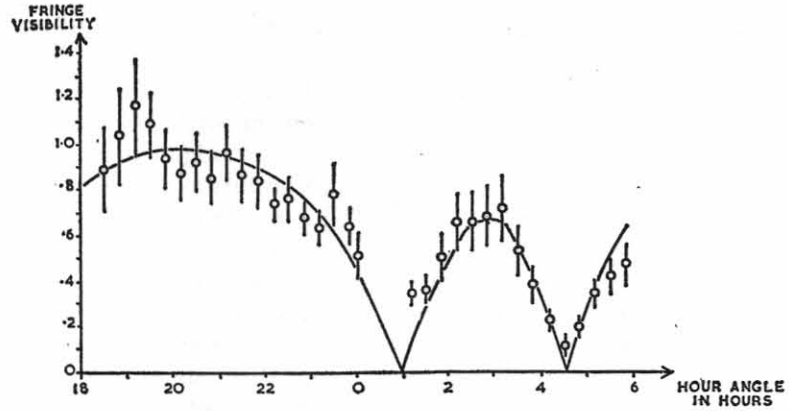


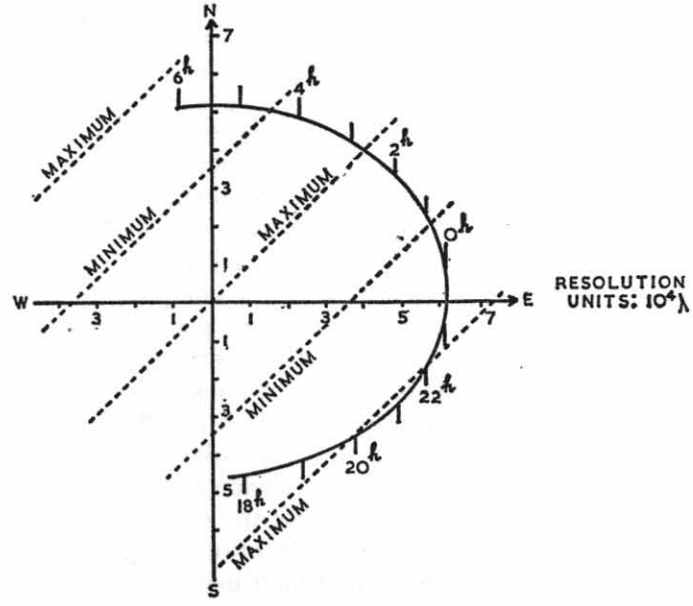
FIG. 15/3.—An optical analogue for deriving the distribution of fringe amplitude from a given source model.

amplitude which is observed with high resolution to that which would be observed at zero resolution with equipment of the same sensitivity.

As an illustration of this method of observation, consider the results obtained from the radio source 3C 295. These are shown in Fig. 15/4, together with the ellipse of observation. It can be seen that the two minima and also the maximum between them occur at much the same resolution and over a position angle range amounting to only 60 degrees. As the fringe visibility is fairly high, this behaviour cannot be reconciled with a source consisting of a single centre of emission and so a double source will be considered. Now since the observations cover a fairly



(*a*) Observations of the source 3C 295. Variation of fringe visibility with hour angle. The curve represents the calculated fringe visibility for a source model having two components of equal intensity.



(*b*) The locus of the observations on the fringe amplitude plane showing the positions of the maxima and minima derived from the source model finally adopted.

FIG. 15/4.

large range of position angles, one of the maxima in the fringe visibility curve must be caused by the fringe amplitude ridge passing through the origin. If we can identify this ridge we can get an approximate value for the position angle of the line joining the two centres of emission since the ridge is perpendicular to this line. (Once a complete interpretation of the results is made, a more accurate value of this position angle can be obtained from the hour angles of the minima.)

If we take the larger maximum as corresponding to the origin ridge the two minima which are observed must correspond to the same valley in the fringe amplitude distribution. Referring to the ellipse diagram it can be seen that this valley would run close to the locus of observations for hour angles of 2 to 3 h and so the fringe visibility over this region could not exceed 0·16. Thus take the maximum at an hour angle of 3 h as the ridge passing through the origin, making the two minima correspond to two separate valleys and giving the two components a separation of 4″ of arc in position angle 135 degrees. Then it is seen that the next ridge in the fringe distribution runs parallel to and nearly coincident with the ellipse for hour angles from 19 to 23 h and so accounts for the large observed width of this maximum. However two aspects of the results remain to be explained: first the large fringe amplitude at 19 h compared with that at 23 h although the distance of the ellipse from the secondary fringe amplitude ridge is similar in the two cases, and second the low fringe visibility for the maximum at 3 h. Both of these effects can be explained if the components have an extension of 1·7″ of arc along a direction perpendicular to the line joining the centres. Also since there is effectively unit fringe visibility at the 21 h maximum the components must have a width of less than 1″ of arc in a direction parallel to this line. The curve drawn in Fig. 15/4 (*a*) is the calculated fringe visibility for this model if the two centres of emission are assumed to have equal intensity. While the curve fits the experimental points reasonably well for most of the range of hour angles there is a suggestion that the minimum at 1 hour may not correspond to zero fringe visibility. If this were subsequently found to be the case it would be necessary to assume that the two components had fluxes that were not quite equal and major axes that were not quite parallel, the more intense component having the smaller major axis position angle.

Tracking observations have also been made on 3C 123, 147, 298 and 405 at this baseline suggesting in each case source models which differ significantly from circular gaussian brightness distributions (Rowson, 1962).

REFERENCES

Elgaroy, O., Morris, D., and Rowson, B., 1962, *M.N.R.A.S.*, **124**, 395.
Moffet, A. T., and Maltby, P., 1961, *Nature*, **191**, 453.
Rowson, B., 1962, *M.N.R.A.S.*, in publication.
Ryle, M., 1962, *Nature*, **194**, 517.
Smart, W. M., 1947, *Spherical Astronomy*, Cambridge University Press.

16

PHYSICAL PROCESSES IN NON-THERMAL RADIO SOURCES

G. R. BURBIDGE

ALTHOUGH radio astronomy has had a chequered history so far as cosmology is concerned, it has certainly produced some fantastic results in high-energy astrophysics if the few identifications of radio sources with optical objects are accepted. There are four distinct types of non-thermal radio source which I shall discuss,* and in each I shall accept that the method of emission is the synchrotron process in which high-energy electrons, and probably also positrons, radiate when they are accelerated in weak magnetic fields.

RADIO EMISSION FROM STARS

Until quite recently the only star from which we could detect radio emission was the Sun, although in the early days of radio astronomy it was thought that most of the radio sources in the sky might be low-luminosity stars. At this time it is not appropriate to go into the many experimental and theoretical arguments that have been advanced to show why this is not so. With the very accurate positions that have been obtained now it has been possible to make identifications of a few sources with stars. The first such identification, which was reported from the Mt. Wilson and Palomar Observatories (by Dr. Sandage to the American Astronomical Society) in the summer of 1961, was that of the radio source 3C 48. The identification was based originally on the position measurement, and more recently the star has been examined both spectrographically and by photo-electric means at Mt. Wilson and Palomar. The object is a 16th magnitude star with a very faint associated nebulosity, so faint indeed that it would be very difficult to reproduce it for publication. Spectroscopically the object is disappointing.

* Not including the very interesting non-thermal emission from Jupiter.

The line at 4686 Å due to He II is the only feature which has been positively identified. The presence of other features in the red and in the ultraviolet (at about 3830 Å) is indicated, but these have not been identified. There is some evidence that the optical intensity of this star has been changing by about one magnitude during the past year. The variations are apparently irregular, and no definite data are yet available.

I am indebted to Dr. Sandage for the following data on the colour and magnitudes of 3C 48 and also of two other objects (3C 196 and 3C 286), which he and Matthews are now fairly sure are good identifications of radio sources with stellar objects (see Table 16/I). In addition there is the object 3C 346 which Mrs. Burbidge has been interested in for some time. Here the identification is much less certain mainly because the radio position is known with much less precision. Fig. 16/1, which was taken by Mrs. Burbidge at Lick, shows this object; the others are very similar in appearance. The faint nebulosity is slightly displaced from the star, but the probability of its being an accidental coincidence is thought to be small. However, this is a new and difficult field, and hitherto astronomers have generally neglected such faint nebulosities.

TABLE 16/I

Colours and magnitudes of three stars that are thought to be radio sources

	V	B–V	V–B
3C 48	16·20	0·40	−0·59
3C 196	17·79	0·57	−0·43
3C 286	17·25	0·26	−0·91

The colour indices for the three stars shown in Table 16/I are very abnormal and as Sandage points out they correspond to the kind of colour one finds for ex-novae, perhaps in some cases, white dwarfs.

I think it is fair to accept the identification of these stars as radio sources, but the absence of any optical or radio evidence of distance and the meagre spectral information make it difficult

to say what contribution stellar radio sources might be making to the background emission. Also at the moment we have only very poor information on the number of sources of high surface brightness which may be stars. None the less these objects may be very significant in the interpretation of the galactic continuum emission, as well as in cosmic ray production. For if they are intrinsically powerful sources at moderately great distances there may be perhaps a hundred of them in the Galaxy, and it is quite possible that they represent a more important source of high-energy particles than do the supernovae.

In view of the uncertainties about the absolute power output, it is not very fruitful to discuss the energetics of radio emission from stars. However, it is clear that conditions are rather different from such objects as the Crab nebula, for the magnetic field may be greater, and much lower energy electrons than in the Crab will then radiate appreciably by the synchrotron mechanism. We are then faced with the problem of the many other energy loss processes which compete with the synchrotron process at lower energies.

SUPERNOVA REMNANTS

There are a number of strong non-thermal sources in our own Galaxy which have been identified with nebulosities that in many cases are thought to be supernova remnants. Minkowski has attempted to determine the ages of these objects by measuring the expansion velocity of the shells, and making assumptions about the deceleration. In the Crab nebula, the light ripples observed by Baade which move with velocities $\sim 0{\cdot}2 - 0{\cdot}3$ c are probably manifestations of hydromagnetic activity (as was suggested by Shklovsky) which originate in the central star. It is clear that the charged particles responsible for the synchrotron emission in the Crab are being produced continuously. Numerous suggestions are current concerning the nature of the central star, one of these being the idea that it is a neutron star. In the other remnants it is not clear whether or not high-energy electrons are still being injected. Calculations by the Russian theoreticians indicate that the supernovae may supply sufficient particle flux to account for the entire cosmic ray density in the Galaxy.

Turning now to the energy considerations in supernovae, we

find that the radio emission obeys approximately a power law of the form

$$\text{flux density} \propto \nu^{-x}$$

From this we can deduce from the theory of synchrotron emission that the energy density in the relativistic charged particles is given by a power law of the form

$$N(E)\,dE \propto E^{-n}\,dE$$

where $N(E)\,dE$ represents the number of particles in the energy range E to $E + dE$, and

$$n = 2x + 1$$

If we now assume that the radio spectrum cuts off at say 10 Mc/s and 10^4 Mc/s, we can use knowledge of the distance and size of the supernova to calculate the total energy.

The power and frequency radiated by an electron depend on the magnetic field; since no independent measure of this is available, it is usual to assume equipartition of energy between the magnetic energy density and the particle energy density. It is now well established that this assumption yields a lower limit to the total energy needed to explain the radio flux. This does not imply that there is any physical basis for expecting equipartition of energy to be realized.

From the synchrotron theory, one calculates only the total energy in electrons and positrons but it is generally thought that any physical process which gives rise to energetic electrons must also produce a considerable flux of protons. The synchrotron energy loss from these charged nucleons is negligible, with the result that the loss of energy of these particles will occur by nuclear collisions at a rate which depends on the density. The net result will be that a cloud of π mesons will be produced. The π_0 mesons will decay to γ rays, and the $\pi+$ and $\pi-$ mesons will decay to μ mesons which in turn decay to positrons and electrons and neutrinos. Thus the energy in the proton flux will be transferred into electrons, positrons, γ rays and neutrinos at a rate depending on the density. These 'secondary' charged particles are the principal source of the synchrotron emission. By going through the calculations involving multiple meson production it is possible to work out that the ratio of energy in the primary protons to the energy in the primary electrons is probably about 100 : 1. If this is so, and we equate the proton

energy density to the magnetic energy density, we arrive at rather higher total particle energies.

Table 16/II shows the total particle energy in some of the

TABLE 16/II

Minimum energies required for sources of acceleration radiation within the Galaxy

	Rate of emission (ergs/sec)	Total energy (electrons + mag. energy) (ergs)	Mean value of H (gauss)	Total energy (protons + mag. energy) (ergs)	Mean value of H (gauss)
*Crab Radio	8×10^{33}	$\sim10^{48}$	$\sim10^{-3}-10^{-4}$	$\sim10^{50}$	$\sim10^{-2}$
*Crab Optical	10^{36}				
Cassiopeia A	$2{\cdot}6\times10^{35}$	$4{\cdot}1\times10^{48}$	2×10^{-4}	$5{\cdot}7\times10^{49}$	1×10^{-3}
IC 443	4×10^{33}	8×10^{48}	1×10^{-5}	$1{\cdot}2\times10^{50}$	4×10^{-5}
*Cygnus Loop	$2{\cdot}5\times10^{32}$	10^{50} magnetic, 3×10^{46} electrons	—	10^{50} magnetic, 3×10^{48} protons	5×10^{-5}
Galactic centre source	$1{\cdot}4\times10^{36}$	$1{\cdot}0\times10^{52}$	1×10^{-5}	$1{\cdot}3\times10^{53}$	4×10^{-5}

* For these sources the equipartition condition has not been used.

supernovae, calculated on the basis that we have discussed. These results were first presented at the Paris Symposium (see references); the situation has not changed much since then.

NORMAL GALAXIES

Galaxies radiating at power levels similar to our own and to the Andromeda nebula are known as 'normal galaxies' and a very nice survey of their characteristics has recently been published by Hanbury Brown and Hazard (1961). It seems that whatever process is going on in our own Galaxy also applies to other normal galaxies. Supernovae are probably the principal source of the synchrotron electrons, with some particles originating in stars, and some arriving from outside the Galaxy.

INTENSE RADIO EMISSION FROM ABNORMAL GALAXIES

Fifteen or twenty strong radio sources have been positively identified with peculiar external galaxies. The peculiarities that appear are very diverse, and I will discuss a few of these identifications. First we have NGC 1068 (3C 71). No other identifica-

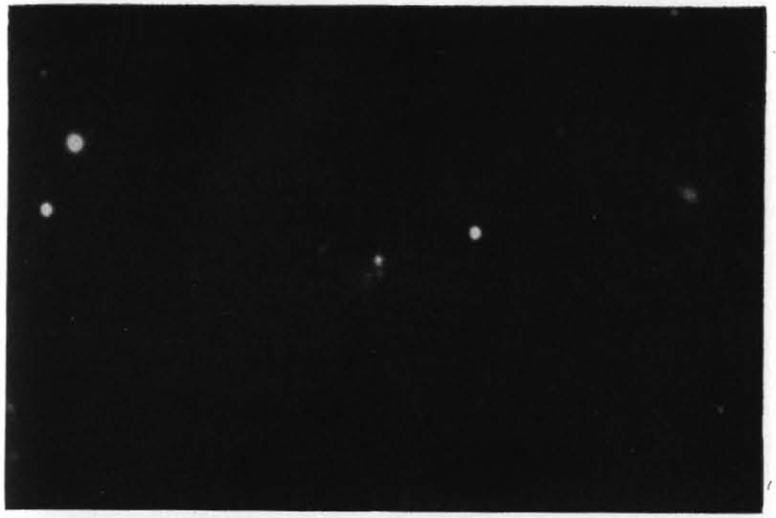


Fig. 16/1. The object tentatively identified with 3C 346. Lick Observatory plate.

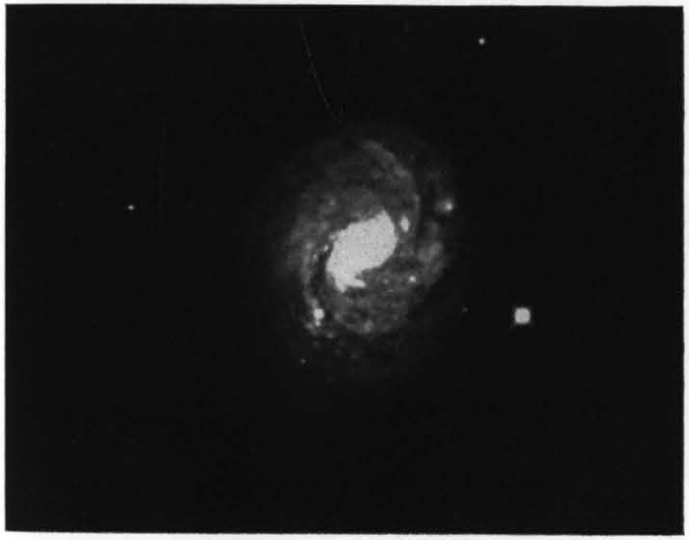


Fig. 16/2.—NGC 1068. McDonald Observatory plate.

FIG. 16/3.—NGC 5128. McDonald Observatory plate.

FIG. 16/4.—NGC 1316. McDonald Observatory plate.

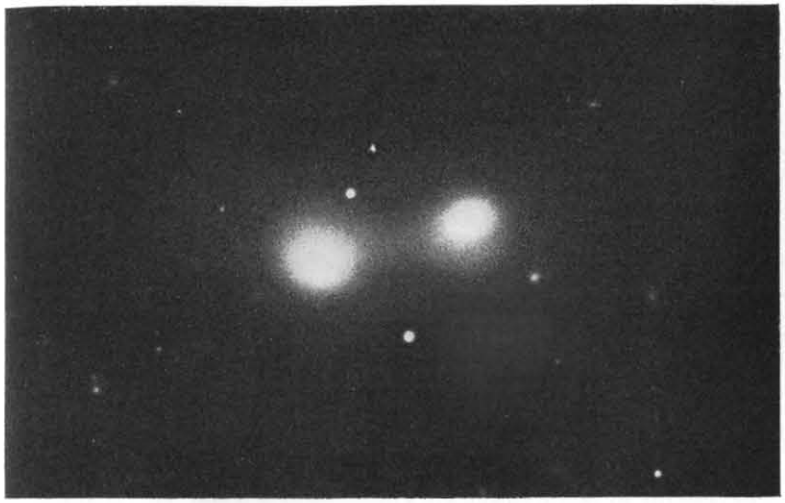


FIG. 16/5.—NGC 4782–83. Plate taken with 120-inch telescope. Lick Observatory.

tion of a radio source has been made with a galaxy of this type. It is an Sb galaxy, and the radio source has an unusually small linear size, being confined to a few kiloparsecs in the condensed central region of the spiral (see Fig. 16/2). The radio source is radiating about 10^{40} ergs/sec, and is thus among the weaker of the extragalactic sources, which have powers in the range 10^{40} to 10^{45} ergs/sec. It is interesting that this galaxy is one of a class of galaxies originally described by Seyfert (1943) as having very broad emission features. They are very rare, only about a dozen being known. One other such galaxy, NGC 1275, is also a radio source. In NGC 1068 studies of the mass distribution and Doppler broadening have shown that in the nucleus the random velocities are greater than the escape velocity, so that material must be escaping at a considerable rate. Secondly, in the Virgo radio source, identified with M87, we have a giant elliptical galaxy which shows a remarkable and highly polarized jet. Such a feature could exist in other galaxies and be undetected on account of the distance or orientation, as M87 is in the Virgo cluster and is thus fairly close to us.

Thirdly, we have a completely different type of object, the famous Centaurus source, NGC 5128. The radio source is gigantic, subtending an angle of some 10 degrees, corresponding to a dimension of $\frac{3}{4}$ Mpc. Fig. 16/3 is a plate of the optical galaxy. Apart from the strong absorption feature, it has the appearance of a bright elliptical galaxy. There is no evidence to support the idea that this is a collision of a small spiral galaxy with a large elliptical. Observations by David Evans at Pretoria (Evans and Harding, 1961) and more recently by the author and Mrs. Burbidge (1962) have shown a small difference between the mean velocity of the rotating gaseous component in and near the dust lane and the mean velocity of the stellar component, instead of the gross effect that one would expect to see if it were a collision. No rotation of the stellar component in the direction suggested by its slightly elliptical outer contour was found, but a small rotation about the same axis as the axis of rotation of the gas and in the same sense may be present. A tentative model may be put forward on the basis of this evidence, in which an elliptical nebula is surrounded by a ring of dust which is rotating rather more rapidly than the galaxy, and also falling in. However this model is unsatisfactory from a dynamical point of view.

As we have seen, the radio source is very large, so that we are dealing with intergalactic dimensions rather than galactic. This seems to be true of the majority of extragalactic radio sources.

Fourthly, in NGC 1316, we have an SO galaxy which is also a radio source. The absence of emission features indicates an absence of gas. Some dust lanes are visible, but there is no reason to think that this source is very similar to the Centaurus source. It is a strong radio source (Fornax) but very little is known about it (Fig. 16/4). Fifthly we have NGC 6166, recently identified with 3C 338 by Minkowski. Four galaxies appear within the same optical envelope. The object lies in a small cluster of galaxies, but it is most probable that the four are associated, and that at least one of them is the radio source. On a short-exposure ultraviolet plate it is apparent that one of the four components has a small bright central core which is probably mostly emission in the [O II] pair of lines at 3727 Å. Another radio source, 3C 310, has also been identified with a group of elliptical galaxies; it is not certain which member of the group is the radio source.

Finally the radio source 3C 278 has been identified with a strange double elliptical system (NGC 4782–83) in which both of the galaxies show an unusual asymmetry (Fig. 16/5). The relationship of this feature to the radio emission is obscure. The structure of the two galaxies is different; one has a greater degree of central concentration than the other.

At this stage I would like to remind you of the calculations one can do on the energy requirements for extragalactic radio sources. In Table 16/III are tabulated the calculations of energy for some of the well-known radio sources. These have recently been brought up to date by Dr. Moffet and Dr. Maltby at California Institute of Technology (1962). These figures for the total energy are so large that to my mind they represent one of the most astonishing contributions of radio astronomy to astrophysics.

For some years it was accepted that the radio source Cygnus A arose from the collision of two galaxies, but it is difficult to maintain this from energetic considerations. A total energy of 10^{60} to 10^{61} ergs is required, and if this is to be produced by the collision of two galaxies each containing a mass of gas of about 10^{10} M$\odot$ at 10^3 km/s, all the kinetic energy must be converted to radio energy with very high efficiency. This is a pretty impossible requirement. The strongest argument against the collision hypothesis is that elliptical systems contain very little gas. In a collision, the stars will not collide. If there is no gas present, it is very difficult to see how any kinetic energy could possibly be converted to radio power. If the collision hypothesis is ruled out, we have to look for a source of energy which can occur in a single galaxy.

TABLE 16/III

Minimum energies required for sources of acceleration radiation in galaxies

	Rate of emission (ergs/sec)	Total energy (electrons + mag. energy) (ergs)	Mean value of H (gauss)	Total energy (protons + mag. energy) (ergs)	Mean value of H (gauss)
*Galaxy	$\sim 10^{38}$	$\sim 3\times10^{54}$ (electrons) $\sim 10^{56}$ (mag. field)	—	$\sim 3\times10^{56}$	7×10^{-6} (disk) 2×10^{-6} (halo)
M 31	$1{\cdot}9\times10^{38}$	$2{\cdot}1\times10^{55}$	8×10^{-7}	$3{\cdot}0\times10^{56}$	3×10^{-6}
Magellanic Clouds	$1{\cdot}3\times10^{37}$	$2{\cdot}5\times10^{54}$	1×10^{-6}	$3{\cdot}4\times10^{55}$	4×10^{-6}
NGC 4038–9	$2{\cdot}1\times10^{39}$	$1{\cdot}7\times10^{56}$	2×10^{-6}	$2{\cdot}3\times10^{57}$	7×10^{-6}
NGC 1068	$7{\cdot}5\times10^{39}$	$3{\cdot}2\times10^{55}$	2×10^{-5}	$3{\cdot}6\times10^{58}$	6×10^{-5}
NGC 5128 (central region)	$2{\cdot}4\times10^{41}$	$3{\cdot}2\times10^{56}$	2×10^{-5}	$4{\cdot}4\times10^{57}$	9×10^{-5}
NGC 5128 (halo)	$2{\cdot}2\times10^{41}$	$5{\cdot}0\times10^{58}$	1×10^{-6}	$7{\cdot}0\times10^{59}$	5×10^{-5}
NGC 1316 (central region)	8×10^{40}	$2{\cdot}1\times10^{56}$	2×10^{-5}	$3{\cdot}0\times10^{57}$	6×10^{-5}
NGC 1316 (halo)	$1{\cdot}6\times10^{41}$	$1{\cdot}8\times10^{58}$	1×10^{-6}	$3{\cdot}2\times10^{59}$	5×10^{-6}
NGC 4486 (jet)	$2{\cdot}3\times10^{42}$	$1{\cdot}7\times10^{54}$	2×10^{-4}	$2{\cdot}4\times10^{55}$	7×10^{-4}
NGC 4486 (central radio source)	$3{\cdot}5\times10^{41}$	$1{\cdot}7\times10^{57}$	1×10^{-5}	$2{\cdot}4\times10^{58}$	4×10^{-5}
NGC 1275	$6{\cdot}4\times10^{41}$	$9{\cdot}4\times10^{56}$	2×10^{-5}	$1{\cdot}3\times10^{58}$	8×10^{-5}
NGC 6166	$7{\cdot}8\times10^{42}$	$1{\cdot}4\times10^{57}$	3×10^{-5}	$1{\cdot}9\times10^{58}$	1×10^{-4}
Hydra A	$1{\cdot}5\times10^{43}$	$1{\cdot}0\times10^{58}$	8×10^{-5}	$1{\cdot}5\times10^{59}$	3×10^{-4}
Cygnus A	$5{\cdot}7\times10^{44}$	$2{\cdot}8\times10^{59}$	4×10^{-5}	$3{\cdot}9\times10^{60}$	2×10^{-4}
Coma Cluster	$1{\cdot}0\times10^{41}$	$2{\cdot}9\times10^{59}$	2×10^{-7}	$4{\cdot}0\times10^{60}$	7×10^{-6}

* For these sources the equipartition condition has not been used.

Professor F. Hoyle has suggested that the energy arises from the rotational energy of the galaxy, the power coming from a galactic flare which occurs after the magnetic field in the central region has been wound up by rotational effects. The main difficulty with this kind of theory is that there is often not much rotation observed in elliptical systems, and very little gas, so that it is not easy to see what coupling mechanism is involved.

Shklovsky (1960) apparently believes that nuclear energy

must be responsible, and postulates that supernovae must occur at a sufficient rate to account for the large number of particles. My own ideas are based on the observation that elliptical galaxies are frequently radio sources and that they contain very little gas. Since I believe that the energy comes essentially from nuclear energy, it is necessary to release the energy over a very short time. It is not possible to build up these tremendous energies over a long time scale from stars. In these ideas I have tried to make use of the fact that in the centre of elliptical galaxies the star density may be extremely high. The suggestion is that in the centre of an elliptical galaxy, a supernova occurs, and this may be able to start a chain reaction of supernovae by triggering nuclear reactions in the atmospheres of nearby stars. The key to the whole question is whether the detonation wave thus started in a stellar atmosphere will propagate into the interior. It seems to me that it will, but it is an exceedingly difficult mathematical problem mainly because of the geometry involved. If the detonation wave does propagate, there will follow a whole series of nuclear chain reactions which will end only when the star density becomes too low. The central regions only would be affected, involving perhaps 10^6 or 10^7 stars.

By one or other of these mechanisms a very large flux of particles is produced in the galaxy. If you accept my process, this is a chance event. If you now calculate the minimum magnetic field in the radio sources, by equating the magnetic and particle energy densities, you find that the magnetic fields are of the order of 3×10^{-5} gauss. This value seems to be unreasonably large, and I have concluded that in fact the magnetic field must be considerably smaller than this. Then the total particle energy must be considerably greater than the lower limit set by assuming equipartition, in order to account for the observed radio emission. We now have a situation in which the particle energy is very much greater than the magnetic energy. Consequently the time scale for the radio sources produced in this way is no longer anything to do with the time scale for the decay of the electrons. You have a cloud of relativistic protons and electrons which simply blows through the magnetic field and the whole radio source is a very short-lived thing. The bulk of the energy in the particles escapes before any reasonable part is radiated as radio emission. The life time one associates with these phenomena is some fraction of the light travel time across

the object, that is to say of the order of 10^5 to 10^6 years. We now have a very short-lived radio source, and in order to account for the observed number of radio sources, one must have galaxies becoming sources very much more frequently than hitherto supposed. If one now considers the local supercluster, in which five of the strong identified sources lie, it is possible to show that with the frequency of radio sources that is needed, the contribution that these sources make to the high-energy particle density in the supercluster is very considerable, and it turns out that the energy density built up over a time of order 10^{10} years is comparable with the energy density of cosmic rays in our Galaxy. There are other arguments which show that the halo of our own Galaxy is open so that particles can escape from it and also come in to it. Consequently it seems that it is not impossible that the strong extragalactic radio sources may make a very considerable contribution to the primary cosmic radiation in the local region of space.

DISCUSSION

In answer to questions, Professor Burbidge made it clear that although at one time the interpretation of Cygnus A as a pair of colliding spiral galaxies seemed reasonable, there has never been any unambiguous observational evidence for this. This galaxy is too remote for classification, and there is at present no direct evidence that any stars are present. The statistics are such that collisions will occur, but there is no known mechanism whereby the collisional energy can be converted into radio energy. Furthermore in a dense cluster like the Coma cluster, for example, one would expect to see many such collisional radio sources, but in fact this does not seem to be the case.

REFERENCES

Burbridge, G. R., 1961, *Nature*, **190**, 1053.
Burbridge, G. R., and Burbridge, E. M., 1962, *Nature*, **194**, 367. Observations of the California Institute of Technology Radio Observatory.
Evans, D. S., and Harding, G. A., 1961, *M.N. Ast. Soc.*, South Africa, **20**, 64.
Hanbury Brown, R., and Hazard, C., 1961, *M.N.R.A.S.*, **123**, 279 and earlier papers.
Moffat, A. T., and Maltby, P., 1962, *An. J. Supp. Ser.*, **7**, 93.
Seyfert, C. K., 1943, *Ap. J.*, **97**, 28.
Shklovsky, I. S., 1960, *A.J.*, U.S.S.R., **37**, 945.

17

THE OPTICAL IDENTIFICATION OF RADIO SOURCES

D. W. DEWHIRST

SUPERPOSED on the smooth but patchy background of radiation that comes from the whole sky at metre wavelengths are a considerable number of discrete sources. They have sizes ranging mostly between a few minutes and a few seconds of arc: a few, found near the plane of the Galaxy, are as much as a degree across. Some of the latter have a thermal spectrum and are H II regions in the plane. We concern ourselves here with those radio sources that have a non-thermal spectrum, and survey the problem of identifying them with optical objects.

There are three reasons which make it important to identify as many radio sources as possible. The first we might call the astrophysical one: with the higher resolving power of optical telescopes and the powerful methods of optical spectroscopy we can learn more about the still imperfectly understood mechanisms of the radio emission. Moreover, the radio observations have directed our attention to some very curious objects whose significance had previously been overlooked, and which are of interest in themselves. The second reason is cosmological. The certain knowledge that many of the discrete sources are distant galaxies has led to the interpretation of source counts in terms of the large-scale structure of the Universe. It is obviously important to know that the sources being counted are really extragalactic. The third reason, more prosaic but not less important, is that a reliably identified galaxy gives us an accurate position. Absolute Right Ascensions and Declinations are determined by radio methods only with great difficulty. On the other hand the optical position of a galaxy can be measured with great accuracy: the identified galaxies provide standard positions scattered over the sky from which further radio positions can be measured by difference or interpolation.

HISTORY OF THE SEARCH FOR IDENTIFICATIONS

The first accurate fundamental determinations of the position of four of the most intense radio sources in the northern sky revealed an obviously puzzling situation: there are no really bright optical objects in the direction of the most intense radio sources. The basic difficulty remains the same today: the objects are optically faint and require really large telescopes for their study. The first four identifications were with two distinct types of objects. (1) Supernova remnants. The Crab nebula, in the position of Taurus A, was already well known to optical astronomers as the remnants of the supernova of A.D. 1054. The much fainter remnant in the position of Cassiopeia A, some 3400 pc away and consequently heavily obscured by dust in the galactic plane, required the Hale telescope for its certain identification. (2) Galaxies. M 87 (Virgo A) was one of the well-known bright elliptical galaxies in the Virgo cluster. The second, in the position of Cygnus A, was optically much fainter, much more distant, and previously unrecognized. Both galaxies showed structural or spectroscopic peculiarities.

For some years these identifications directed thinking about, and searches for, further identifications. It was thought that there were only these two types of objects (supernovae and galaxies), an idea well supported by the distribution of the radio sources in the sky. At first the idea that galaxies that were radio sources must necessarily show structural or spectroscopic peculiarities led to a random search for associations between 'freak' galaxies and the still unreliable radio positions, with resulting mistaken identifications. From about 1957 onwards it became clear that freak galaxies are not necessarily abnormal radio emitters, and equally that galaxies with abnormally strong radio emission may show neither obvious structural nor spectroscopic abnormality. It may also be remarked that the two principal condensations of light that constitute the Cygnus A galaxy were originally interpreted as two spiral galaxies in collision, the reasoning at that time being that (*a*) the optical spectrum shows intense emission lines which must arise in the interstellar gas, (*b*) only Population I objects (spirals) contain interstellar gas, (*c*) only a collisional process could provide the enormous energy output that is observed. Our more extensive knowledge of the nature of identified galaxies (e.g. Dewhirst,

1959; Bolton, 1960; Minkowski, 1961), of the gaseous content of elliptical galaxies (e.g. Osterbrock, 1960), and of the possible mechanisms of radio emission (e.g. Burbidge, 1961) now make this interpretation of radio source galaxies as collisional processes less likely, either for Cygnus A in particular or radio galaxies in general.

THE RADIO POSITIONAL DATA

Early experience showed the necessity of radio positions of the greatest possible accuracy if further identifications were to be made, but the need for surveys to indicate the number, flux, and approximate position of all the sources brighter than some given flux limit, and for the most accurate positions alone, were conflicting requirements. So far there have been only two major surveys of the whole sky published, the 3C (3rd Cambridge) survey covering roughly the northern sky between $+50°$ and $-10°$ Declination, with some brighter sources N and S of these approximate limits (Edge *et al.*, 1959), and the MSH survey extending from $+10°$ to $-20°$ (Mills, Slee and Hill, 1958) with more recent extensions to cover most of the southern sky (Mills, Slee and Hill, 1960, 1961). The 3C survey was made with an interferometer at 159 Mc/s and lists 471 sources: the MSH survey with a 'Mills cross' at 86 Mc/s lists 1159 sources in the first 30°-wide Declination zone. In both surveys the 'probable error area' of each source (i.e. the area of the rectangle defined by the quoted errors in the two coordinates) depends principally on the brightness of the source, increasing as the sources become fainter. It is of the order of 20 sq. min. of arc for an average 3C source and rather larger for MSH sources of the same brightness.

More recent work has concentrated on getting more detailed information—better positions, diameters and intensity distributions, spectra, etc.—for the brighter sources discovered in the two surveys. Because of the location of the instruments the sources studied in detail have been mostly in the northern sky. The most accurate positions published have been of 64 3C sources measured by Elsmore, Ryle and Leslie (1959). Studies of intensity distributions have been made by the California Institute of Technology group at Owens Valley with a twin dish interferometer at 960 Mc/s (Maltby and Moffet, 1962) and by Allen *et al.* (1962) at Jodrell Bank. Further accurate positions,

determined by the C.I.T. group, and at Cambridge, Nançay, Pulkova and elsewhere, remain largely unpublished at present with work still in progress.

The probable error area of the 64 ERL sources averages about 2 sq. min. of arc. In principle the achievement of still greater accuracies is possible, though very difficult in practice: the C.I.T. observers have reported error areas of only a fraction of a square minute of arc. As we shall see, the greater the positional accuracy the fainter the optical objects that may safely be identified. It now seems, however, that some limit to this might be in sight, since many radio sources have been found to exhibit complex structure and may have an asymmetrical intensity distribution: the linear separation of the radio components may be several times greater than the size of the associated optical object. There is no guarantee that the 'centres of gravity' of the radio and optical isophotes coincide.

Later in this paper, and in discussing the cosmological significance of source counts in Chapter 8, we shall meet problems of sampling. It cannot be too strongly stressed that the recognition of effects of observational selection in a published list of radio sources is a most difficult matter. A short list of sources selected for detailed study already contains any defects in the original surveys. In these surveys, some parts of the sky are less well studied than others, because of solar interference, confusion of the records by side lobes of intense sources, and the difficulties of analysis in complicated regions both in the galactic plane and in high latitude where two or three sources are close together. The selection of a short list for detailed study may be affected by such practical considerations as the difficulty of tipping aerials for sources of similar Right Ascensions, whilst it would be misleading, for example, to attach physical significance to the percentage of radio sources identified with galaxies in the ERL list of 64 positions, since some sources suspected to be galaxies were purposely included in the observing programme.

IDENTIFICATIONS WITH SUPERNOVAE IN THE GALAXY

So far 13 sources have been identified with optical objects that are probably the remnants of supernovae. Much of our knowledge is surveyed in a thesis by Harris (1961). In addition there are a few radio sources in the plane whose intensity distribution

and size suggest supernovae remnants that are optically invisible because of dust obscuration.

The optical objects are usually more or less circular arcs of emission nebulosity, often with a characteristic filamentary structure. Sufficient quantitative information is available for five of them to permit an interpretation of their physical nature. Thus the angular expansion θ radians/sec can be determined from direct photographs spaced by several years, whilst Doppler displacements in the spectra give line of sight components of the velocity V in km/sec. Reasonable assumptions about the symmetry give the distance $D = V/\theta$. The general picture is of an expanding mass of gas, from the original catastrophe, decelerated by the stationary interstellar medium. Equating the momentum of the original material with the observed momenta of the shells (which are largely swept-up interstellar matter) leads to original ejecta of the order of one solar mass and ages and diameters for the old remnants, like the Cygnus Loops, of around 50,000 years and 40 pc. Although the general picture is satisfactory there are many unsolved problems. The observed remnants are of great diversity: Zwicky has recently cast further doubt on the generally accepted concept that there are two types of supernovae associated with the two extreme stellar populations. It is probably an over-simplification. The Crab nebula, usually regarded as a typical remnant of 'Type I', is anomalous in many respects. In other galaxies we see supernovae of different types in both spiral and elliptical galaxies, sometimes apparently at large distances from the equatorial plane. Why are all the supernovae remnants in our own Galaxy concentrated in the plane? Does a supernova remnant far from the plane have a shorter radio and optical life because the expanding gas is less efficiently contained?

IDENTIFICATIONS WITH GALAXIES

It was fortunate that the completion of the first surveys of radio sources coincided with the completion of the National Geographic Society—Palomar *Sky Survey*. The plates, taken with the 48-inch Schmidt and each covering a square area of sky of side 6°·6, cover the sky from the North pole down to Declination −33° on a scale of 1 mm = 67″·1. The faintest objects recorded have an m_{pg} of about 21^m. The uncertain earlier radio

positions could conveniently be examined on the plates and many tentative identifications were made in this way. Accurate modern positions enable the limited field of a large reflector to be used directly, but the *Sky Survey* is still valuable for a first inspection and statistical studies.

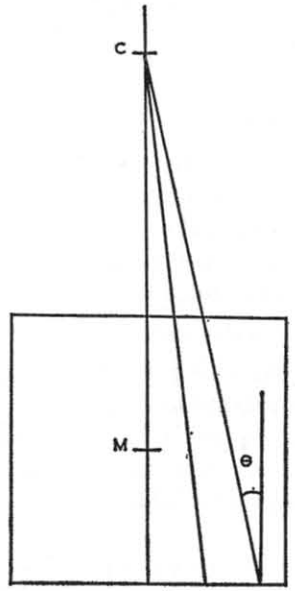


FIG. 17/1.—Great circles of constant Right Ascension on a classical Schmidt plate.

Identifications based on the coincidence of radio sources showing, for example, structural or spectroscopic peculiarities, may be correct but contain an assumption about the nature of galaxies that are radio sources: it is preferable to make identifications on a positional basis alone, disregarding the type of galaxy involved.

For a uniform spatial distribution of galaxies, the number $N(\mathrm{m})$ of galaxies brighter than a limiting magnitude m varies directly as the volume of space V penetrated to distance D:

$$\log N(\mathrm{m}) + \text{constant} = \log V = 3 \log D$$
$$= 3 \log \left(\frac{1}{\text{luminosity}}\right)^{\frac{1}{2}}$$

From the definition of optical magnitude

$$\left[\frac{I_1}{I_2} = 10^{-0\cdot4(m_1-m_2)}\right]$$

it follows that this last quantity

$$= \frac{3}{2}(0\cdot4)\ \mathrm{m}$$

$$= 0\cdot6\ \mathrm{m}$$

This is the optical analogue of the familiar $\frac{3}{2}$ power law frequently referred to in discussions of the source count data for radio sources. The only comparable optical counts date from the 1930's (Hubble, 1936): the best value of the constant fitting Hubble's counts to modern magnitude scales is due to Sandage (quoted by Minkowski, 1961) and gives the relation

$$\log N(\mathrm{m}) = 0\cdot6\ \mathrm{m_{pg}} - 8\cdot9$$

for the number of galaxies per square degree brighter than a given apparent photographic magnitude. We use this to calculate the area (in square minutes of arc) in high galactic latitude in which on average we shall always find one galaxy of a given $\mathrm{m_{pg}}$ or brighter (Table 17/I).

TABLE 17/I

Area, in square minutes of arc, containing one object brighter than limiting apparent magnitude $\mathrm{m_{pg}}$

$\mathrm{m_{pg}}$	Area (sq. min. of arc) containing one object		
	Galaxies	Stars	
		$b = 0°$	$b = 90°$
14·0		6·9	70
15		2·7	39
16		1·4	21
17	181	0·53	11
18	45·4	0·23	7·0
19	11·4		4·5
20	2·8		2·9
21	0·72		
22·0	0·18		

Thus in a probable error area of 2·8 sq. min. (a typical ERL source position) we shall always find one galaxy brighter than 20·0. For each magnitude step brighter, the probability of finding an object of that brighter magnitude in the same area of 2·8 sq. min. is decreased by a factor of very nearly 4 (antilog 0·6), and if we limit ourselves to making identifications with galaxies brighter than 17·0 we shall make only about one mistake in 64 identifications arising from chance coincidences.

Such a criterion cannot be very precise, because of the uncertainty of the counts and the effects of clustering of galaxies. (The present state of galaxy counts as a function of magnitude is very unsatisfactory, but it is easier to criticize Hubble's work than to embark on the labour of improving it!) However, some restriction of this kind must be imposed if entirely erroneous identifications are to be avoided.

Taking all the published source data, and applying a rather cautious criterion of the kind described above, we may now regard at least 40 radio sources as fairly certainly identified with galaxies that are peculiar in having an abnormally strong radio emission (i.e. omitting normal spiral galaxies like M31).

IDENTIFICATIONS WITH STARS

Until quite recently no truly stellar object had been identified as a radio source. There is evidence, however (Allen *et al.*, 1962) that some sources have very small angular diameters, even less than 1″; no galaxy has been found in these places and there are reasons for believing that such small diameters are unlikely to be associated with galaxies. It now seems certain that at least one of these sources (3C 48) is a stellar object in the Galaxy (Sandage *et al.*, 1960). The object, of $m_v = 16$, has a small area of excessively faint nebulosity adjacent to it. The spectrum (of the stellar object) shows broad emission lines of uncertain origin. Smith and Hoffleit (1961), whose description of the object is however rather misleading, have proposed that it may be the stellar remnant of a supernova.

The difficulty of making reliable identifications with stellar objects on a purely positional basis is much greater than for galaxies because of the greater number of stars and their less favourable number-magnitude distribution. Sample figures (again based on rather antiquated data) for two extreme

galactic latitudes are given alongside the figures for galaxies in Table 17/I. Even in high latitude an error area of 3 sq. min. will lead to about one chance coincidence with a 15th magnitude star in every 13 source positions.

PROPORTION OF IDENTIFIED SOURCES

It is difficult to find any homogeneous sample of radio sources, free from selection effects, for which we have at the same time positions of sufficient accuracy to lead to any reasonable expectation of making identifications. Perhaps the most favourable group that has been studied in any detail is the 70 most intense sources of the 3C catalogue (sources with flux density $\geqslant$ 25 units), which may be thought reasonably complete for a large area of the northern sky. Taking all the available information (much of it unpublished) we obtain the following result for the present state of the identification problem:

TABLE 17/II

Identifications: Analysis of the 70 most intense sources of 3C

(*a*) Sources certainly in the Galaxy, or unidentified in low latitude.

Supernova remnants	6	
H II regions	2	
Stars	1	
Unidentified in low latitude	14	Total 23

The large proportion of unidentified sources here arises from the difficulties of making identifications in crowded star fields and where obscuration effectively hides extragalactic objects.

(*b*) Sources certainly extragalactic, or unidentified in high latitude.

Certain identifications with galaxies	23	
Probable identifications with galaxies	7	
Unidentified in high latitude	17	Total 47

The identified galaxies here do not include any 'normal' spirals like M31. 'Probable identifications' include situations where the source lies in a cluster of galaxies, but the position is not good enough to identify the individual galaxy responsible.

It would seem that we can identify some 60 per cent of the intense sources in high galactic latitude: this percentage is likely to increase as more precise positions become available for some of the as yet unidentified sources. Indeed for sources brighter than 50 flux units in the 3C catalogue the identification list is almost complete, whilst adding the next 30 brightest

sources to the above 70 (100 sources $\geqslant$ 20 units) adds only 4 identifications—a result largely of the poorer positional information available for the weaker sources.

It seems significant that radio sources tend to be associated with clusters of galaxies, but this may mean no more than that since galaxies cluster, and some galaxies are radio sources, radio sources will also be associated with clusters of galaxies. This interpretation appears to be supported by van den Bergh's study (1962). It also seems significant that of the 23 certainly identified galaxies in the above analysis, identified on a basis of position and magnitude alone without appeal to their structural peculiarity as an additional criterion of reliability, 10 are on examination found to be close pairs or multiple objects—an appreciably higher proportion than for galaxies selected at random.

Postscript added in proof (*March 1963*)

The source 3C 48 has been reidentified as an extremely bright extragalactic object, with an optical red-shift $\frac{\partial\lambda}{\lambda_0} = 0{\cdot}36$. See Greenstein and Matthews, *Nature*, **197**, 1041, 1963.

REFERENCES

Allen, L. R., Anderson, B., Conway, R. G., Palmer, H. P., Reddish, V. C., and Rowson, B., 1962, *M.N.*, **124**, 477.

Bolton, J. G., 1960, 13th General Assembly of U.R.S.I., London, *Observations of the California Institute of Technology Radio Observatory*, No. 5.

Burbidge, G. R., 1961, *Nature*, **190**, 1053.

Dewhirst, D. W., 1959, *Paris Symposium on Radio Astronomy* (*I.A.U. Symposium No. 9*), ed. R. N. Bracewell, Stanford Univ. Press.

Edge, D. O., Shakeshaft, J. R., McAdam, W. B., Baldwin, J. E., and Archer, S., 1959, *Mem. R.A.S.*, **68**, 37.

Elsmore, B., Ryle, M., and Leslie, P. R. R., 1959, *Mem. R.A.S.*, **68**, 61.

Harris, D. E., 1962, *Ap. J.*, **135**, 661.

Hubble, E., 1936, *Ap. J.*, **84**, 517.

Merton, G., 1953, *M.N.*, **113**, 428.

Moffet, A. T., and Maltby, P., 1962, *Ap. J., Supp. Ser.*, **7**, 93.

Mills, B. Y., Slee, O. B., and Hill, E. R., 1958, *Aust. J. Phys.*, **11**, 360.

Mills, B. Y., Slee, O. B., and Hill, E. R., 1960, *Aust. J. Phys.*, **13**, 676.

Mills, B. Y., Slee, O. B., and Hill, E. R., 1961, *Aust. J. Phys.*, **14**, 497.

Minkowski, R., 1961, *Proc. 4th Berkeley Symposium on Mathematical Statistics and Probability*, ed. J. Neyman, Univ. of California Press, Berkeley.

Osterbrock, D. E., 1960, *Ap. J.*, **132**, 325.

Sandage, A. R., *et al.*, 1961 paper at 107th Meeting of Am. Astron. Soc., New York, see *Sky and Telescope*, **21**, 148.

Smith, H. J., and Hoffleit, D., 1961, *P.A.S.P.*, **73**, 292.

Van den Bergh, S., 1961, *A.J.*, **66**, 562.

APPENDIX

Notes on the use of the Palomar—N.G.S. Sky Survey

1. *Introduction*

The original plates of the Sky Survey cover all the sky north of Declination −33°·3. The scale of the contact prints is 1 mm = 1′·118; the square fields centred at 6° Declination intervals (+90°, +84°, +78° . . .) have a side of 6°·6. The identification and measurement of positions on the prints of the Sky Survey, which bear no co-ordinates other than those of the plate centre, are often a source of difficulty to users at first. The following is a brief description of one method that has been found useful by the writer. It was used in a practical class in the course of the Summer School.

2. *Maps and Star Catalogues*

Auxiliary maps and star catalogues used to locate and measure the positions of objects are:

(1) The *B.D.* (*Bonner Durchmusterung*) charts. Atlas of stars to Declination −23°, showing about 460,000 stars.

(2) The *B.D.* catalogues. *Bonn Astron. Beobachtung.* Vols. 3–5, 8. Approximate positions and catalogue numbers (e.g. *BD* +27° 1379) of these stars. Both charts and catalogues are epoch 1855·0, and the Sky Survey plate centres are for this reason also 1855·0.

(The *C.D.—Cordoba Durchmusterung*—is similar for the southern sky.)

(3) *AGK_2—Zweiter Katalog der Astronomischen Gesellschaft.* Hamburg—Bonn. Precise 1950·0 places for stars north of −2°, with *B.D.* numbers.

(4) Yale Zones. *Trans. Yale Univ. Obs.* Precise 1950·0 places for stars between +30° and −30°, with *B.D.* numbers and proper motions. Some other zones.

3. *Procedures*

There are two practical problems.

(i) To identify an area of given R.A. and Dec. (1950·0, usually) for preliminary examination:

Apply precession corrections to convert (1950·0) to (1855·0) place, with moderate accuracy—about ±1′—only. Plot 1855·0 place and relevant plate centre on *B.D.* charts, identify area by visual inspection with respect to nearby stars.

(ii) To measure an accurate 1950·0 place for any selected optical object:

Identify the 4 or 5 nearest stars by their *B.D.* catalogue numbers.

AGK_2 or Yale Zones then provide accurate 1950·0 places for most of these stars. The distance of the object from each of these stars can be measured in millimetres with an overlaid transparent graticule. The distances can be converted directly into differences of R.A. (in time) and Dec. (in arc) if the axes of the millimetre grid can be adjusted to be parallel to the lines of constant R.A. and Dec. in the immediate vicinity of the area being measured.

In a classical Schmidt camera great circles of constant R.A. project very nearly as straight lines: the central meridian is perpendicular to the base of the plate (Fig. 17/1). Great circles of constant R.A. on a plate centred at Dec. δ_z intersect the central meridian at C, where

$$CM = f.\ \cot \delta_z \quad (f = \text{focal length})$$

If the object being measured is α° of R.A. from the central meridian CM, then that line of constant R.A. makes an angle θ with the side of the plate, where

$$\tan \theta = \tan \alpha \sin \delta_z$$

By tilting the millimetre grid by the angle θ the object can be measured as (e.g.) 'x mm N. of, and y mm preceding' a nearby star. If the millimetre grid is large enough to overlap one side of the plate it is convenient to know the value of tan θ, as practice will

TABLE 17/III

Tabular values of tan θ for 48-inch Schmidt plates (*see Fig. 17/1*)

R.A. from C.M. / δ zone	5m	10m	15m	20m	25m	30m	35m	40m	
±0°	0·000	0·000	0·000						
6	·002	·005	·007						
12	·005	·009	·014						
18	·007	·013	·020						
24	·009	·018	·027	·036					
30	·011	·022	·033	·044					
36	·013	·026	·038	·051					
42	·015	·029	·044	·058					
48	·016	·032	·049	·065	·081				
54	·018	·035	·053	·071	·090				
60	·019	·038	·057	·076	·095	·114			
66	·020	·040	·060	·080	·100	·120	·140	·161	
	10m	20m	30m	40m	50m	60m	70m	80m	90m
72	·042	·083	·125	·168	·211	·255			
78	·043	·086	·129	·172	·217	·262	·308	·356	·405
84	·043	·087	·131	·175	·220	·266	·313	·362	·412

immediately show. Two tables give tan θ as a function of R.A. (in minutes of time) for the various declination zones of the Sky Survey (Table 17/III), and conversion factors from millimetres to minutes of arc and seconds of time (Table 17/IV).

TABLE 17/IV

Conversion factors for 48-inch Schmidt plates

$\delta°$	1 mm =	10^s of R.A. =	$\delta°$	1 mm =	10^s of R.A. =	$\delta°$	1 mm =	10^s of R.A. =
	s	mm		s	mm		s	mm
0	4·472	2·236	30	5·164	1·936	50	6·957	1·437
2	4·475	2·235	31	5·217	1·917	51	7·106	1·407
4	4·483	2·230	32	5·274	1·896	52	7·264	1·377
6	4·497	2·224	33	5·332	1·875	53	7·432	1·346
8	4·515	2·215	34	5·394	1·854	54	7·609	1·314
10	4·540	2·202	35	5·459	1·832	55	7·797	1·283
12	4·572	2·187	36	5·527	1·809	56	7·996	1·251
14	4·609	2·170	37	5·600	1·786	57	8·211	1·218
16	4·653	2·149	38	5·675	1·762	58	8·439	1·185
18	4·702	2·127	39	5·754	1·738	59	8·686	1·152
20	4·759	2·101	40	5·837	1·713	60	8·943	1·118
22	4·823	2·073	41	5·925	1·688	61	9·224	1·084
24	4·894	2·043	42	6·017	1·662	62	9·526	1·050
26	4·975	2·010	43	6·115	1·635	63	9·851	1·015
28	5·065	1·974	44	6·217	1·608	64	10·202	0·980
30	5·164	1·936	45	6·324	1·581	65	10·583	0·945
			46	6·437	1·553	66	10·995	0·909
			47	6·557	1·525	67	11·445	0·874
			48	6·683	1·496	68	11·937	0·838
			49	6·817	1·467	69	12·480	0·801
			50	6·957	1·437	70	13·074	0·765

1 mm = 1′·118 1′ = 0·895 mm

The above method is quick, and the error will be greater than 6″ only when there is an unusually unfavourable field star. Greater accuracy may be attained by more tedious procedures (e.g. Merton, 1953). There is a limit to the accuracy attainable on the prints set by the distortion of the paper base. Further limitations arise from the uncertain proper motions of the reference stars (reject discordant places) and from the complicated geometry of the Schmidt camera optics (elastic plate distortion).

An alternative procedure is to use a digital computer to calculate the rectilinear coordinates of the several (say 10) nearest AGK_2 stars with respect to the plate sides and to the nominal radio source position, using the standard formulae of photographic astrometry. Each of the errors mentioned above will be reduced in the resulting 'best fit' between the calculated positions and their respective images on the plate. This method is used at Jodrell Bank.

18

RADIO SOURCE OBSERVATIONS AND THE APPLICATION OF THE RESULTS IN COSMOLOGY

M. RYLE

THE survey of radio sources carried out in Cambridge between 1953 and 1957 which is known as 3C (Edge *et al.*, 1959) indicated that the radio sources of small angular diameter were isotropically distributed in the sky, and that counts of the sources lying in various ranges of flux density did not seem compatible with an isotropic distribution in depth. It seemed probable that many of the radio objects were very powerful sources at great distances, and that the non-uniformity was connected with the red shift. If this were the explanation, then an extension of the observations to fainter sources should reveal more marked effects. At the same time, it was important to examine various possible selection effects in more detail than was possible with the 3C instrument. For these reasons we decided in 1956 to construct a new instrument with sufficient sensitivity and resolution to extend the observations to considerably weaker sources and also to investigate the angular structure and the angular distribution of the sources with greater precision than was possible in 3C. With this new instrument we have been able to make a more detailed analysis of the luminosity function of the sources, and also of the number/flux density relationship.

In establishing this relationship it is necessary to carry out a number of separate observing programmes. The difficulty is that you must carry out a fairly rapid survey of the whole sky to get enough statistical accuracy for the more intense sources (that is to say that for these sources 4π steradians is not really enough!) and the angular structure of the more intense sources must also be determined so that due regard can be paid to the probable angular structure of the weaker sources in interpreting the results. In the main survey, which takes much longer, the full resolution and sensitivity is realized, but whole sky coverage is not necessary to obtain results of comparable statistical accuracy.

TOTAL POWER AND INTERFEROMETRIC SURVEY OF THE MORE INTENSE SOURCES

In the first series of observations (Leslie, 1961*a*) the long element of the instrument (dimensions 1450 ft × 65 ft oriented E–W) was used as a total power instrument, which had a beam of 13·5 min. arc by 4·5 degree. In the second series, the same element was used in conjunction with a smaller element at a spacing of 469λ to the East, giving effectively the same polar diagram but with an 8′ arc interference fringe system inside the polar diagram. This equipment was used to observe 900 sources having flux densities greater than 6×10^{-26} w m^{-2} (c/s)$^{-1}$, of which about 600 lay more than 20° from the galactic plane.

This set of observations confirmed the 3C result that very few radio sources are appreciably resolved by an interferometer of this spacing (469λ), and indicated that 82 per cent of the sources had surface brightness temperatures $T_s \geqslant 6 \times 10^4$ °K and that 16 per cent lay in the range 10^4 °K to 6×10^4 °K. Only about 2 per cent were completely resolved, having brightness temperatures of about 200°K, and these are identified with nearby normal galaxies.

If these statistics of the structure of the brighter sources are taken to be typical of that of the weaker (more remote) sources, we can use the data to correct the number/intensity counts of the full survey for the effects of partial resolution. It turns out that this correction is extremely small.

THE FULL SURVEY BY APERTURE SYNTHESIS

The main survey was made by combining the long east–west element with the smaller element and making 24-hour observations with it in each of 24 successive positions on its north–south railway track. In this manner an instrument was obtained with an envelope resolution of 25′ × 35′, again with an 8′-fringe pattern (Scott, Ryle and Hewish, 1961). The areas of sky which we have covered to date are declinations:

02°–07°
17°–34°
40°–44°
48°–54°

In each case virtually the whole 24 hrs of right ascension was

observed, apart from ¾ hr which was needed for changing the aerial position. The intermediate regions are now being filled in.

In the analysis we generally read off sources which have flux densities greater than 2×10^{-26} w m^{-2} (c/s)$^{-1}$ at 178 Mc/s. At this intensity there is about one source to every thirty beam areas so that the effects of confusion are negligible, and the signal to noise ratio is better than 25 : 1. The computations for the synthesis were carried out in the Cambridge University computer, EDSAC II, and the results are in the form of printed

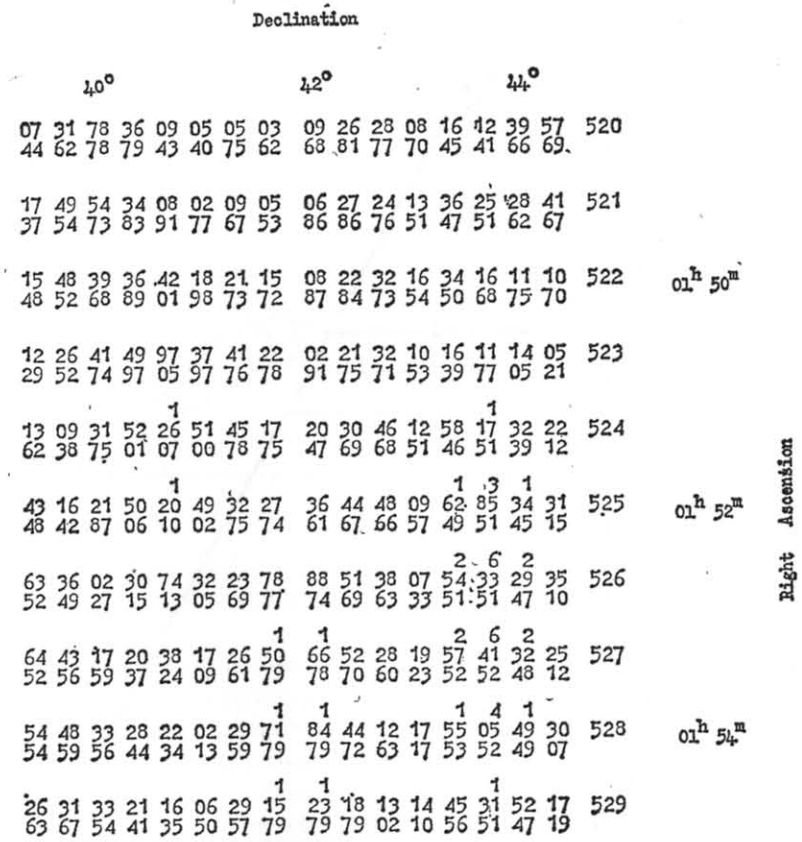


FIG. 18/1.—Sample of printed output from the computer, showing amplitude (upper row) and phase (lower row) of the record at a series of grid points. The vertical columns represent different declinations within the 4½° primary pattern, and right ascension runs vertically. Numbers greater than 100 are printed on the uppermost lines. The scale of flux density is 71 units ≡ 10^{-26} w m^{-2} (c/s)$^{-1}$.

sheets showing the fringe amplitude and phase at a set of positions in the sky, the printing being arranged so that sources stronger than 2×10^{-26} w m^{-2} (c/s)$^{-1}$ are easily picked out. An example of the print-out is shown in Fig. 18/1. A subsidiary computer programme is used to determine the coordinates (α, δ 1950·0 and α, δ 1855·0 and l^{II}, b^{II}) of the sources and their normalized flux density.

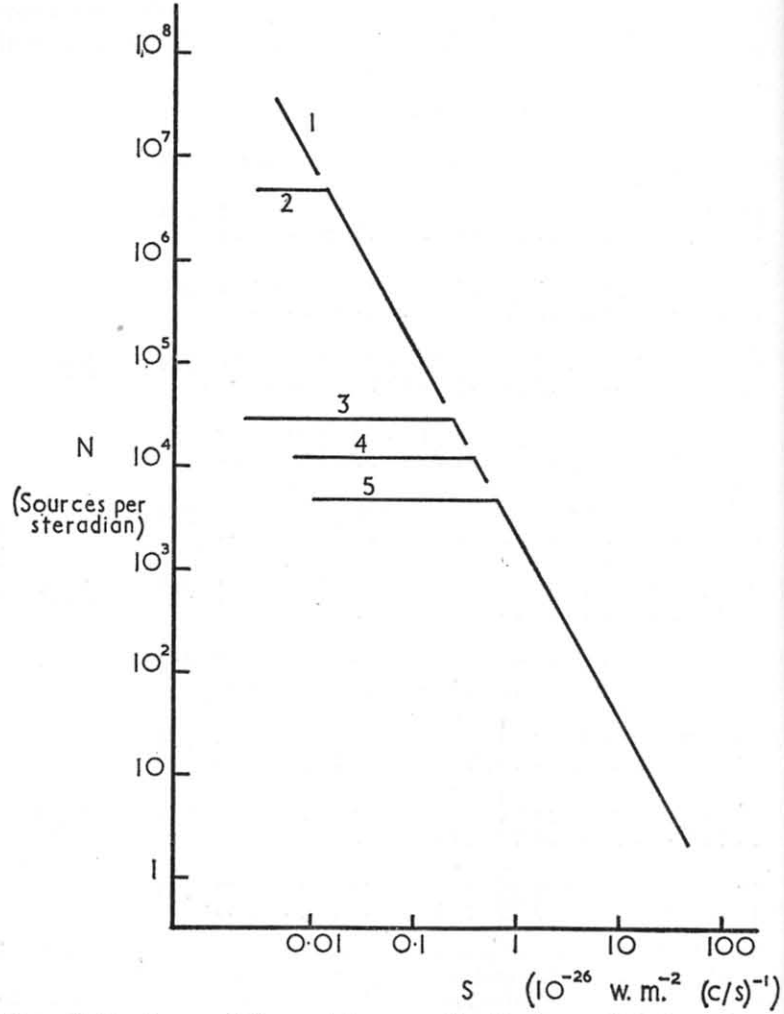


FIG. 18/2.—Some of the model source distributions which have been investigated by Hewish; the corresponding $P(D)$ curves are shown in Fig. 18/3.

Now the large signal to noise ratio at 2×10^{-26} w m^{-2} (c/s)$^{-1}$ suggests that it should be possible to apply the statistical method developed by Scheuer (1957) to extend the flux density range over which we can get information on the numbers of sources. By this means the number/flux density statistics can be extended to well beyond the point at which the measurement of individual sources becomes unreliable.

Typical $P(D)$ diagrams are shown in Fig. 18/3. The ordinate is proportional to the relative frequency with which a deflection on the chart, D, lies in the range D to $D + dD$. D has the same

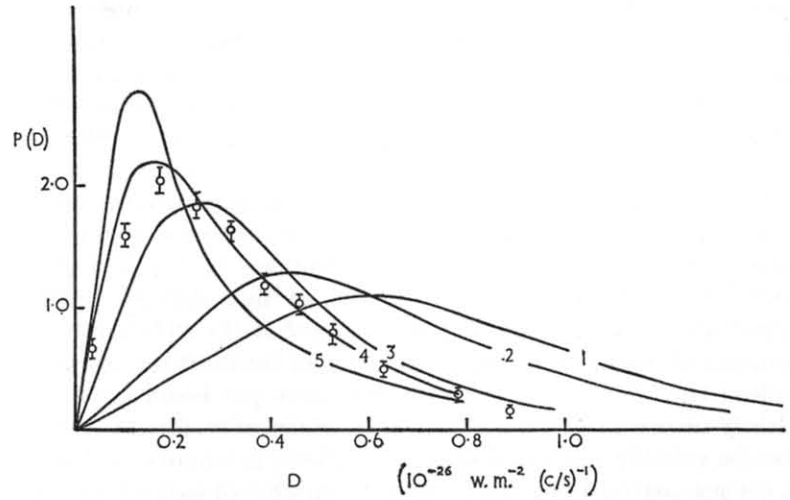


FIG. 18/3.—The $P(D)$ distributions derived for the model log N/log S relationships shown in Fig. 18/2. The observed points are also shown.

dimensions as flux density, but is given a different symbol because we are discussing a signal which arises from the vector addition of a number of small fringe patterns from sources which are not necessarily situated in the maximum of the primary beam. The $P(D)$ function depends on the distribution of sources in the sky and on the reception pattern of the instrument. In the case of the present survey the median value of D in the experimentally determined $P(D)$ function is about $0{\cdot}2 \times 10^{-26}$ w m^{-2} (c/s)$^{-1}$. Thus all sources measured from this survey are subject to an uncertainty of this order of magnitude in flux density and a corresponding phase (i.e. position) error.

If we are to make use of the actual shape of the observed $P(D)$ curve to extend our knowledge of the number/flux density relationship, we must find out how sensitive the form of the $P(D)$ distribution is to changes in the number/flux density distribution of the sources beyond the point where individual sources can be measured. Hewish (1962) has recently been making an investigation both to determine how sensitive the $P(D)$ function is to various changes in the source distribution and also to find what range of models is compatible with the observed $P(D)$ curve. To carry out the calculations analytically is extremely tedious but they can be done quite rapidly in a computer using a Monte Carlo type method. Hewish considered a number of possible extrapolations of the observed number/flux density distribution, and for each calculated the expected $P(D)$ curve. This was done, in the Monte Carlo method, by making the computer generate sources randomly in accordance with the desired number/flux density relationship and then, still in the computer, scanning these with the instrumental polar diagram to determine the corresponding $P(D)$ curve. Fig. 18/3 shows $P(D)$ curves computed for a number of log N/log S distributions having cut-offs at various flux density limits (illustrated in Fig. 18/2). It will be seen that the $P(D)$ curve is remarkably sensitive to the distribution of the faint sources and indeed the effects of sources at 50 sources per beam area are clearly discernible in the $P(D)$ curve. Since individual sources can be reliably measured only when there are about 25 *beam areas per source*, the $P(D)$ method is capable of extending our knowledge of the log N/log S curve to include information on a thousand times as many sources as can be individually measured, provided that the signal to noise ratio is adequate. Hewish worked with a large number of log N/log S models and determined which were compatible with the observed $P(D)$ distribution in the present survey (shown by the points in Fig. 18/3). By this means he was able to extrapolate the observed log N/log S curve to flux density levels of the order of $0{\cdot}1 \times 10^{-26}$ w m^{-2} (c/s)$^{-1}$.

THE NUMBER/FLUX DENSITY RELATION

The examination of individual sources was carried out to a flux density limit of 2×10^{-26} w m^{-2} (c/s)$^{-1}$. In using these to form a log N/log S distribution, account has to be taken of the partial resolution of a small fraction of these sources by the aerial beam. As already discussed, the necessary statistics on source structure were obtained from measurements of the brightest 600 sources, and the corrections were only a few per cent. Another important effect which might affect the number/flux density count would be any tendency for the sources to occur in clusters. If clustering of radio sources occurred, objects in nearby clusters would appear as separate sources, whereas clusters at a distance might appear as a single source, and an error would then occur in the counts. In order to determine the importance of such an effect Miss Leslie (1961*b*) made a comprehensive examination of the evidence for a non-random distribution of the sources. Her analysis covered associations of sources having angular separations of between 3·5 and 200 min. of arc. This range was sufficient to include all selection effects which might have been important in the survey. She concluded that there was no clustering beyond what would have have been expected for a random distribution of sources. Of course this may conceal a weak tendency to clustering, to which an upper limit was put by Miss Leslie; the correction which must be applied to the log N/log S curve if this upper limit to clustering is taken is not important, and we can conclude that clustering is not significantly affecting the results.

Errors can arise also if strong sources in the instrumental side-lobes are accidentally counted as weak sources in the main beam. Since the response more than about 1° from the principal planes containing the main beam is always less than 10^{-4} and in most places 10^{-5} of the main beam, corrections for this effect in the survey are very small.

Fig. 18/4 shows the number/flux density distribution of the sources stronger than 2×10^{-26} w m^{-2} (c/s)$^{-1}$, including the small corrections we have discussed (Scott and Ryle, 1961). The ordinate shows the number of sources per steradian having a flux density greater than the value plotted on the abscissa. The open circles refer to the total power survey and the crosses represent the main survey covering about three steradians. The limits of error include both the statistical errors and the uncertainties introduced by angular diameter effects, etc. The extension of this distribution by the $P(D)$ method to weaker flux densities is also shown in Fig. 18/4 which gives three possible

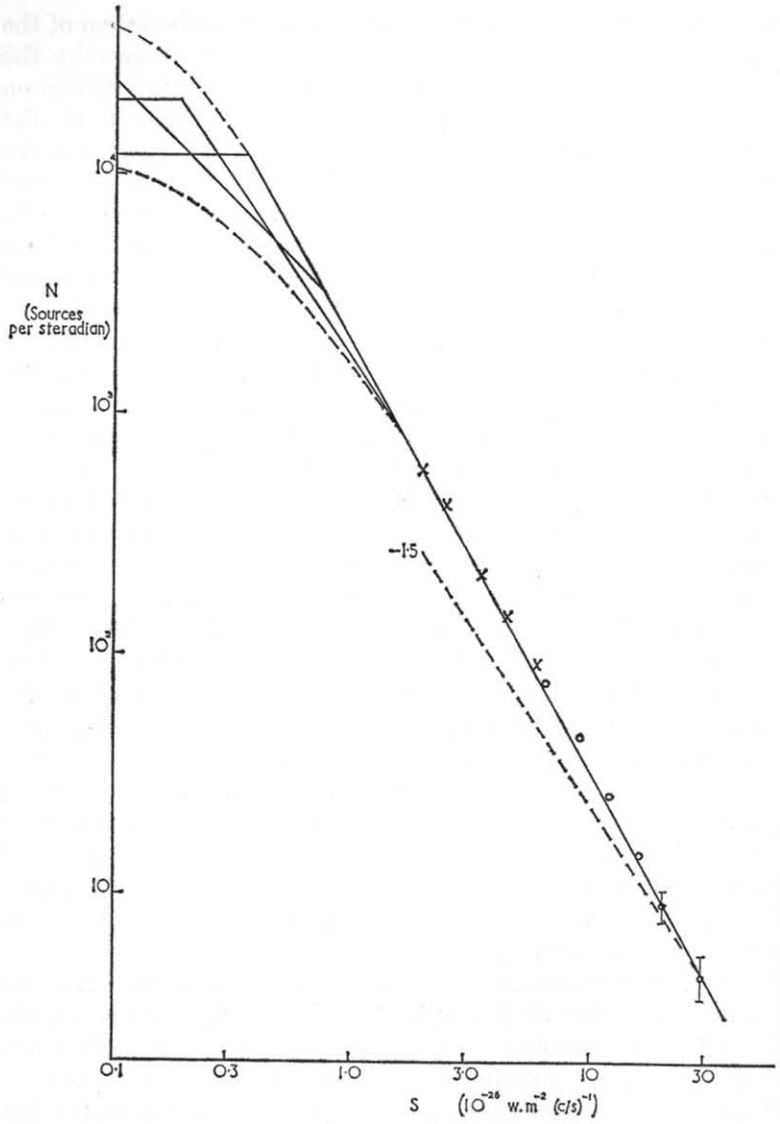


FIG. 18/4.—The log N/log S relationship derived from the counts of individual sources, together with the extrapolation to weaker sources using the statistical method of analysis. The circles represent the 'whole sky' observations, and the crosses the smaller areas which have been observed with the full resolving power.

extensions which Hewish found to give a $P(D)$ curve compatible with that observed. Many extensions of the log N/log S curve did not give permissible $P(D)$ curves, and the conclusion is that the extrapolation of the log N/log S curve must lie within the limits indicated in the figure by dotted lines.

THE RADIO LUMINOSITY FUNCTION

We shall discuss the distribution of radio sources in absolute luminosity in terms of the function $n(P)dP$, which represents the number of sources in the range of power P to $P + dP$ in a sample selected according to flux density. P is the power radiated by the source in watts per steradian per c/s at 178 Mc/s. This quantity is clearly of considerable importance in interpreting the number/flux density counts.

If it were possible to obtain from optical measurements the distances of all sources down to a given limiting flux density then we could derive the luminosity function $n(P)$. Unfortunately the number of sources included in a sample limited in flux density by the requirement that all should be observable optically is too small to allow any statistically reliable derivation of $n(P)$. As soon as you include sources of smaller flux density, where the numbers are sufficient, those which can be identified only represent a small fraction of the total, so that it is again impossible to obtain a definite curve of $n(P)$.

As a compromise we can examine sources down to an intermediate limit of flux density of 25×10^{-26} w m^{-2} (c/s)$^{-1}$ where nearly two-thirds of the sources can be identified with reasonable certainty; these should at least provide some idea of the true luminosity function. Using the data compiled by Dr. Dewhirst and by the workers at the California Institute of Technology,* estimates of distance and hence P are available for some 24 of the 40 sources having $S \geqslant 25 \times 10^{-26}$ w m^{-2} (c/s)$^{-1}$. The results plotted as a histogram of $n(P)$ are shown in Fig. 18/5.

It is evident that severe selection effects are present in this derivation of the luminosity function. A source having an optical luminosity comparable with the other identified sources, and a value of $P > 10^{26}$ w ster^{-1} (c/s)$^{-1}$, would be fainter than

* I am indebted to Dr. D. W. Dewhirst and Dr. G. J. Stanley for providing details of these sources prior to publication.

$m_{pg} = 19$ when at a distance such that $S = 25 \times 10^{-26}$ w m^{-2} (c/s)$^{-1}$. Since this optical magnitude represents about the limit of search, most of the sources having $P > 10^{26}$ w ster^{-1} (c/s)$^{-1}$ are unlikely to be identified.

The fact that the three sources with $P > 5 \times 10^{26}$ w ster^{-1} (c/s)$^{-1}$ produce flux densities considerably greater than the limit of $S = 25 \times 10^{-26}$ w m^{-2}(c/s)$^{-1}$ (8100, 325 and 73 $\times 10^{-26}$ w m^{-2} (c/s)$^{-1}$), indicates that the larger volume of space within which similar sources would not be identified optically might well contain 10 or more such sources having $S \geqslant 25 \times 10^{-26}$ w m^{-2} (c/s)$^{-1}$—a figure comparable with the remaining 16 sources not identified.

It is therefore probable that the true median value of P lies appreciably above the median value ($\sim 10^{25}$ w ster^{-1} (c/s)$^{-1}$) derived for the 24 sources shown in Fig. 18/5.

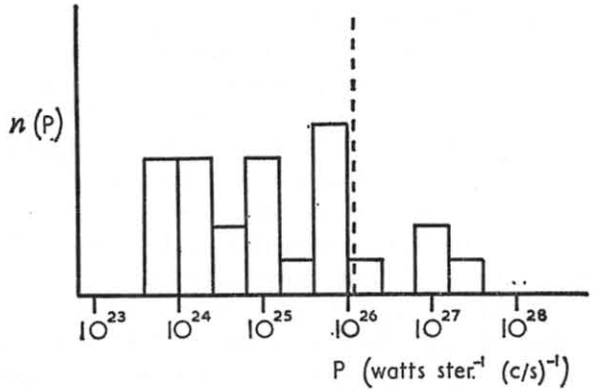


FIG. 18/5.—Histogram of the 'radio luminosity', P, for 24 of the 40 sources which have a flux density $S \geqslant 25 \times 10^{-26}$ w m^{-2} (c/s)$^{-1}$. The dashed line indicates the approximate limit of the present optical search for sources having $S \simeq 25 \times 10^{-26}$ w m^{-2} (c/s)$^{-1}$.

An independent determination of the luminosity function may be made by considering the radio data alone, by using a number of different types of observation (Ryle and Clarke, 1961). Suppose for example that we adopt a luminosity function of the form

$$n(P) = \rho_0 P_0^{3/2} \exp\left[-\tfrac{1}{2} (\log_{10} P/P_0)^2\right]$$

where ρ_0 represents the weighted mean space density of the sources. For distances at which the effects of the red-shift are

unimportant, the number of sources (N) per unit solid angle having flux density greater than S is given by

$$N = \tfrac{1}{3} \rho_0 P_0^{3/2} S^{-3/2}$$

From the observations that relate N and S we can therefore derive a value of

$$\rho_0 P_0^{3/2}$$

for the local region of space. It is in fact $1{\cdot}4 \times 10^{14}$ if P_0 is in w ster^{-1} (c/s)$^{-1}$ and ρ_0 in number per cubic parsec. Another quantity we can use is the degree of isotropy in the radio source distribution, for if an appreciable number of the radio sources lay within the Galaxy, we would expect to observe considerable anisotropy among sources at distances beyond a few kiloparsecs. In order to study this we have made some special observations at $+05°$ declination, a strip of sky which covers a region towards the centre at $b^{II} = 45$ and another region towards the anti-centre at $b^{II} = -55°$; these directions represent areas of sky where there is a large and small extent of the galactic halo. The results were examined by the $P(D)$ method and show no evidence of anisotropy out to 2×10^4 sources per steradian. This means that if the sources lie in the Galaxy, the space density ρ_0 must be at least 3×10^{-3} sources per cubic parsec. Another important number may be obtained from an examination of the sky brightness temperatures. From these we can derive the volume emissivity of various regions of the Galaxy, and this must clearly represent an upper limit to the value of $\rho_0 P_0$, the total power radiated per unit volume by sources, otherwise the radio sky would appear brighter than it is.

The three quantities that we have just discussed may be used to define limits to possible luminosity functions from the radio data alone. This is illustrated in Fig. 18/6, in which $\log P_0$ is plotted against $\log \rho_0$. We shall first exclude the possibility that most sources lie within the Galaxy.

The line (i) having a slope of $-3/2$ corresponds to the condition that

$$\rho_0 P_0^{3/2} = 1{\cdot}4 \times 10^{14}$$

which is derived from the source counts. Line (ii) indicates the minimum value of the space density ρ_0 that can occur if the majority of sources lie in the Galaxy, a result derived from isotropy considerations. Likewise, if most of the sources are in the Galaxy, a consideration of the total sky brightness shows

that $\rho_0 P_0$ must lie below line (iii). It is clearly impossible to satisfy these conditions; therefore our presumption that the sources lay in the Galaxy was wrong. The majority of the

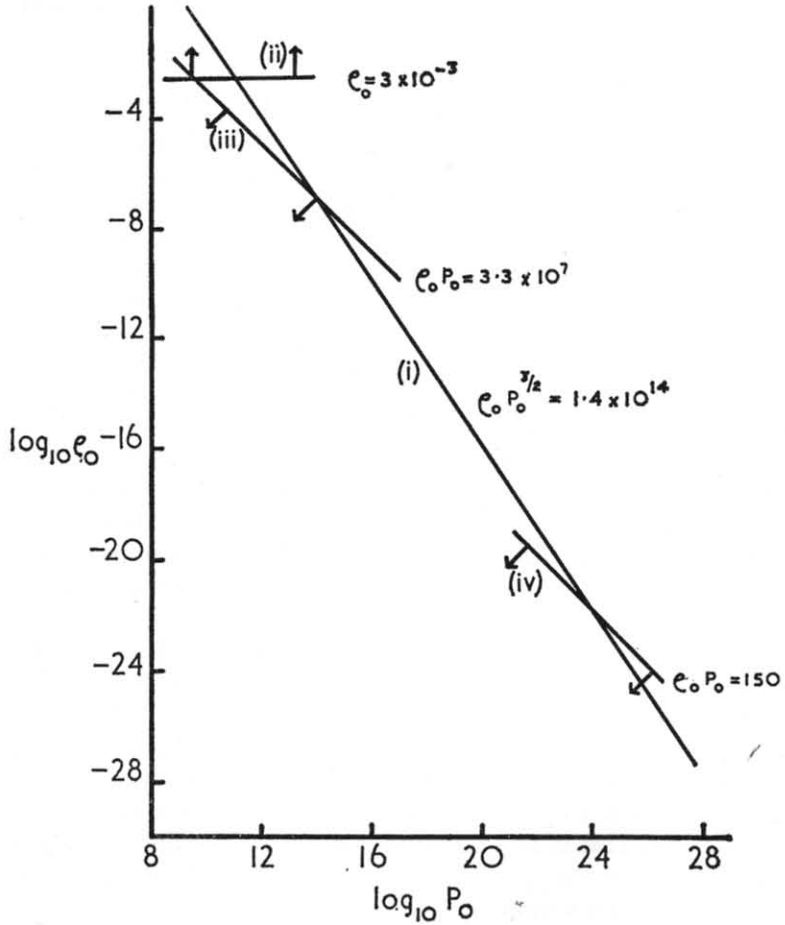


FIG. 18/6.—Diagram relating the values of the radio luminosity P_0 and space density ρ_0 which are compatible with the counts of sources falling in different ranges of flux density (i); the restrictions set by the observed isotropy are shown by (ii) and the limits set by the contribution of the sources to the integrated emission by (iii) and (iv).

sources must be extra-galactic, and not more than a few per cent lie in the Galaxy. Having established this point it is possible to place a lower limit on P_0 from considerations of the

contribution to the sky brightness temperature of the integrated extragalactic emission (line (iv) in Fig. 18/6). An upper limit to this is 25°K, which leads to a lower limit to P_0 of 10^{24} w ster^{-1} (c/s)$^{-1}$. This is clearly consistent with values obtained from the optical identifications (Fig. 18/5).

A third approach to the question of the radio luminosity function comes from the measurements of angular diameter. A large fraction of the identified sources are double, the two components having a separation of some 100 kpc. Assuming this to be typical of more distant sources, we can use the diameter measurements of the California Institute of Technology and Jodrell Bank to determine the distance and hence luminosity of a large number of sources. Quite apart from the intrinsic uncertainties in the physical scale of the sources, we have the difficulty that the relation between distance and angular diameter depends on the cosmological model. Clarke (1962) has shown, on the basis of the available data, that P_0, according to the Einstein-de Sitter model, may have a value of either about 10^{26} w ster^{-1} (c/s)$^{-1}$ or 3×10^{28} w ster^{-1} (c/s)$^{-1}$; the former figure would also be consistent with the Steady-State model.

CONCLUSIONS

Summarizing these results then, we find that the radio sources show no evidence of clustering, and are distributed isotropically. The median luminosity is certainly greater than 10^{24} and is probably about 10^{26} w ster^{-1} (c/s)$^{-1}$ at 178 Mc/s. We have a number/flux density relationship which shows a marked excess of sources in the range 1 to 5×10^{-26} w m^{-2} (c/s)$^{-1}$ compared with an Euclidean universe. There is some evidence that this excess is not continued to flux densities as small as $0{\cdot}1 \times 10^{-26}$ flux units. These facts I think rule out the possibility of explaining the log N/log S curve in terms of a local deficit of sources, for the region concerned would have to have dimensions of 350 Mpc or greater, a scale which cannot reasonably be regarded as a local irregularity. If we are to explain something which has a minimum scale of this order, it seems that we must bring in the general effect of the red-shift, and we can then see whether or not the different cosmological models would or would not explain what we have observed.

DISCUSSION

The problem of identification of sources was discussed, and the point was made that although the weaker sources are in general not identified with the present accuracy in positional information, improved accuracy may not greatly increase the number of sources which can be identified; the difference between the optical and radio emitting regions in a number of identified sources is already greater than the uncertainty in the radio position. Identification of large numbers of sources will be greatly assisted when the emission process is better understood.

Professor Lovell asked if the luminosity function discussed in the lecture was markedly different from that deduced by Professor Hanbury Brown (1962) from the Jodrell Bank diameter measurements.

In reply *Professor Ryle* said that the luminosity function derived by Hanbury Brown fell within the range of those he had discussed; the conclusions reached by Hanbury Brown concerning the interpretation of the form of the log N/log S relation were however entirely different, and did not appear compatible with the observations. Hanbury Brown supposed there to be a deficit of sources out to 50 Mpc, associated with a concentration of local sources to the supergalactic equator. Not only was there no evidence for such a concentration—even on the basis of Fig. 7 in Hanbury Brown's paper—but the curves of N/N_0 derived from this model had a maximum at a value of flux density approximately ten times that of the observed curve. The magnitude and form of the variation of N/N_0 could not be explained by such a region of low source density unless it had a scale very much greater than that of the supergalaxy, of the order of 350 Mpc (Clarke, Scott and Smith, in preparation).

In answer to a number of questions about clustering, *Professor Ryle* said that the range of angular separation of the component members which had been studied was 3′·5 to 200′ arc; this range was chosen as sufficient to cover the possible selection effects in the observations. The double nature of many of the sources corresponds to separations much smaller than 3′·5 arc; angular separations as small as this introduce negligible selection effects in observations with a 469λ interferometer.

Aggregations of sources into external superclusters having

dimensions of the order of 10^9 pc and subtending many degrees in the sky had been considered by Hoyle and Narlikar (1961); the investigation of any such assemblies necessitated the completion of a survey covering a large area of sky, and the aperture synthesis observations were now being continued in a way to allow such a study.

REFERENCES

Clarke, R. W., 1962, *Nature*, **194**, 171.
Clarke, R. W., Scott, P. F., and Smith, F. G. (in preparation).
Edge, D. O., Shakeshaft, J. R., McAdam, W. H., Baldwin, J. E., and Archer, S., 1959, *Mem. R.A.S.*, **68**, 37.
Hanbury Brown, R., 1962, *M.N.*, **124**, 35.
Hewish, A., 1962, *M.N.*, **123**, 167.
Leslie, P. R. R., 1961a, *M.N.*, **122**, 51.
Leslie, P. R. R., 1961b, *M.N.*, **122**, 371.
Ryle, M., and Clarke, R. W., 1961, *M.N.*, **122**, 349.
Scheuer, P. A. G., 1957, *Proc. Camb. Phil. Soc.*, **53**, 764.
Scott, P. F., and Ryle, M., 1961, *M.N.*, **122**, 389.
Scott, P. F., Ryle, M., and Hewish, A., 1961, **122**, 95.

19

COSMOLOGICAL THEORIES, A SURVEY

W. H. McCREA

THE purpose of this paper is to survey the foundations, scope and achievements of the main modern cosmological theories and also some of the possibilities of future developments. The discussion is intended to be as general and simple as possible; very little of the mathematical treatment is included. I aim rather to give an impression of what has been and is being done than to give a critical appraisal of the outcome.

COSMOLOGY

Cosmology is often described simply as the study of the astronomical universe *in the large*. This may mean either of two different things. (*a*) It may mean, as a convenient division for practical purposes, the part of astronomy that deals with systems of galaxies and with intergalactic matter, but not directly with anything on a smaller scale. With this interpretation, cosmology is not concerned with the physical behaviour of a single galaxy except in so far as it assists in the observation of systems of galaxies or in so far as it affects interactions between galaxies. (*b*) It may denote a study that is different in character from ordinary physics and astronomy. An ordinary physicist, chemist or astronomer may be interested in the properties of, say, an α-particle, water, or the interior of a star. Any of these things is of interest to him only because it exists in a physical sense. But he is not interested in how much or how many of them actually exist. This, however, is precisely what is of interest to the cosmologist. His concern is *the universe as it actually exists*. The reason he studies the universe in the large is primarily because it seems simpler to do so, or at any rate it seems to be possible to form a significant simplified picture of the universe by retaining only large-scale features.

The second interpretation *may* mean that the ordinary methods of science are not applicable to cosmology. We shall

return to this consideration later, but for the present we simply try what can be done using 'ordinary' methods.

We have then to ask what is the purpose of cosmological theory. We may say that it is to infer the past state and to predict the future state of the universe in the large. There seem to be quite specific questions in this context to which we believe answers ought to be obtainable. For instance, had we performed certain observations 10^9 years ago, following, so far as possible, exactly the same procedure as we do now, should we have got the same or different results? A rather different sort of problem for cosmological theory is, if we claim that a particular system of reference is inertial, what makes it so? In particular, with respect to what particular large-scale features of the universe is such a system not rotating? Or again, are the various large-scale characteristics of the universe, the rate of apparent expansion, the mean density of material, and so on, connected together by knowable relations? Again, are these characteristics connected by knowable relations with the basic properties of matter, i.e. the properties of the 'elementary particles' of physics?

COSMOLOGICAL MODELS

Our procedure is to seek to construct a theoretical *model universe*. We claim to have understanding of the actual universe to the extent to which our model reproduces our experience of the actual universe. Unless the contrary is evident, in this paper we are all the time speaking of some theoretical model or another and not directly of the actual universe.

Strictly speaking, we cannot even specify a model without using a theory. For the terms we should want to use, time, density, gravitation, and so on, have no precise meaning apart from a particular physical theory. But there are certain characteristics that, when described in sufficiently general terms, are common to all the models we shall consider. Such characteristics are:

(*a*) The contents of the model, matter and radiation, are distributed without local irregularities. So we say that it is a model of the *smoothed-out* universe. In fact, this is another reason why cosmology as here understood does not deal with the physics of an individual galaxy.

(*b*) The model is postulated to satisfy a *cosmological principle*

(CP). Thus we postulate the existence of a triple infinity of *fundamental observers* (FOs) such that no two ever meet (unless they all meet in one or more singular events) and such that the experience of the rest of the universe got by one can be exactly superimposed upon the experience of the rest of the universe got by any other. This is the sense in which we postulate the universe to be *homogeneous*.

(*c*) We postulate also that the universe has spherical symmetry about some FO. Then it follows that there is spherical symmetry about every FO. This is the sense in which we postulate the universe to be *isotropic*.

From the postulates, it is clear that a FO is moving with the material in his neighbourhood.

We can say that at any event in his experience a FO sees a world-picture. In general in the models to be considered two world-pictures seen by any one FO are not identical. A model in which world-pictures are not all the same, and which cannot be described as periodic, will be called *evolutionary*. Periodic models appear not to be of much interest and we shall expressly exclude them from further discussion.

The models we consider are such that if any two world-pictures seen by some FO are identical, then all world-pictures seen by all FOs are the same. We then have a *steady-state* model. This is said to satisfy the perfect cosmological principle (PCP).

If we consider any world-picture got by one FO, then according to our postulates, there is a corresponding world-picture in the experience of every FO. We may assign the same *cosmic epoch* to the events in the experiences of the FOs at which this picture is got. In this way, evolutionary models admit the existence of *cosmic time*. In other words, an evolving universe serves as its own universal clock. On the other hand, a steady-state universe does *not* admit the existence of a unique cosmic time. In this and other respects, it is misleading to regard a steady-state model simply as a particular case of models in general.

COSMOLOGICAL THEORIES: NEWTONIAN COSMOLOGY

We find we can construct a set of model universes entirely within the scope of Newtonian mechanics and gravitation.

Let t be Newtonian time reckoned from an arbitrary zero. Let F be a Newtonian frame of reference with origin O. Let there be matter whose distribution and whose state of motion is spherically symmetric about O. Let the matter obey Newton's law of gravitation, and let the total amount of matter be finite so that there is no doubt as to the applicability of Newton's law in its standard form. Let the matter be in the form of an ideal fluid of negligible pressure.

Let $\mathbf{q}$ be the position-vector relative to O of any particle of the matter. We wish to consider motion of the matter such that

$$\mathbf{q} = \mathbf{r}R(t) \quad . \quad . \quad . \quad . \quad (4.1)$$

where $R(t)$ is a function of t only, the same for all particles, and $\mathbf{r}$ is a fixed vector, characteristic of the particular particle considered. We then have

$$\frac{d\mathbf{q}}{dt} = \frac{R'}{R}\mathbf{q}, \quad \frac{d^2\mathbf{q}}{dt^2} = \frac{R''}{R}\mathbf{q}, \text{ etc.} \quad . \quad . \quad (4.2)$$

All this is to say that, knowing by the hypothesis of spherical symmetry that the motion has to be radial, we consider the particular sort in which at any instant the speed is proportional to the distance from O.

Further, we tentatively consider the case in which at some epoch t_0 the material has density

$$\begin{aligned} \rho &= \rho_0 \quad (r \leqslant r_0) \\ &= 0 \quad (r > r_0) \end{aligned} \quad . \quad . \quad . \quad (4.3)$$

where $r = |\mathbf{r}|$ and r_0, ρ_0 are given constants. Density is used here in the Newtonian sense in which matter is conserved. Therefore, from (4.2), (4.3), conservation is satisfied if and only if at epoch t

$$\begin{aligned} \rho &= \rho_0/R^3 \quad (q \leqslant r_0R) \\ &= 0 \quad (q > r_0R) \end{aligned} \quad . \quad . \quad . \quad (4.4)$$

where for convenience we take

$$R_0 \equiv R(t_0) = 1 \quad . \quad . \quad . \quad (4.5)$$

The mass $M(q)$ within distance q of O at epoch t is

$$M \equiv M(q) = \frac{4}{3}\pi\rho q^3 = \frac{4}{3}\pi\rho_0 r^3 \quad (q \leqslant r_0R) \quad . \quad (4.6)$$

which is constant for given r. Since the distribution is spherically

symmetric in a Newtonian frame, according to Newton's law of gravitation the attraction at $\mathbf{q}$ is

$$-GM\mathbf{q}/q^3 = -\frac{4}{3}\pi G\rho\mathbf{q} \quad (G = \text{gravitational constant}) \quad (4.7)$$

Hence, pressure being neglected, the equation of motion of a particle of the fluid at q is

$$d^2\mathbf{q}/dt^2 = -\frac{4}{3}\pi G\rho\mathbf{q} \quad . \quad . \quad . \quad (4.8)$$

Using (4.2), (4.4), this becomes

$$R^2R'' = -\frac{4}{3}\pi G\rho_0 \quad . \quad . \quad . \quad (4.9)$$

This has a first integral

$$R(R'^2 + kc^2) = \frac{8}{3}\pi G\rho_0 \quad (kc^2 = \text{constant of integration}) \quad (4.10)$$

If then $R(t)$ is any solution of (4.9), the whole system in motion is a possible one in accordance with Newtonian theory. In particular, our tentative hypothesis about the density is justified as being compatible with the type of motion considered. Also, we note that the behaviour at any point inside the fluid is entirely independent of the parameter r_0.

Any particular system of the sort contemplated is specified by the value ρ_0, the function $R(t)$ which must be a solution of (4.9) and so depends upon two arbitrary constants of integration, and the independently assignable r_0.

Conversely, if an observer sees fluid all around himself having motion specified by the relation $\mathbf{q} = \mathbf{r}R(t)$ and density ρ_0/R^3, where ρ_0 is constant and $R(t)$ satisfies equation (4·9), then he can claim that all he sees is consistent with the hypothesis that he is in a Newtonian frame, that the distribution is spherically symmetric about himself, and that the material obeys Newton's law of gravitation.

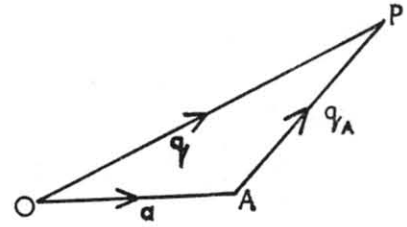


FIG. 19/1.—A triangle of velocities in Newtonian mechanics.

Now let A with $\mathbf{q} = \mathbf{a}$, say, be any particle of the fluid different from O. Let G be a frame, origin A, not rotating relative to F. Let P be a general particle of the fluid and let $\mathbf{q}_A$ be its position vector in G. Then by the triangle of velocities, the velocity of P relative to A is $(R'/R)\mathbf{q}_A$, and, of course, the density is ρ_0/R^3 as before. Thus the observer moving with A is precisely in the situation of the observer just mentioned.

This means that every observer moving with the fluid is equivalent to every other so far as his observation of the fluid is concerned, and each sees the behaviour of the fluid as spherically symmetric about himself.

So far as we have gone, the only feature that does not conform to this equivalence is the boundary of the fluid. Any observer other than O, if he sees the boundary, does not see it as symmetrical about himself. (He would then conclude that he is not moving with a Newtonian frame, but we need not pursue this now.) However, we have seen that the parameter r_0 enters into none of the equations. Also we may suppose the range of vision of any observer to be finite. If we consider the case where r_0 is arbitrarily large compared with this, then only an arbitrarily small fraction of all observers will have even this evidence of lack of symmetry. For all the remaining observers, the only feasible hypothesis for any one is that he is a centre of symmetry for the whole system. Making this hypothesis and treating his frame as Newtonian, he will give a correct account of the motion of the fluid.

On the same hypothesis, as it is easy to show, he will also give a correct account of the motion of a free test-particle.

Thus we have a set of effective Newtonian frames each in accelerated motion relative to any other.

Were we to postulate a constant light-speed everywhere relative to the original frame F, then this would be a property that would not allow equivalence of the observers. However, we find that we can instead postulate that the light-speed has the constant value c relative to the local 'Newtonian' frame determined as above. Obviously, this again leaves all the frames equivalent to each other and it is found to introduce no inconsistencies. Indeed, it enables us to define 'horizons' for any observer and in effect it imposes an inevitable limitation upon the range of vision. So it reinforces the features to which we are calling attention.

MODEL UNIVERSE AND ACTUAL UNIVERSE

Now let us treat the system as a model universe. For the moment, we ignore the boundary.

The observer moving with the fluid at any point is then a FO. He describes the rest of the universe as receding ($R' > 0$) from himself with speed proportional to distance. From (4.2), the factor of proportionality is

$$R'/R = 1/T, \text{ say} \quad . \quad . \quad . \quad (5.1)$$

We may then call T the *Hubble time* for the model; it is in general a function of the epoch of observation t. The model behaves in accordance with a known system of mechanics and gravitation, provided $R(t)$ satisfies an equation of the form (4.9). From this equation we see that if $\rho \neq 0$ then $R'' \neq 0$ and therefore $R' \neq 0$; *thus there is no static non-empty model.* We can go further and say that there is no non-empty model for which T is independent of t.

Using (4.2), (5.1) equation (4.10) may be written in two instructive forms

$$\tfrac{1}{2}\left(\frac{dq}{dt}\right)^2 = \frac{GM}{q} - \tfrac{1}{2}kc^2r^2 \quad . \quad . \quad . \quad (5.2)$$

$$G\rho T^2 = \frac{3}{8\pi}(1 + kc^2/R'^2) \quad . \quad . \quad (5.3)$$

The first allows us to state that the particle of the fluid at $\mathbf{q}$ has speed greater than, equal to, or less than the escape speed from the material inside distance q according as $k < 0$, $k = 0$, $k > 0$. The last case, $k > 0$, is one where the material would fall back again even if it is now receding. This makes it periodic in our sense and we have excluded this case from present consideration. Thus we are concerned only with cases $k \leqslant 0$ whence, using (5.3),

$$G\rho T^2 \leqslant \frac{3}{8}\pi \quad . \quad . \quad . \quad . \quad (5.4)$$

Consider this model as a representation of the actual universe. As already said, it can be a model only of the smoothed-out universe; so we are supposing that if we consider the actual universe on a sufficiently large scale we may treat the contents in a significant manner as a continuous fluid and that the local irregularities in the actual distribution are not of essential importance on this scale. Or, rather, amongst other things, we are testing the hypothesis that the contents may be so considered.

Without going into details, we may say that treated in this way the actual universe does appear to conform to the CP with $\rho \neq 0$. Also the apparent recession of the galaxies to the first order, which here means to a very good approximation, conforms to the simple linear 'Hubble' law of speed proportional to distance. The empirical value of T got in this way is approximately

$$T = 10^{10} \text{ years} \quad \text{or} \quad T = 3 \times 10^{17} \text{ seconds} \quad . \quad (5.5)$$

Then (5.4) gives approximately

$$\rho \leqslant 2 \times 10^{-29} \text{ g. cm}^{-3} \quad . \quad . \quad . \quad (5.6)$$

The mean density of known matter in the universe is in fact estimated to be not more than about 10^{-29} g. cm^{-3}.

Finally, so far as we can tell, what we treat as an inertial frame in the actual universe is in some way determined by the large-scale distribution of matter as we have seen to be also a property of the model.

All this is satisfactory. But we can judge how satisfactory it is only by asking if we can reasonably imagine things to have been otherwise.

In fact, we could imagine the distribution of matter to have been found to be so irregular that no smoothing would leave anything significant. Or, were smoothing significant, on general grounds we should indeed expect the radial motion of the smoothed out medium to be to a first approximation linear with distance, but (*a*) we could imagine the approximation being good to only a cosmically small distance; (*b*) we could imagine the factor of proportionality with distance to be different in different directions. In the actual universe we do get the same factor in all directions. Even so, we could still imagine the apparent expansion to be some 'accidental' local phenomenon as viewed on a cosmic scale, or we could imagine the apparent expansion to be produced in whole or in part by optical effects having nothing to do with the mechanical properties of matter. In neither case would we then expect the rate of apparent expansion to be related in a simple manner to the mean density of matter 'in the universe'.

In fact we have here a theory which tells us that the model

universe cannot be in a static state and tells us that the mean density and the rate of expansion are related in a definite simple way (since any particular model gives a definite value to the left-hand side of (5.4)). Then when we find that the actual universe is apparently not static and that the agreement with (5.4) is good we conclude both that the apparent motion is rightly treated as an ordinary motion and we claim that our account of it possesses significance.

It should be mentioned that another way of regarding (5.4), or the corresponding precise relation for a particular model, is that it is a determination of the value of the gravitational constant G by the contents of the universe and their behaviour as summarized by the values of ρ, T.

RELATIVISTIC COSMOLOGY

In relativistic cosmology we start again with the same general notions as before about the existence of FOs. Using simply the basic concepts of what we can call the kinematics of general relativity, i.e. that the aggregate of events are the points of a Riemann 4-space and that the world-lines of light-signals are null geodesics, we find that the only space-times conforming to these notions are of the form

$$ds^2 = c^2dt^2 - R^2(t)(dx^2 + dy^2 + dz^2)/(1 + \tfrac{1}{4}kr^2)^2 \quad . \quad (6.1)$$

where $r^2 = x^2 + y^2 + z^2$, these being the Robertson-Walker metrics.

The FOs have the world-lines $x = \text{const.}$, $y = \text{const.}$, $z = \text{const.}$ Let O be the one with $x = y = z = 0$. Then the 'ordinary' distance of any other from O is

$$q = rR(t) \quad . \quad . \quad . \quad . \quad (6.2)$$

where r is fixed for the other FO.

If we now appeal to Einstein's field-relations, they tell us that the density ρ and pressure p in the universe (6.1) are given by

$$8\pi G\rho = \frac{3(R'^2 + kc^2)}{R^2} \quad . \quad . \quad . \quad (6.3)$$

$$8\pi\frac{Gp}{c^2} = -\frac{2RR'' + R'^2 + kc^2}{R^2} \quad . \quad . \quad (6.4)$$

Now put $p = O$ as before and (6.4) gives

$$2RR'' + R'^2 + kc^2 = 0$$

which has the first integral

$$R(R'^2 + kc^2) = C \quad . \quad . \quad . \quad (6.5)$$

where C is an integration constant. Then from (6.3)

$$\frac{8}{3}\pi G\rho = C/R^3 \quad . \quad . \quad . \quad . \quad (6.6)$$

Writing $C = \frac{8}{3}\pi G\rho_0$ where ρ_0 is a constant, (6.5), (6.6) become

$$R(R'^2 + kc^2) = \frac{8}{3}\pi G\rho_0 \quad . \quad . \quad . \quad (6.7)$$

$$\rho = \rho_0/R^3 \quad . \quad . \quad . \quad . \quad (6.8)$$

Equations (6.2), (6.7), (6.8) *are seen to be exactly the same as* (4.1), (4.10), (4.5).

This is not the place to rehearse the curious history of modern cosmology. Broadly speaking, it is true to say that the expansion of the universe was predicted by general relativity. When Hubble discovered it as apparently an empirical property of the actual universe—the most extensive property known—this was naturally hailed as a triumph for relativity theory. This theory was much more difficult to understand than classical theory but all that it had yielded hitherto had been some minute departures from classical theory that had been only uncertainly verified. However, when it came to the large-scale behaviour of the universe, general relativity predicted this stupendous phenomenon of the expansion of the whole system. This was something that classical theory had been powerless to treat. Only afterwards it was found that classical theory can do something about it—to the extent of yielding exactly the same equations!

Now the homogeneity of the universe means that its large-scale behaviour is the outcome of the operation everywhere of the same local behaviour. But the local behaviour is such that we expect it to be adequately described by classical theory. So it is not as astonishing as it appears at first sight that the behaviour in the large is also so described (within limitations that will appear).

There are differences in other respects:

(*a*) In the classical treatment we never quite dispose of the

boundary. The relativistic theory treats every FO as strictly equivalent to every other in every respect.

(*b*) The geometry is affected by the presence of matter in relativity theory but not in classical theory. Therefore certain integral properties are different even though differential properties are the same. The consequent observable differences enter only beyond the first approximation.

(*c*) If we do not neglect *stress* in the model, then the equations are no longer the same. This may be important, but we cannot do more than mention it here.

To state the position briefly, and perhaps rather loosely, relativistic cosmology is a more satisfying treatment of the same physical phenomena as those treated in Newtonian cosmology. Thus the physical discussion of the actual universe according to relativistic cosmology is effectively the same as that according to Newtonian cosmology. In particular we are led to a system of *evolutionary cosmology* as will be seen in the next section.

EVOLUTIONARY COSMOLOGY

We can write (6.5) in the form

$$R'^2 = \frac{C}{R} - kc^2 \quad . \quad . \quad . \quad (7.1)$$

where $C = \frac{8}{3}\pi G\rho_0 > 0$ if $\rho_0 > 0$. We are interested only in $R \geqslant 0$. Without solving (7.1) explicitly (in elliptic functions) we can see that the solutions are of the forms shown qualitatively in Fig. 19/2.

As already stated, we do not wish to consider the case $k > 0$. Then we are left with *evolutionary models*. We see that these start in a singular state; this is the 'big bang' of such models.

For definiteness, take the case $k = 0$. Then we find from (7.1) that $R(t) = \text{constant } t^{2/3}$, giving in this case

$$\frac{R'}{R} = \frac{2}{3t} = \frac{1}{T} \quad \text{or} \quad t = \frac{2}{3}T$$

where we have recalled the definition (5.1). This gives the time from the big bang to the epoch of observation, or what we call

the *age of the universe*. Also we can show that for all these evolutionary models the age is between $\frac{2}{3}T$ and T.

If the actual universe were built on one of these models and we use the value (5.5), we find the age of the universe to be

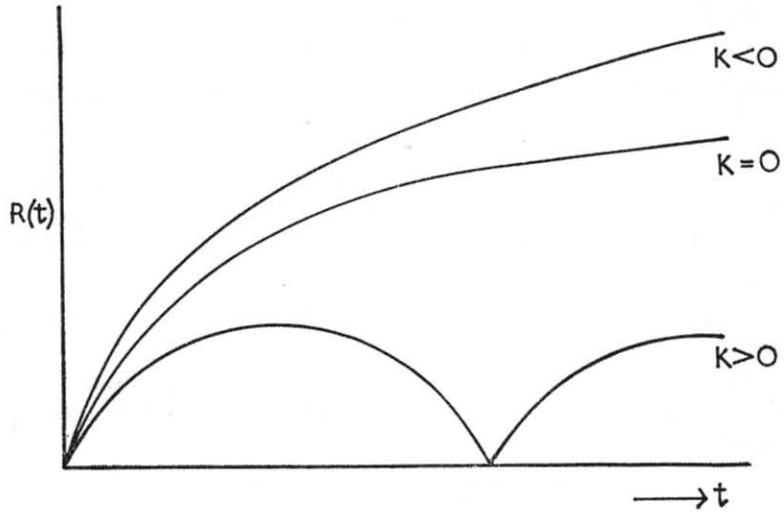


FIG. 19/2.—The form of solutions for $R(t)$ for different values of k.

between about 7 and 10×10^9 years. This appears to be probably too short to accommodate what we know about stellar evolution. Moreover, there is no known feature of the actual universe that gives any indication of its ever having been in a state of extreme congestion as required by the models.

STEADY-STATE COSMOLOGY

For reasons such as those just stated, Bondi and Gold and Hoyle proposed in 1948 a system of steady-state cosmology. There were also more philosophical reasons. For instance, the concept of a changing universe behaving in accordance with unchanging laws is hard to accept because we naturally think of the universe and its laws being interdependent. While we cannot say that the universe must be in a steady state, for the reasons just indicated it is of much interest to examine the possibility.

We return to the treatment used at the beginning of section 4. We are now going to postulate that *all characteristics of the model are independent of the epoch.* So in the first place if we write as before

$$R'/R = 1/T \quad . \quad . \quad . \quad . \quad (8.1)$$

then T is now a constant independent of t. Therefore we obtain

$$R = e^{t/T} \quad . \quad . \quad . \quad . \quad (8.2)$$

where nothing is gained by inserting a constant of integration.

Again, we must now have a constant value for ρ. Consider then a sphere about O of any given radius q. The matter flows across its surface with speed q/T and so the total rate of outflow is $4\pi\rho q^3/T$. Hence new matter has to be created inside the sphere at the rate $3\rho/T$ per unit volume per unit time. *This must be the universal rate of creation in the model.*

Let $f(t)\rho$ be the density of material of age greater than or equal to t. Consider a sphere of radius q whose surface is moving with the medium. In time Δt the volume changes from V to $V + \Delta V$, say, and the material that was of age t becomes of age $t + \Delta t$. So we have

$$f(t + \Delta t)(V + \Delta V)\rho = f(t)V\rho$$

giving

$$f'/f + V'/V = 0$$

But

$$V = \frac{4}{3}\pi q^3, \; V' = 4\pi q^3/T$$

and so we obtain

$$f'/f + 3/T = 0$$

Hence

$$f(t) = e^{-3t/T} \quad . \quad . \quad . \quad . \quad (8.3)$$

since we must have $f(0) = 1$. From this it follows that the mean age of the matter in any neighbourhood is

$$\frac{3}{T}\int_0^\infty t\, e^{-3t/T} dt = \frac{1}{3}T \quad . \quad . \quad . \quad (8.4)$$

This is also the mean age of systems belonging to any general category in the model.

These properties are all 'local' properties and so they are the same whether we use the space-time of classical or of relativity theory.

If we use (8.2) in (6.1) and put $k = 0$ we have the space-time

$$ds^2 = c^2dt^2 - e^{2t/T}(dx^2 + dy^2 + dz^2) \quad . \quad (8.5)$$

Now change t to $t + t_0$ where t_0 is any constant; this metric is unchanged if we also change x into $e^{-t_0/T}x$, etc. Thus the geometry is unchanged by a change in the origin of t. So we verify that the space-time (8.5) is indeed appropriate for the description of a steady-state universe. We should not have got this result with $k \neq 0$.

If we attempt to apply Newtonian gravitation or the field-relations of general relativity, again taking $p = 0$, we find that the motion is not in accordance with these laws. In fact, the material is accelerated away from 0 with acceleration $\mathbf{q}/T^2$ and so it would be interpreted as showing gravitational repulsion rather than attraction.

Many attempts have been made to use modified field-relations of gravitation or of electromagnetism in order to give a theory of the behaviour of the model and, in particular, to give a relationship between ρ, T. Most such attempts do in fact give a value of the order of unity for the quantity $G\rho T^2$. However there is nothing especially compelling about any of these attempts, which suggests that a satisfactory approach has not yet been discovered. The *simplest* suggestion appears to be that of the present writer who has pointed out that we can obtain the desired behaviour *without* modifying the field-relations of general relativity provided we admit the possibility of a suitable zero-point stress.

PHYSICAL DISCUSSION

If the actual universe were built in the steady-state model, the mean age of any characteristic entity is $\frac{1}{3}T = 3 \times 10^9$ years, approximately. However, a fraction e^{-3} or about 5 per cent of such entities are of age greater than T, about 1 per cent are of age greater than $1{\cdot}5T$, and so on. This avoids the objection to the evolutionary models in which nothing at all is older than some definite age less than T. Also, there never was a singular state like that required by evolutionary models. The rate of creation of new material is $3\rho/T = 10^{-46}$ g. cm^{-3}s^{-1}, approximately, or one atom of hydrogen per cubic centimetre in about 5×10^{14} years.

There are observable differences between the steady-state universe and any expanding evolutionary model. For, if we look at a region of the universe τ light-years away we see it as it was τ years ago. According to evolutionary cosmology the region must therefore look to be more congested than our own region; according to steady-state cosmology, the congestion must be the same as that in our own region. But it can be shown that the difference is only of the order of the fractional red-shift for the region considered. Again, according to evolutionary cosmology there is an age-distance effect but not according to steady-state cosmology. At present, we do not know how to apply this since we do not know how individual galaxies change in appearance with age.

In practice it appears that no thoroughly satisfactory observational test has yet been possible although stupendous efforts have been made in this field.

To sum up, we may say that steady-state cosmology possesses practically all the satisfactory general features that we previously noted for Newtonian or relativistic cosmologies, while it avoids the unsatisfactory features associated with a definite age of the universe. Steady-state cosmology possesses also the philosophically agreeable feature of yielding an effectively unique model, instead of the infinite set yielded by other theories. Its predictions appear not to have been contradicted by observation. At present it has the disadvantage of not resting upon any accepted 'field-theory'.

OTHER POSSIBILITIES

We have not considered the possibility of including the so-called cosmical constant and *cosmical terms* in Einstein's field-relations. This should be considered but it is not possible to say anything conclusive about it at present.

Milne's *kinematic relativity* offers a different approach to the subject, but the model to which it leads is included amongst those of relativistic cosmology.

The possibility of cosmical electrodynamical effects has been proposed, but a critical discussion shows that nothing characteristically electrodynamic survives in the resulting cosmological model.

CRITICAL COMMENTS

If we know exactly how any two portions of matter interact then, if we make all possible observations of the matter in the universe and its motions, we could use our knowledge in order to predict the outcome of future observations. But I have pointed out elsewhere that there are apparently unavoidable limitations upon the information that we obtain from our observations. This means that there are apparently inescapable uncertainties in our predictions, these uncertainties increasing with distance from ourselves in time and space.

In particular, we should not expect to be able to escape from this situation by using any of the theory described in the earlier part of this paper. Therefore I conclude that the theory is deficient in certain fundamental respects.

I should expect a more satisfactory theory to make precise predictions when applied to cosmically small regions around us in time and space, but that the precision would decrease as we seek to consider larger and larger regions until in dealing with the largest regions accessible to observation we can only describe what we do see and predict almost nothing of what we shall see.

Thus I should expect much of the discussion of this lecture to remain valid but within limitations that should become better understood. The understanding of these limitations ought to include an understanding of the lack of success in discriminating observationally between various sets of predictions.

There is at any rate one thing about which we are almost certain that no certain prediction can be made, and that is the future development of cosmological theory.

20

COSMOLOGY AND RADIO ASTRONOMY

A Discussion

Contributors: G. C. McVittie, W. H. McCrea, M. Ryle, H. P. Palmer

Aspects of Cosmology by G. C. McVittie

The expansion of the universe

I would like you to consider why we think that the universe is expanding. It may seem pointless to state the obvious, but we notice that the configuration of the galaxies in the sky never changes. Although the distant galaxies are receding at great speed, so great is their distance that even in a thousand years no change in their apparent size, brightness or relative separation could be detected. Thus the only evidence we have for the expansion of the universe is the shift towards the red of the lines in the spectra of faint galaxies. I would like to tell you how the two parameters describing the red-shift phenomenon are obtained. The first of these is the Hubble constant which is denoted by the letter H and which gives the present rate of expansion in our neighbourhood. The second parameter, q_0, measures the acceleration or retardation of the expansion at the present time. In order to obtain the values of these parameters, the red-shifts of all measured galaxies can be plotted logarithmically against the apparent magnitude of the object. A diagram (see Humason, Mayall and Sandage, 1956) with a cloud of points on it is thus obtained and through the cloud, with the eye of faith, a straight line may be drawn. Such rough approximate treatments are the origin of the belief, held with an almost superstitious tenacity, that the velocity of recession of a galaxy is proportional to its distance. A more refined analysis is based on clusters of galaxies. Red-shifts of galaxies in a cluster are measured and their average is taken to be the red-shift of the cluster as a whole. To obtain comparative distances of clusters,

it is necessary to use the apparent brightness of some object in each cluster which is believed to have the same intrinsic luminosity in every cluster. Clearly the integrated intensity of the entire cluster will not do as clusters vary in size and membership. But a composite brightest galaxy in a cluster can be set up by observing the apparent magnitudes of the 1st, 3rd, 5th and 10th brightest members of the cluster. This procedure amounts to establishing a rudimentary luminosity function for galaxies in clusters. If the intrinsic luminosity of the composite brightest member of every cluster is assumed to be the same, then one has a relative scale of distances for the clusters. Using the apparent magnitude of this composite cluster member in conjunction with the mean red-shift of the cluster, one can determine directly, not the Hubble constant itself, but the rate at which the expansion is changing. This is because, to find H, one must also know the intrinsic luminosities of the brightest cluster members, whereas to find the acceleration factor q_0 it is only necessary to assume that the same (unknown) luminosity function applies to clusters at all distances.

A model of the universe predicts a relationship between photographic apparent magnitude, m, and the red-shift $d\lambda/\lambda = \delta$ of the form

$$m - 5 \log c\delta = a(1 + b\delta + c\delta^2 + d\delta^3 + \ldots + \ldots)$$

in which a is effectively the Hubble constant, and b is related to the acceleration parameter. There is some difference of opinion as to the value of 'b', but the extreme limits give for the acceleration factor,

$$0{\cdot}5 \leqslant q_0 \leqslant 3$$

This expression corresponds to a retardation of the acceleration, and a reasonable interpretation leads to rather closer limits on q_0, i.e.

$$1 \leqslant q_0 \leqslant 2$$

The Hubble constant may be said to lie between 75 and 150 km/sec per Mpc, a more accurate determination depending on a knowledge of what the intrinsic luminosity of the brighter galaxies in clusters is. The values of H, and more particularly of q_0, enable us to limit the range of models which adequately represent that part of the universe accessible to our observations.

Number/intensity counts in cosmology

I would like to make a few comments on the optical analogue of what Professor Ryle and others are doing when they count radio sources to successive intensity limits. In the 1930's Edwin Hubble made a very strong and determined effort to deal with this problem. He counted optical galaxies in five ranges (from magnitude 18 to magnitude 21·3). When he counted the number N of galaxies brighter than magnitude m, Hubble (1936) obtained a number/intensity curve of the form

$$\log N = 0{\cdot}501m + \text{const.}$$

i.e.
$$N \propto S^{-1{\cdot}25}$$

whereas in an Euclidean universe, the law would be

$$\log N = 0{\cdot}6m + \text{const.}$$

i.e.
$$N \propto S^{-1{\cdot}5}$$

where S is the observed flux density of optical radiation.

Thus we see that what Hubble thought he had found, a deficit of faint galaxies compared with the Euclidean universe, was the very opposite of the result obtained by Ryle from the radio data. What Hubble did was to assume that the departure of the factor in his number/magnitude relation from the value 0·6 was due to a red-shift effect, and he recast the relationship in the form

$$\log N = 0{\cdot}6(m - A\delta) + \text{const.}$$

where A is a constant and δ the red-shift. This is a very restricted way of fitting the data, and is not by any means the only interpretation (McVittie, 1938, 1956). However it is said that Hubble himself later rejected the data on which these relationships were based. We may here notice a curious feature with respect to cosmological data: they have a weight of either one or zero! An intermediate weight, such as 0·5, is apparently unacceptable. Be this as it may, I would warn you that, if ever you want to use Hubble's data or his interpretation of them, do not simply copy out formulae from his papers. Look at his work in its entirety, and also study the literature of the 1930's to which it gave rise, before making up your own minds. If indeed all this work of Hubble's, which was the chief basis for our belief that the distribution of galaxies is uniform, is unreliable, I am puzzled to know why we continue to accept with such faith the notion of uniformity.

Discussion

In answer to questions, Professor McVittie discussed the desirability and difficulties of making a new number versus apparent magnitude count of galaxies. It would be a heavy task. A telescope beyond the Earth's atmosphere would help, but the greatest difficulty lay in finding an enthusiastic band of young astronomers and a suitable telescope even on Earth.

REFERENCES

Hubble, E. P., 1936, *Ap. J.*, **84**, 517.
Humason, M. L., Mayall, N. U., and Sandage, A. R., 1956, **61**, 97, Figs. 3, 6, 10.
McVittie, G. C., 1938, *Observatory*, **61**, 209.
McVittie, G. C., 1956, *General Relativity and Cosmology*, p. 172. London: Chapman and Hall.

Observational Tests in Cosmology

W. H. McCrea

In setting up evolutionary models of the universe, it is usual to consider a 'smoothed-out' distribution of matter. A problem arises in the formation of galaxies from this homogeneous material under gravitational forces. If ρ is the density, the only time scale you can construct dimensionally is $(\rho G)^{-1/2}$, and this is of the order of the age of the universe on any model at all. Thus it is very difficult on any of these theories to see how galaxies can have been formed in the time available. It seems to me that this is a problem invented by us, not given to us by nature, for how do we know that the universe was ever smoothed out? Unfortunately it is mathematically very difficult to handle models having non-uniform initial states. (At this point Professor Burbidge said that this amounted to avoiding the problem entirely by assuming proto-galaxies. Professor McVittie said that the 'smoothed out' universe was a convenient simple description of the universe, and not an exact description of an early stage in its evolution; Professor McCrea agreed.)

The next point that I would like to comment on is very fundamental indeed, although again theory has not really coped with it. As Professor McVittie said, what you would like to deal with is our own observable part of the universe. He went on to

discuss the m, δ relationship, and what you can get out of it about the acceleration and so on. But you get the relations from a homogeneous smoothed-out model and nobody knows whether you would get any sort of similar relations if you had our own local observable part of the universe in front of some different sort of background. I have looked at this problem from time to time, but do not even see how to begin. One would like to be able to say that what happens beyond the observable universe is of no consequence, but that is not at present possible, and there is a danger of an inconsistency in using the formulae for retardation of the expansion, etc.

Many of the tests which can be applied to cosmological theories are difficult and round-about. In principle what we would like most I think is a way of distinguishing between an evolutionary and a steady-state model of the universe, by a means which does not depend on a particular geometry. If you could look at any distant part of the universe at something that you know is a standard object, and measure its angular size (for example the angular separation of the components of a radio source), you could then compare this angle with the angular separation of the nearest neighbour in space (having the same red-shift) and thus obtain a direct measure of the spatial density of galaxies. This procedure would enable you to determine whether the density of galaxies is the same at remote distances as it is locally, in a manner which does not depend on the geometry. The difference between the predictions of spatial density on the evolutionary and steady-state theories is a factor of $(1 + z)$, so that your observations have to be accurate to the order of z to give any information at all. (Here $z = d\lambda/\lambda$ is a measure of the red-shift.)

I think these ideas are significant in principle in another way. In 1959 Metzner and Morrison published a paper on the application of information theory to cosmology. They showed that if you take unity to represent the flow of information received from nearby parts of the universe then the theory gives you zero as the flow of information you get from, so to speak, the horizon. In fact the flow of information does contain the factor $(1 + z)^{-1}$.

Thus we seem to be faced with the paradoxical situation that the information available from the universe falls off in exactly that ratio $(1 + z)$ that represents the ratio of quantities predicted by the various cosmological theories. This I think is a point of fundamental importance in applying observational tests to cosmology because it may mean that a decisive test is not possible in principle.

Discussion

Professor Lovell: 'Are world models with a positive cosmological constant no longer considered seriously? The models we have so far heard about all have a singularity near the Hubble time, which in the past has caused difficulty with the apparent great age of galaxies.'

Professor McCrea: 'In ordinary classical mechanics the absolute value of the stress has no significance. Only gradients and discontinuities are important. The cosmological constant is just the analogue of the arbitrary constant which appears in the stress in classical mechanics. I do not see how to take this into account in cosmology, and it is not easy to see how anything so trivial in principle can have much significance.'

Professor McVittie: 'The cosmological constant may be determined from Einstein's equations. If you accept that the density is $\sim 10^{-30}$ gm cm^{-1} and that the retardation factor lies between 0·5 and 1·0, then Λ is negative, and we have a singularity at $t = 0$. The negative value of Λ implies that a force other than gravitation holds matter together in the universe. This is generally disregarded by cosmologists, on the basis of a paper by Einstein in 1932 in which he showed that the data *then available* were consistent with what we now call the Einstein-de Sitter universe. Einstein might have said something quite different were he alive today.'

Professor McCrea: 'I think it is much more complicated. Λ is an arbitrary constant which can be determined only through atomic physics, if there is any significance in zero point energies and zero point stresses. It cannot be determined from astronomical considerations and, so far as I can see, it denotes a degree of arbitrariness inevitably present in relativistic cosmology.'

REFERENCE

Metzner, A. W. K., and Morrison, P., 1959, *M.N.R.A.S.*, **119**, 657.

Radio Astronomical Tests of Cosmological Models

M. Ryle

I shall discuss four methods which are available to the radio astronomer for making observational tests in cosmology. At present it turns out that there are more variables in the theories than can be determined from the observations, and you have to try to fit a plausible situation to all the evidence you have. The most promising methods are based on:

i. *The integrated sky brightness*

If one could measure the contribution to the total sky brightness which arises outside the Galaxy, and also measure the contribution to the emission per unit volume from all the nearby extragalactic sources, then the observed ratio of these two quantities could be compared with that predicted on various cosmological models. Obviously there are practical difficulties: the determination of the relative contributions of the Galaxy and extragalactic sources to the sky brightness is very difficult, but there is some hope of doing it, at least at the longer wavelengths, because the two components have different spectral indices.

ii. *Measurement of the density of intergalactic hydrogen*

In principle it should be possible to determine the density of neutral hydrogen in space by observing the 21 cm line in absorption. The continuum spectrum of a distant source like Cygnus should show an absorption feature extending from 1420 Mc/s to the doppler-shifted value corresponding to the red-shift of the source. This is a very difficult experiment, quite apart from the fact that the intergalactic material may be ionized.

iii. *Radio source counts*

Here one would like to measure ρ, the spatial density of sources, as a function of the red-shift. This cannot be done in the absence of distance measurements, but by making the general assumption that the fainter sources are further away, we can get somewhere from the observed number/flux density counts. This naturally involves some knowledge of the typical luminosity, P, of a radio source, and if

the universe is evolutionary, so that P varies with epoch, it may not be possible to interpret the $\log N/\log S$ curve uniquely at all. However, it is possible that the effects will be so remarkable that we can already exclude some models of the universe. If this can be done it is of course a very worthwhile result in the present state of cosmology.

iv. *The angular diameter of radio sources*

Ideally one would like to measure the angular diameters of radio sources as a function of their red-shift, whereas the observations yield directly a measure of surface brightness and flux density. The surface brightness of a source of given physical size varies in some way, depending on the model, with distance; the intensity at the receiving frequency also varies with distance in a manner depending on the spectrum. For sources of constant linear size, some cosmological models predict a monotonic decrease of angular diameter with increasing red-shift while others predict a minimum angular diameter at a certain red-shift, beyond which the angular diameter increases. If one could measure the angular diameters of a large number of radio sources, this would therefore provide evidence to distinguish between certain models. Dispersion in the luminosity and in the physical size of sources, complexity of structure and orientation make this approach difficult, although in principle it should be improved by combining the diameter measurement with number counts.

In order to proceed further, we must compare the predictions of various cosmological models with these observations. Suppose for example we could measure in detail the relationship between surface brightness and flux density. This, on any given model, allows us to deduce something about the luminosity function. We can then use this function in conjunction with our source counts and optical information to see if a plausible picture emerges. I think myself that the present results show so large a departure from those predicted by the steady-state model that we must recognize that we live in an evolving universe, in which the luminosity or number of the sources depends on the epoch. The alternative seems to be to regard the results as arising from a 'local' irregularity as proposed by Hoyle and Narlikar (1961), but such an irregularity would have to extend

to distances of about 350 Mpc, a distance which is scarcely local, and also maintain an angular distribution which is very nearly isotropic.

Discussion

Dr. R. D. Davies: 'We have made hydrogen line measurements which set an upper limit to the optical depth of intergalactic hydrogen in the direction of Cygnus A. The result appears to be in conflict with the Einstein-de Sitter model, the density coming out to 3×10^{-31} gm cm^{-3} if a kinetic temperature of 1°K is assumed. We have also attempted to measure the intergalactic hydrogen in emission, by looking for a discontinuity in the continuum spectrum at the H-line frequency. This gives an upper limit for the density of H I of 3×10^{-29} gm cm^{-3}, and is independent of assuming a kinetic temperature. It should be possible to improve this limit by a factor of ten.'

Professor McVittie: 'I have always understood that in the Einstein-de Sitter universe the pressure is zero by definition. If you introduce intergalactic gas at a finite temperature, you must be satisfied that the equations defining the model do not lead to negative pressure, or some other horror at some stage in the evolution.

Dr. Margaret Burbidge: 'What is the present state of the knowledge of hydrogen line emission from extragalactic radio sources?'

Dr. R. D. Davies: 'I think it is now clear that the radiation field near most sources is so high that all the hydrogen will be excited to the upper state, and no absorption will occur. H-line absorption might be observed in sources of very low surface brightness.'

REFERENCE

Hoyle, F., and Narlikar, J. V., 1961, *M.N.R.A.S.*, **123**, 133.

The Angular Diameters and Luminosity Function of Extragalactic Radio Sources

H. P. Palmer

In a recent paper Ryle and Clarke (1961) describe four distinct methods of deriving the luminosity function of extragalactic radio sources. The diagrams of Fig. 20/1*a* and 1*b* show the two extreme functions discussed by them. Fig. 20/1*a* shows

the limiting function which is implied by the observed isotropic distribution of the sources and the low measured value of the residual radio temperature of the sky after the radiation from our Galaxy has been subtracted. The sources must have an absolute luminosity $P > 10^{24}$ watts (c/s)$^{-1}$ster^{-1}. If the median value of P were lower than that, the spatial density of sources would be higher, and the background temperature would be greater than observed. This argument therefore defines the minimum possible values of P. The functions which attribute the

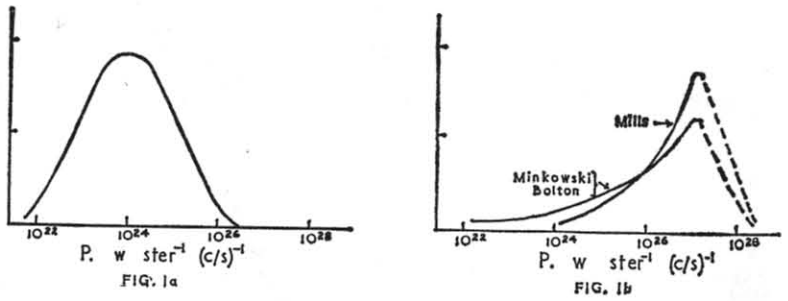


FIG. 20/1.

(*a*) The limiting distribution of the luminosities of extra-galactic radio sources, as derived by Ryle and Clarke.
(*b*) The distributions derived by Bolton and Minkowski, and by Mills from consideration of the optically identified sources.

highest values of luminosity to the sources have been derived by Bolton and Minkowski (1960) and by Mills (1960) from a study of optically identified sources and are reproduced in Fig. 20/1*b*. They indicate a median value of about 10^{27} watts (c/s)$^{-1}$ster^{-1}. Ryle and Clarke conclude that the sources have values of P in the range 10^{24}–10^{27}, and they tend to favour the higher values.

Measurements of the angular diameters of radio sources can also be used to derive a luminosity function, provided information is available about the angular sizes of all, or almost all, the sources in a given flux range. The measurements we have made and analysed in this way refer to 133 sources in the 3C catalogue with flux $S \geqslant 12$ f.u. and galactic latitude $b^{\text{I}} \geqslant 12°$. I described in Chapter 14 the method of making these measurements, and extracting from them the angular size of the radiating regions of the sources. A luminosity function can be derived if it is then assumed that these regions are of constant linear size, the value used being the average value of the linear sizes of 14 identified

sources, for which optical measurements of red-shift are available. The exact value depends slightly on the method of analysing the diameter data; but it is close to 25 kpc, and this value has been used in our calculations. The results are shown in Fig. 20/2, where each point represents a source in this group of

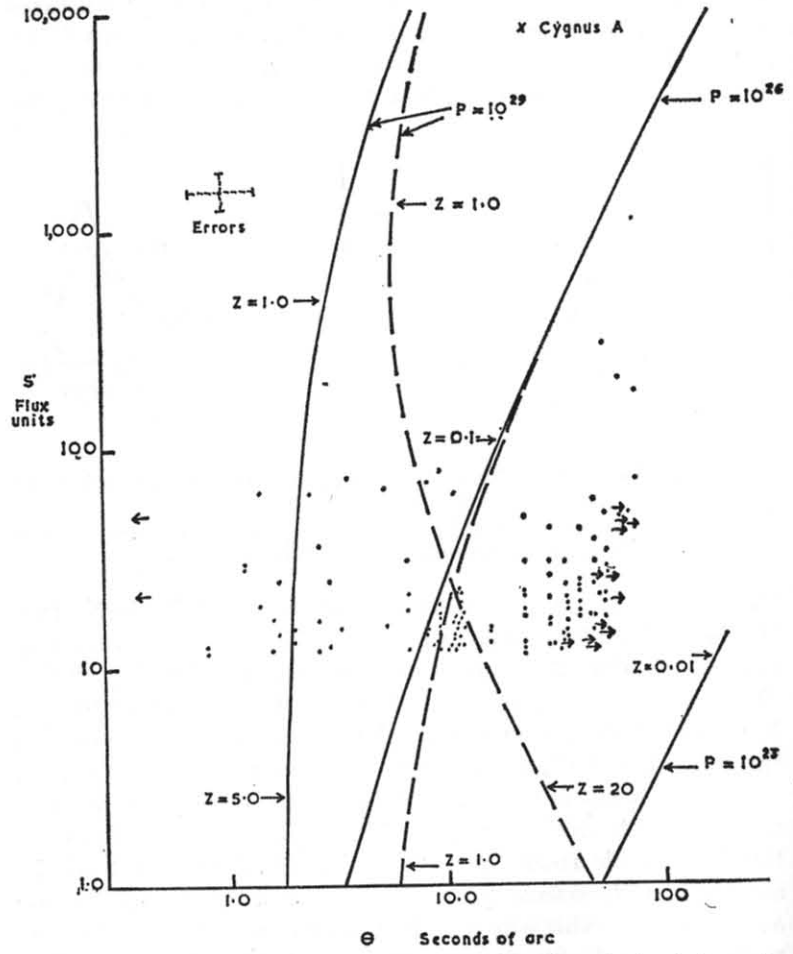
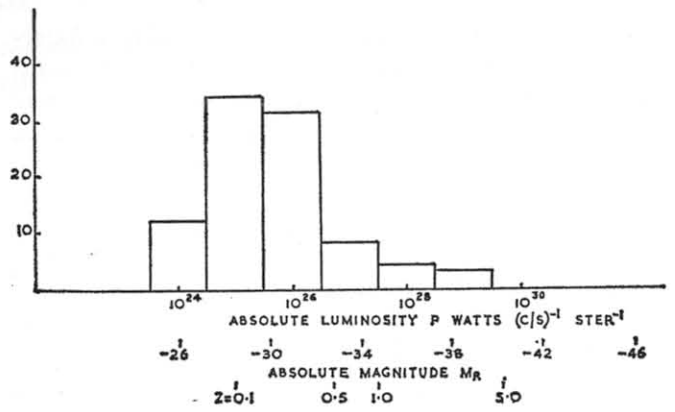


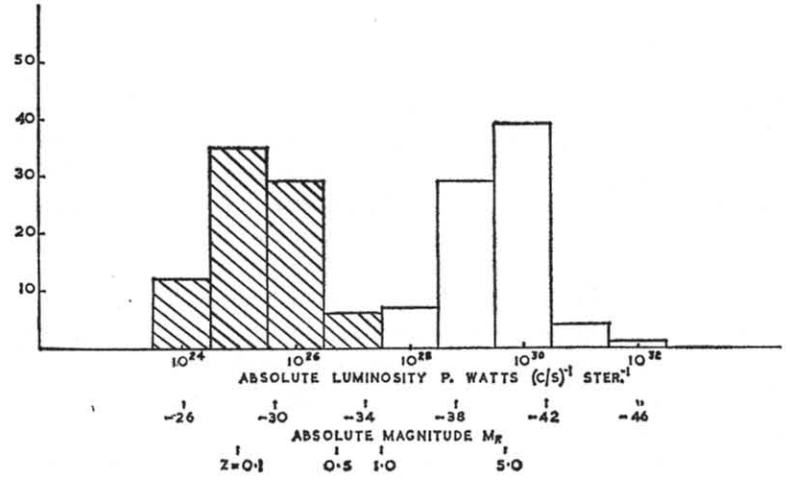
FIG. 20/2.—The observed component angular sizes derived from the Jodrell Bank data, plotted against the values of flux density S from the 3C catalogue, for 133 sources having $S > 12$ f.u. and $b^{1} > 12°$. The full lines show the calculated relations for the steady state cosmology for three values of absolute luminosity P assuming the linear size of the sources to be 25 kpc. The dashed lines give the equivalent relations for the Einstein-de Sitter cosmology.

133, the component angular diameter being plotted on a logarithmic scale horizontally, and the flux, as listed in the 3C catalogue, on the vertical scale. The approximate size of the errors in flux and angular diameter associated with these points is also shown near the vertical axis. The symbol X represents the position on this diagram for Cygnus A ($P = 10^{27\cdot5}$) which is not included in this group of 133 as it is near the galactic equator. All the sources included in this analysis had a value of flux $S \geqslant 12$ f.u., and this observational cut off sets the horizontal base to the array of dots. The full lines show where the points would fall on this diagram according to the theory of the steady-state universe, for values of $P = 10^{29}$, 10^{26} and 10^{23} watts(c/s)$^{-1}$ster^{-1}. The angular diameters, for the highest value of P, tend to an asymptotic limit of 1·4″, while the lowest value of P gives a straight line. In a static Euclidean universe, the corresponding lines would be straight, parallel to that for $P = 10^{23}$ watts (c/s)$^{-1}$ster^{-1}. The dotted lines show the results of similar calculations for an Einstein-de Sitter universe, in which, with increasing values of red-shift z, the angular diameters decrease to a minimum value until the 'equator' is reached at $z = 1{\cdot}25$, beyond which they increase again. There is no obvious relationship between the points and the lines in this figure, even when one takes account of the errors (as indicated) in the positions of the points. It follows that in the real universe there are wide dispersions in the absolute power, and/or the linear size for the radio sources represented by points on this diagram.

The same information can be replotted (Fig. 20/3) as a luminosity function of the same form as Fig. 20/1 (still assuming the linear size of each source to be 25 kpc). The histogram in Fig. 20/3*a* shows the function for a steady-state universe, while the shaded area of Fig. 20/3*b* is the function for an Einstein-de Sitter universe in which all the sources are assumed to be this side of the equator. About 90 of the 133 sources appear on these diagrams. There are also about 30 sources bigger than 47″ for which P is less than 10^{24} watts (c/s)$^{-1}$ster^{-1}, which would appear, therefore, if they could be plotted, on the left of the histograms. As only a small number of sources have values of $P > 10^{26}$ the histograms are almost identical in either cosmological model. The clear area of the histogram of Fig. 20/3*b* shows the luminosity function calculated if it is assumed that all the sources are beyond the equator of the Einstein-de Sitter



(*a*)



(*b*)

FIG. 20/3.—The distribution of absolute luminosities P inferred for sources of flux density $S \geqslant 12$ f.u. and $b^{1} \geqslant 12°$.

(*a*) Assuming a steady-state model universe.
(*b*) Assuming an Einstein-de Sitter model universe, in which there are two values of the absolute luminosity P associated with each observed component angular diameter greater than 5·5″. The absolute luminosities calculated from the lower value ($z < 1{\cdot}25$) are plotted in the shaded area of the diagram and the absolute luminosities calculated for the higher value ($z > 1{\cdot}25$) are plotted in the unshaded part. The scale of red-shift z refers to sources having the limiting flux density $S = 12$ f.u.

universe. It will be seen that this interpretation leads to values of P which are very much larger than $10^{27{\cdot}5}$, the value of Cygnus A, and correspondingly enormous values of red-shift $z > 50$.

As this seems a rather extreme conclusion to draw from the evidence available, I tend to support the shaded histogram of Fig. 20/3.

Discussion

Professor Ryle emphasized that even for a value of $P \simeq 10^{24}$, his observations, going down to 2 flux units, extended out to distances of 500 Mpc ($z = 0{\cdot}12$) where discussions of 'local' clustering were scarcely relevant. If the mean absolute photographic magnitude of the sources was $M_p = -20{\cdot}5$, he could only ask 'where are the identifications?' If M_p were taken as -17 for the majority of sources, for instance, he wondered whether the sources identified so far might not be atypical in absolute diameter also, and if so how was one to use angular measurements in cosmological studies? The other interpretation, that the majority of the sources were beyond the equator in an evolutionary universe, avoided both these difficulties. Dr. Burbidge pointed out that, in that case, Cygnus A would be a rather weak and insignificant source, and perhaps the fraction identified was already too large. He wondered whether the Cygnus identification was incorrect, and referred to the success of the very recent attempts at Cal. Tech. and Lick, to identify some of the more intense sources with stars. Professor Ryle emphasized again that the arguments based on background temperature showed that only a very few per cent of sources could be of this type.

REFERENCES

Bolton, J. G., 1960, 13th General Assembly, U.R.S.I., London.
Mills, B. Y., 1960, *Aust. J. Phys.*, **13**, 550.
Minkowski, R., 1960, Proceedings of the 4th Symposium on Mathematical Statistics and Probability, Berkeley.
Ryle, M., and Clarke, R. W., 1961, *M.N.*, **122**, 349.

Participants in the Summer School at Jodrell Bank.

Summer School students gaining practical experience in the analysis of records.

A 'Brains Trust' at the Summer School: R. C. Jennison, G. Burbidge, R. D. Davies, M. I. Large, H. P. Palmer and Professor Sir Bernard Lovell.

INDEX

RADIO ASTRONOMY TODAY