As you are now the owner of this document which should have come to you for free, please consider making a donation of £1 or more for the upkeep of the (Radar) website which holds this document. I give my time for free, but it costs me in excess of £300 a year to bring these documents to you. You can donate here https://blunham.com/Radar

Please do not upload this copyright pdf document to any other website. Breach of copyright may result in a criminal conviction.

This document was generated by me Colin Hinson from a document held at Henlow Signals Museum. It is presented here (for free) and this version of the document is my copyright (along with the Signals Museum) in much the same way as a photograph would be. Be aware that breach of copyright can result in a criminal record.

The document should have been downloaded from my website https://blunham.com/Radar, if you downloaded it from elsewhere, please let me know (particularly if you were charged for it). You can contact me via my Genuki email page: https://www.genuki.org.uk/big/eng/YKS/various?recipient=colin

You may not copy the file for onward transmission of the data nor attempt to make monetary gain by the use of these files. If you want someone else to have a copy of the file, please point them at the website (https://blunham.com/Radar).

Please do not point them at the file itself as the file may move or be updated.

I put a lot of time into producing these files which is why you are met with this page when you open the file.

In order to generate this file, I need to scan the pages, split the double pages and remove any edge marks such as punch holes, clean up the pages, set the relevant pages to be all the same size and alignment. I then run Omnipage (OCR) to generate the searchable text and then generate the pdf file.

Hopefully after all that, I end up with a presentable file. If you find missing pages, pages in the wrong order, anything else wrong with the file or simply want to make a comment, please drop me a line (see above).

It is my hope that you find the file of use to you personally – I know that I would have liked to have found some of these files years ago – they would have saved me a lot of time !

Colin Hinson
In the village of Blunham, Bedfordshire.

RRE JOURNAL

APRIL 1958

R.R.E. JOURNAL APRIL 1958

R. R. E. JOURNAL

CONTENTS

Issued by the Royal Radar Establishment, Ministry of Supply

No. 41 April 1958

ON THE NUMBER OF 0's FOLLOWED BY 0 IN A RANDOM BINARY SEQUENCE

by P. M. Woodward

1. Introduction

In the course of testing a programme for generating a sequence of "artificially random" binary digits, it was noticed that the standard deviation of the number of 0's followed in the sequence by 0 was more than twice as great as the deviation of the number of 0's followed by 1. The present paper explains this theoretically, and shows that the asymptotic ratio of the two standard deviations (in a long sequence) is $\sqrt{5}$.

2. 0's followed by 1 and 1's followed by 0

Let the total length of the sequence be N digits, so that the average number of 0's followed by 1 is

$$\bar{n}_{01} = \frac{1}{4}N .$$

The average number of changes, from 0 to 1 and 1 to 0, is N/2. These changes are alternately of the first and second kind, so that apart from end-effects, we have

$$n_{01} = n_{10} .$$

The distribution of changes is binomial with p = 0.5, whence

$$\bar{n}_{ch} = Np = \frac{1}{2}N$$

$$\sigma^2_{ch} = Np\,(1 - p) = \frac{1}{4}N.$$

Thus

$$\sigma_{01}^2 = \frac{1}{16}N$$

or

$$\sigma_{01} = \frac{1}{4}\sqrt{N}.$$

This is the first result, and is elementary.

3. 0's followed by 0 and 1's followed by 1

The expected numbers of 0's followed by 0 and 1's followed by 1 are each N/4, and the actual numbers in any one experiment satisfy (apart from end-effects) the relation

$$n_{00} + 2n_{01} + n_{11} = N .$$

Our problem is to determine the variance

$$\sigma_{00}^2 = \sigma_{11}^2$$

which would be obtained if counts were made in a large number of different sequences, each of length N.

First we shall represent the sequence of 0's and 1's in terms of a sequence of a's, b's, c's and d's representing the transitions

$$a = 00$$

$$b = 01$$

$$c = 10$$

$$d = 11$$

Thus for example

... 00101110110001010110O ...

would become

... abcbddcbdcaabcbcbdca ...

Of the sixteen possible pairs of letters, eight are forbidden transitions (ac, ad, cc, cd, ba, bb, da, db). This new ergodic process can be represented as a graph, as shown in Figure 1.

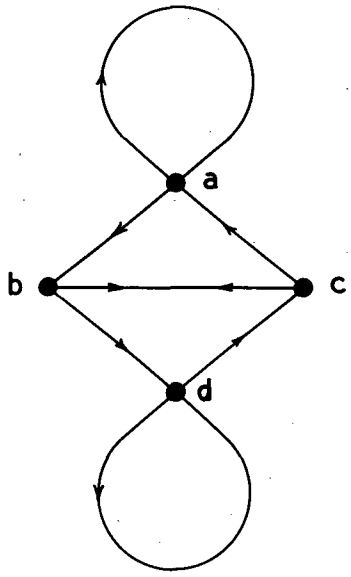

Figure 1 Graph of the ergodic process

Instead of attempting at first to find the distribution of numbers of a's in a long sequence, we shall start by investigating the distribution of the number of symbols which intervene between two successive a's. (The desired result can then be readily obtained, provided that we confine our final attention to large values of N.) Thus we require to find the probability p(r) that, starting from an a, the next a will be reached after r "steps". For example, taking r = 6, we can have

abcbdca, abdcbca, abdddca

and these are all the possibilities. Each has probability 1/64, whence

$$p(6) = 3/64 .$$

Before considering a general value of r, we may also note that

$$p(1) = 1/2 .$$

This corresponds to the sequence aa.

By studying Figure 1, we see that between any two successive a's, all possible routes can be broken down into a succession of smaller ones chosen from

bc, bdc, bddc, bdddc, etc.

Thus, to find the number of routes for r = 6, we have to enumerate all the ordered partitions of the number 5, excluding terms less than 2. In this case we have the three partitions

$$2 + 3, \quad 3 + 2, \quad 5$$

which lead to the sequences already given for illustration. To find the number of partitions of any integer, consider the generating function

$$(x^2 + x^3 + x^4 + \ldots) + (x^2 + x^3 + x^4 + \ldots)^2 + \ldots.$$

The coefficient of x^n in the first term gives the number of partitions into one part, in the second the number of partitions into two parts, and so on. To obtain a general expression for the total coefficient of x^n, we note that the generating function may also be written

$$(x^2 + x^3 + \ldots)\left[1 + (x^2 + x^3 + \ldots) + (x^2 + x^3 + \ldots)^2 + \ldots.\right].$$

By comparing coefficients in the two forms, we find that

$$c_n = 1 + c_{n-2} + c_{n-3} + \ldots. + c_2$$

and similarly

$$c_{n-1} = 1 + c_{n-3} + \ldots + c_2 .$$

Subtracting, we obtain the Fibonacchi recurrence relation

$$c_n = c_{n-1} + c_{n-2}$$

having the general solution

$$c_n = A\left(\frac{1 + \sqrt{5}}{2}\right)^n + B\left(\frac{1 - \sqrt{5}}{2}\right)^n .$$

Using the fact that $c_2 = 1$, $c_3 = 1$, we may simply verify that the arbitrary constants are given by

$$A = \frac{1}{2}\left(1 - \frac{1}{\sqrt{5}}\right)$$

$$B = \frac{1}{2}\left(1 + \frac{1}{\sqrt{5}}\right).$$

The number of partitions of n is therefore

$$c_n = \frac{1}{\sqrt{5}} \left(\frac{1 + \sqrt{5}}{2} \right)^{n-1} - \frac{1}{\sqrt{5}} \left(\frac{1 - \sqrt{5}}{2} \right)^{n-1}$$

and finally, the probability of r steps from one a to the next is

$$p(r) = c_{r-1} / 2^r .$$

We may observe that this relation holds good for r = 1, representing the sequence aa, since $c_0 = 1$. Further, we find from the expression for c_n that $c_1 = 0$, which represents the impossibility of any route from a to a in two steps on Figure 1.

The next stage in the calculation of σ_{00}^2 is to calculate the variance of r, viz.

$$\sigma_r^2 = \text{Av}\,(r^2) - (\text{Av}\, r)^2$$

$$= \sum_1^\infty r^2 \, p(r) - 16 .$$

The 16 derives from the fact that a occurs one quarter of the time and the mean number of steps from one a to the next is thus 4.

The required sum may be written in the form

$$\sum_1^\infty r^2 \, p(r) = \frac{1}{4\sqrt{5}} (S_+ - S_-)$$

where

$$S_+ = \sum_1^\infty r^2 \left(\frac{1 + \sqrt{5}}{4} \right)^{r-2}$$

and S_- is defined similarly, except for the sign of the square root. Denoting $(1 + \sqrt{5})/4$ by x, we have

$$S_+ = \sum_1^\infty r^2 \, x^{r-2}$$

which is easily summed, and gives

$$S_+ = \frac{1 + x}{x \, (1 - x)^3} .$$

Substituting the numerical value for x, we find

$$\frac{1}{4\sqrt{5}}(S_+ - S_-) = \text{Av}(r^2) = 36 .$$

Subtracting the square of the mean, we finally obtain

$$\sigma_r^2 = 20 .$$

Since each interval between successive a's is independent of the rest, the variance of the length of a sequence containing exactly n a's is simply 20n. The mean length of the sequence is of course 4n. Thus n a's are likely to overfill or underfill a sequence of exactly 4n digits by about $2\sqrt{(5n)}$ digits, this being the standard deviation. We now have in a sense to invert our result, and find out how many a's too many or too few we should expect to find in a sequence of a given number of digits.

If n a's occupy $4n \pm 2\sqrt{(5n)}$ digits, N/4 a's occupy $N \pm \sqrt{(5N)}$ digits, and the $\sqrt{(5N)}$ digits can be expected to contain $\frac{1}{4}\sqrt{(5N)}$ a's. Thus N digits will contain a mean number of N/4 a's, with a standard deviation

$$\sigma_{00} = \sigma_{11} = \frac{1}{4}\sqrt{(5N)}$$

for large values of N. This is $\sqrt{5}$ times larger than the value of σ_{01} or σ_{10} obtained in section 2, and completes the proof.

THE PHYSICS OF THE SOLID STATE MASER

by D. J. Howarth

1. Introduction

The Maser is a name given by Townes and co-workers[1] to describe a device for Microwave Amplification through Stimulated Emission of Radiation. In any application of an amplifier to a particular microwave circuit, interest attaches to performance figures such as gain, bandwidth and, particularly, the noise figure. We shall comment briefly on these properties in the present article, in order to demonstrate the value of such a device; in particular, we shall show that very low noise figures are theoretically obtainable. The major part of the article, however, will aim at indicating the physical principles underlying the operation of a Maser, particular attention being focussed on the three level solid state Maser. These principles are generally unfamiliar to electronic engineers, and study of recent literature on solid state Masers can be very confusing unless there is a clear understanding of the underlying physics and of the meaning of the expressions used.

We shall therefore start by outlining the general theory of emission and radiation processes, particularly as they apply to the microwave spectrum. These processes are generally described in terms of wave mechanics, and a brief account of the relevant portions of this theory and the investigation thereby of transitions between energy levels is given in Section 3.

Various suggested methods of obtaining an emissive state are described in Section 5, and of these only the three level Maser will be considered in detail. For use in such a Maser, the energy levels of paramagnetic ions in crystals prove suitable, and we give an account of the relevant properties of such substances in Section 4. The three level Maser will then be considered in detail, and operating characteristics are obtained in Section 6. We can see quite clearly from this work the necessary conditions to obtain Maser action; the calculation from

this data of possible gain bandwidth and noise figures obtainable in an amplifier is briefly reviewed in Section 7, and the microwave requirements for efficient action are outlined.

It should be emphasised that the solid state Maser is still in an experimental stage. At the time of writing, three Masers are known to be in existence, one of which has been operated as a stable C.W. amplifier. The existing achievements, and the work in progress at R.R.E., are described in Section 8.

2. Emission and Absorption of Radiation

In an isolated atom or molecule, it is well known that the electrons occupy discrete energy levels, or quantum states. The same is approximately true when the atom is not completely isolated, but interacts with its neighbours sufficiently strongly to be in thermal equilibrium. In fact, these levels will be broadened into bands due to this interaction, but we shall here assume this broadening to be small.

Suppose there are two levels, 1 and 2, such that the atom has energies E_1 and E_2. An atom in level 1 can absorb radiation and be excited to level 2, the frequency of the absorbed radiation, ν_{12}, being given by

$$h\nu_{12} = E_2 - E_1 \qquad (2.1)$$

where h is Planck's constant. Suppose now we have a large number of atoms. Then in thermal equilibrium, the number N_2 in level 2 is related to the number N_1 in level 1 by the Boltzmann Law

$$N_2 = N_1 \exp(-h\nu_{12}/kT) \qquad (2.2)$$

where k is the Boltzmann constant. Interest lies in absorption of radiation in the microwave region, wavelengths 1 to 10 cm. At these wavelengths, insertion of numerical values shows that

$$\frac{h\nu_{12}}{kT} < 1$$

even at low temperatures, and we then find that

$$\frac{N_1 - N_2}{N_a} = \frac{h\nu}{2kT} \qquad (2.3)$$

when $N_a = N_1 + N_2$ is the total number of atoms. We thus see that, when the two energy levels are separated by energies corresponding to the

microwave range of frequencies, there is a small excess of population in the lower energy level.

We shall now consider how these populations are affected by the emission and absorption of radiation at the frequency ν_{12}. The population N_2 of level 2 can obviously be augmented from level 1 by absorption of radiation, this effect being proportional to N_1 and to the intensity of the radiation field, i.e.

$$\frac{dN_2}{dt} = B\rho(\nu)\, N_1$$

where $\rho(\nu)$ is the energy density of the radiation field in the frequency interval $d\nu$. N_2 can also be decreased by spontaneous emission of radiation, a process not dependent upon the presence of a radiation field, i.e.

$$\frac{dN_2}{dt} = -AN_2$$

There is a third, not so obvious, process which can alter N_2. Einstein showed that in the presence of a radiation field, the probability of a transition from the upper to the lower level is increased by an amount proportional to the field, i.e.

$$\frac{dN_2}{dt} = -C\,\rho(\nu)\, N_2$$

This process is known as 'stimulated emission', and is of prime importance in understanding the principles of Maser action. The constants A, B and C are essentially transition probabilities, which we may consider to be time independent.

Combining these three processes, we have

$$\frac{dN_2}{dt} = B\,\rho(\nu)\, N_1 - (A + C\rho(\nu))\, N_2 \qquad (2.4)$$

Now in thermal equilibrium, N_2 must be constant, and we have already seen that N_1 and N_2 are related by the Boltzmann law (2.2). Thus, inserting these conditions, we see that the energy density of the radiation field is

$$\rho_o(\nu) = \frac{A}{B \exp(h\nu/kT) - C} \qquad (2.5)$$

But $\rho_o(\nu)$ is given by the well known Planck formula

$$\rho_o(\nu) = \frac{8\pi h^3}{c^3} \frac{1}{\exp(h\nu/kT) - 1} \qquad (2.6)$$

Hence, equating (2.5) and (2.6), we find that B = C, that is, the transition probability for stimulated emission is equal to that for absorption. We shall later see that this result can also be derived from a wave mechanical approach. We also see that

$$A/B = 8\pi h\nu^3 / c^3$$

Thus, in thermal equilibrium, the ratio of the probability of spontaneous emission to that of stimulated emission is

$$\frac{A}{B\rho_o(\nu)} = \exp(h\nu/kT) - 1 = h\nu/kT \quad \text{(approx.)} \qquad (2.7)$$

The approximation holds good in the microwave region, where $h\nu/kT$ is small.

We then see that in the microwave region we may neglect spontaneous emission and consider merely the stimulated emission and absorption in the radiation field. This is an important difference between the microwave and optical region of the spectrum. In the latter, $h\nu/kT$ is large, and spontaneous emission will predominate. In the microwave region, we see that almost all the transitions depend directly on the radiation field, the spontaneous emission contributing a small noise factor. Henceforth, spontaneous emission will be neglected until we come to study the noise figures of Masers.

2.1 The Emissive Condition

One simple fact emerges immediately from the above formulae. Neglecting A and putting B = C in (2.4) gives

$$\frac{dN_2}{dt} = B\rho(\nu)(N_1 - N_2) \qquad (2.8)$$

Thus, if there are more atoms in state 1 than state 2, N will increase, the system absorbing radiation. This is the situation occurring in undisturbed thermal equilibrium when equation (2.3) holds. If, however, N_2 is greater than N_1, the system will emit radiation, thus giving amplification of the incident radiation. This condition can never obtain in thermodynamic equilibrium; the problem of designing an amplifier using the stimulated emission process is the problem of

upsetting the thermodynamic equilibrium and obtaining an emissive state, with N_2 greater than N_1. Another way of describing this condition is by the concept of 'effective temperature'. We can regard the 'effective temperature' of the system as being defined by equation (2.2).

In thermal equilibrium, this effective temperature is obviously the actual temperature of the system. The condition for emission may now be expressed as the existence of a negative effective temperature. The process of obtaining this condition is frequently referred to as 'inversion', a term which should be used with some care. Several suggested methods for obtaining an emissive state consist in 'inverting' the levels; the method to be described here, however, achieves an emissive state by alteration of the equilibrium population of the levels, and the term 'inversion' does not truly describe this process.

2.2 Saturation

Suppose we consider a radiation field applied to a system in thermal equilibrium with N_1 greater than N_2. Absorption occurs, and equation (2.8) shows that N_2 will increase and N_1 decrease; at first sight it would appear that after a certain time, the levels would be equally populated, and absorption would cease. This condition is called saturation. However, there must exist processes such as collisions etc. tending to restore thermodynamic equilibrium. These may be regarded as having a relaxation time T_1 such that the disturbed system would decay exponentially back to equilibrium. Thus,

$$\frac{d}{dt}(N_1 - N_2) = -\frac{N_1 - N_2 - (N_{10} - N_{20})}{T_1} \qquad (2.9)$$

where N_{10} and N_{20} are the populations in thermal equilibrium. Equilibrium will occur when this process balances out the process described by equation (2.8), and the equilibrium values of N_1 and N_2 are easily found to be

$$N_1 = \frac{B\rho(\nu)N_a + N_{10}/2T_1}{2B\rho(\nu) + \frac{1}{2}T_1}$$

$$N_2 = \frac{B\rho(\nu)N_a + N_{20}/2T_1}{2B\rho(\nu) + \frac{1}{2}T_1}$$

where $N_a = N_1 + N_2$. Thus if T_1 is short, and $B\rho(\nu)$ small (i.e. a low power radiation field is incident, and the transition probability is small) the equilibrium populations are only very slightly altered from the populations in thermal equilibrium. If, however, the incident power is increased, or T_1 is raised, the populations tend to the

saturated values $\frac{1}{2}N_a$. We shall see that this process of saturation is of basic importance in the particular Maser we shall be considering. It is important to note, however, that even with large radiation fields, we can never do more than saturate - we cannot achieve directly an emissive condition.

3. Quantum Mechanical Description of Transition Probabilities

The energy levels and transition probabilities B of a system are calculated by application of quantum mechanics. The results are generally quoted without proof, and to readers unfamiliar with the ideas and notation of quantum mechanics, confusion can arise at just this point. We shall here give a very brief (and not very rigorous) outline of the necessary quantum mechanical ideas, and as an example, shall derive a formula for the transition probability B.

In quantum mechanics, the properties of a system, an electron, an atom, or a molecule, are characterised by a wave function $U(r,t)$. This is defined as a solution of the time dependent Schrödinger equation

$$H_o U = -\frac{ih}{2\pi}\frac{\partial U}{\partial t} \qquad (3.1)$$

H_o is an operator known as the Hamiltonian. It is directly related to the Hamiltonian of classical dynamics, and is formed by writing down the classical energy of the system, and replacing the co-ordinates by operators according to certain simple rules. It represents the sum of kinetic and potential energies of the system.

If we now consider a stationary system, where H_o is independent of time, a solution of (3.1) is

$$U = \sum_n c_n \psi_n(\underline{r}) \exp\left(-\frac{2\pi i}{h} E_n t\right) \qquad (3.2)$$

where

$$H_o\psi_n = E_n\psi_n \qquad (3.3)$$

Here the c_n are arbitrary constants, (3.3) is the time independent Schrödinger equation, and E_n is the energy of the stationary state ψ_n. It is a familiar 'eigenvalue' equation, and in general there are a set of discrete energies E_n for which solutions ψ_n are possible. The sum in (3.2) is taken over all such states. The eigenfunctions ψ_n form an orthonormal set, i.e.

$$\int \psi_m(\underline{r})\,\psi_n(\underline{r})\,d\underline{r} = \begin{cases} 0, & m \neq n \\ 1, & m = n \end{cases} \qquad (3.4)$$

The wave function U has the property that the probability of finding the particle in volume $d\underline{r}$ is

$$U^{*}(\underline{r},t)\, U(\underline{r},t)\, d\underline{r}$$

where * represents the complex conjugate. Integrating over all space must give unity, and using (3.2), this shows that

$$\sum_n |c_n|^2 = 1$$

$|c_n|^2$ is in fact the probability that the system is in station state ψ_n.

Suppose now we have such a system defined by the Hamiltonian H_o, and we apply a time dependent perturbation to the system, giving rise to a potential energy $V(\underline{r},t)$. We now seek solutions of the Schrödinger equation

$$(H_o + V)U = -\frac{ih}{2\pi}\frac{\partial U}{\partial t} \qquad (3.5)$$

and by analogy with (3.2), we shall write

$$U = \sum_n c_n(t)\, \psi_n(\underline{r})\, \exp(-2\pi i E_n t/h) \qquad (3.6)$$

Substituting into (3.5), multiplying by ψ_k^{*} and integrating, we find that

$$\frac{hi}{2\pi}\frac{dc_n}{dt} = \sum_k c_k(t)\, V_{nk}(t)\, \exp(2\pi i \nu_{nk} t) \qquad (3.7)$$

where $h\nu_{nk} = E_n - E_k$

and $$V_{nk} = \int \psi_n^{*}\, V\, \psi_k\, d\underline{r} \qquad (3.8)$$

Now suppose that at time $t = 0$, the system is in state ψ_m. Then $|c_n(t)|^2$ will give the probability that the system is in state ψ_n at time t, and hence gives the transition probability. If this probability is small, the dominant term in (3.7) is when $k = m$, and so

$$c_n(t) = \frac{2\pi}{ih}\int_o^t V_{nm}(\underline{r},t)\, \exp(2\pi i \nu_{nm} t) \qquad (3.9)$$

Now suppose the perturbations to arise from an oscillating magnetic field, say $\underline{H} = (H\cos 2\pi\nu t, 0, 0)$. The potential energy due to this is

$$V = -\underline{\mu}.\underline{H}$$

where $\underline{\mu}$ is the magnetic moment of the system. (In the quantum theory, $\underline{\mu}$ is an operator). Thus

$$V_{nm} = -H \cos(2\pi\nu t).(\mu_{nm})_x \qquad (3.10)$$

where

$$(\mu_{nm})_x = \int \psi_n^{*} \mu_x \psi_m \, d\underline{r} \qquad (3.11)$$

Terms such as μ_{nm} form a two-dimensional array of matrix elements and obey the laws of matrix algebra. In particular, we can show that the matrix is Hermitian, which means that we have $(\mu_{nm})^{*} = \mu_{nm}$.

Combining (3.10) with (3.7) gives a strong resonance term when ν is approximately equal to ν_{nm} corresponding to absorption of radiation, and in fact

$$c_n(t) = \frac{iH\pi}{h}(\mu_{nm})_x \frac{\exp(2\pi i(\nu_{nm}-\nu)t) - 1}{2\pi i(\nu_{nm}-\nu)} \qquad (3.12)$$

We have so far assumed an idealised case of discrete stationary states ψ_n. For an electron in a solid, when there are interactions with the surrounding electrons and with the lattice, the resonant frequencies are spread over a band, which can be made narrow by reducing the interactions. This situation can be described by a distribution function $g(\nu)$ such that there are $g(\nu)d\nu$ states with resonant frequency in the range $d\nu$. $g(\nu)$ is generally called the line shape, and is normalised so that

$$\int_0^{\infty} g(\nu)\, d\nu = 1$$

The probability that the system is in state ψ_n after time t is now

$$P_{nm}(t) = \int_0^{\infty} |c_n(t)|^2 \, g(\nu_{nm}) \, d\nu$$

and to a close approximation, this reduces to

$$\frac{P_{nm}(t)}{t} = \frac{\pi^2 H^2}{h^2} |(\mu_{nm})_x|^2 \, g(\nu_{nm}) \qquad (3.12)$$

A Lorenz line shape is generally assumed, in which case $g(\nu) = 2T_2$, T_2 being inversely proportional to the line width, and being the transverse relaxation time. Equation (3.12) gives the transition probability per unit time, which we have previously denoted by B. An immediate consequence of the Hermitian property of μ_{rm} is that

$$P_{nm}(t) = P_{mn}(t)$$

as was deduced in Section 2. The transition probability is proportional to $|(\mu_{nm})_x|^2$. If this should be zero for any pair (n,m), the transition is non-allowed. In simple cases, most of the transitions fall into this class.

The magnetic moment of an electron arises from two sources, the orbital motion of the electric charge, and the intrinsic magnetic moment resulting from the spin. We shall now consider what happens when electrons in paramagnetic ions are placed in a crystal lattice, and shall consider in a simple case the values of $|(\mu_{nm})_x|^2$ and the corresponding 'allowed' transition.

4. Theory of the Energy Levels of Paramagnetic Crystals

To determine the properties of a paramagnetic ion in a crystal, we set up a Hamiltonian representing the total energy, and solve the Schrödinger equation (3.3). Except in simple cases, exact solutions of (3.3) cannot be obtained, so we proceed from a simple case and add on terms representing different interactions in descending order of magnitude, treating them as independent effects. If we omit magnetic interaction with the nucleus, the terms in the Hamiltonian may be written as [2,3]

$$H = H_f + V + H_{LS} + H_H \qquad (4.1)$$

Here

H_f is the energy of the free ion

V is the effect of the crystal field

H_{LS} is the magnetic interaction of the spin and orbital magnetic moment

H_H is the interaction of an external DC magnetic field with the orbital and spin magnetic moments.

In the iron transition group, H_f is very large compared with V, which is in turn large compared with the last two terms. We shall therefore consider them in order.

(a) Energy of the free ion The theory of the energy levels of a many electron atom is well known, and can be described as follows:-

Direct solution of the Schrödinger equation for a single electron moving in a central field of force shows that the states are defined by quantum numbers, n, ℓ, m, all integers. The energy depends upon n and ℓ only, ℓ being the orbital quantum number. The state is then labelled by a letter corresponding to ℓ (s,p,d,f for $\ell = 0,1,2,3$) preceded by the value n. The outer electrons in the iron transition group are in 3d and 4s states.

The quantum number m can take values such that $|m| \leqslant \ell + 1$, and we find that the projection of the angular momentum along any axis has value $mh/2\pi$. There are thus $(2\ell + 1)$ levels (with different values of m) having the same energy.

In addition, the electron possesses a spin S, taking values $\pm \frac{1}{2}$, such that the spin angular momentum along any axis has value $Sh/2\pi$.

In a many electron atom, the angular momentum of the consistent electrons couple together to give a total orbital quantum number L and a total spin quantum number S. L takes integer values, S takes integer or half integer values. The state can then be described by the letter corresponding to L, and a number corresponding to S. There are many such states, corresponding to different ways of combining the individual angular momenta, and we select by use of Hunds rule the 'term' giving the lowest energy. The energy will be the same for the (2M) values of M (or L_z) and for the corresponding (2S + 1) values of S_z (the components of the angular momenta along any axis).

As an example, consider the ion Cr^{+++} with 3 electrons in the 3d state. The lowest 'term energy' is for the term 4F, which has $S = 3/2$, $L = 3$ and has 7-fold orbital degeneracy and 4-fold spin degeneracy.

(b) Crystal field The crystal field acting on a paramagnetic ion contains a large term of cubic symmetry and smaller terms of trigonal, tetragonal or rhombic symmetry. Electrostatic interaction between the orbital motion and the cubic field will be the largest term in V. This effect splits the (2L + 1) - fold degeneracy according to simple rules. For the Cr^{+++} ion, it leaves an 'orbital singlet' lowest, with a 4-fold spin degeneracy. The effect of the fields of lower symmetry is much smaller and is comparable with the effects of the other term in (4.1).

(c) Magnetic interaction of the spins Associated with the spin S of an ion there is a magnetic moment

$$-\frac{e}{mc}\underline{S}$$

Also associated with the angular momentum $\underline{L}$ there is a magnetic moment

$$-\frac{e}{2mc}\underline{L}$$

These two can interact with one another, leading to spin-orbit coupling proportional to $\underline{L}.\underline{S}$, given by H_{LS} in (4.1). When a DC magnetic field is applied, the potential energy of the system is given by

$$H_H = \frac{e}{2mc}\underline{H}.(\underline{L} + 2\underline{S}) = \beta\underline{H}.(\underline{L} + 2\underline{S})$$

where β is the Bohr Magneton. The crystal fields of lower symmetry interact with the orbits and affect the term H_{LS}.

In calculating the energy due to these terms in the Hamiltonian, we are helped by knowing the properties of the orbital part of the motion. It is possible, therefore, to carry out the calculation involving the $\underline{L}$ and we are left with the 'Spin Hamiltonian', which can be written in the form[4]

$$H = \sum_{i,j} S_i\, d_{ij}\, S_j + \beta \sum_{i,j} H_i\, g_{ij}\, S_j \qquad (4.2)$$

the sums being over the three co-ordinates x, y, z. Here d and g are complicated functions of the orbital wave functions. This function must reflect the overall symmetry of the problem. Suppose the crystal field has trigonal symmetry about an axis z. Then (4.2) must reduce to

$$H = DS_z^2 + g\beta\,\underline{H}.\underline{S}$$

omitting all terms such as $|\underline{S}|^2$, which is a constant, $S(S + 1)$, for all the $2S + 1$ levels. For mathematical reasons, this is generally written as

$$H = D\left[S_z^2 - \frac{1}{3}S(S + 1)\right] + g\beta\,\underline{H}.\underline{S} \qquad (4.3)$$

The entire properties of the lowest lying system of energy levels can thus be described by two parameters D and g. For a crystal field of lower symmetry, more parameters are necessary.

Our requirement is now to determine the properties of the system defined by this Hamiltonian. In the absence of the perturbation H,

there are $(2S + 1)$ states corresponding to the values of S_z, all with the same energy. Suppose the wave functions of these states are ψ_m where $m = S_z$. These functions can be shown to be ortho-normal for different m. It can be shown[5] that the effect of the spin operators S_x, S_y, S_z on these functions ψ_m are given, omitting the constant $h/2\pi$, by

$$S_z\psi_m = m\psi_m \tag{4.4}$$

$$(S_x \pm iS_y)\psi_m = \sqrt{S(S+1) - M(M \pm 1)}\ \psi_{m \pm 1} \tag{4.5}$$

Now suppose we look for a solution of (4.3) in the form

$$U = \Sigma c_m \psi_m \tag{4.6}$$

Then the constants c_m are given by the set of equations

$$\sum_m c_m H_{mn} = Ec_n \text{ for all } n \tag{4.7}$$

where

$$H_{mn} = \int \psi_m^* H \psi_n d\underline{r} \tag{4.8}$$

H_{mn} can be calculated from (4.5) using the orthnormal property of ψ_m. (4.7) have a non zero solution only if

$$|H_{mn} - E\delta_{mn}| = 0$$

where

$$\delta_{mn} = \begin{cases} 0, & m \neq n \\ 1, & m = n \end{cases}$$

This equation determines a set of $2S + 1$ values E, the energy levels, and (4.7) gives the corresponding wave functions.

This way of describing the resulting states is illuminating. Suppose in (4.3) H is taken along the z-axis. Then the only spin operator occurring is S_z, and by (4.5) we see that all H_{mn} are zero unless $m = n$. Thus from (4.7), the wave function for state n has only one term in (4.6) and is, when renormalised, simply ψ_n. If terms involving S_x and S_y are included in (4.3), by choosing H at an angle to the crystal axis or by inclusion of terms of lower symmetry, by virtue of (4.5), not all the H_{mn} $(m \neq n)$ will be zero, and the wave function of each state will be a mixture of ψ_m for different m.

Let us now consider transitions between the levels. We saw in (3.12) that these depend on the magnetic moment operator μ_x which here is simply

If the wave functions of the states were just ψ_m, we see from (4.5) that the only non-zero elements $|(\mu_{nm})_x|^2$ are

$$|\mu_{m\pm1,m}|^2 = \frac{g^2\beta^2}{4}(S(S+1) - M(M \pm 1)) \qquad (4.9)$$

If, however, mixture of states is allowed, other elements $|(\mu_{nm})_x|^2$ will be non-zero, and hence other transitions are allowed. In many cases, the contribution to U comes mainly from a single term in (4.6), say ψ_m, and we then refer to '$\Delta m = 1$ transitions' being allowed, according to (4.9), whilst under some conditions, we may find '$\Delta m = 2$' and higher transitions. When the states contain comparable contributions from several ψ_m, the distinction between allowed and non-allowed transitions becomes less clear.

5. Methods of obtaining an Emissive State

One method of obtaining a negative effective temperature is to remove altogether the atoms in the lower level. This method has been used in the ammonia Maser developed by Townes.[1] The ammonia molecule exhibits absorption at frequencies around 24 k Mc/s (K band). The transition involved is between two states having opposite signed electric dipole moments resulting in a neutral state when the atom oscillates between the two states. By passing a beam of ammonia molecules through an inhomogeneous electric field, the molecules in the lower state are deflected into the strong field region. Those in the upper state are collected in a cavity resonator where they drop to the lower energy, giving up energy to the field.

The gas Maser suffers a serious drawback in that the bandwidth obtainable is exceedingly small. The resonance in ammonia is only a few kc/s wide, and has, indeed, been used as a highly stable frequency standard. Broader bandwidths are only obtainable by using a higher density of atoms or molecules, thus increasing the interaction and broadening the energy levels. Attention thus turns to solid state devices.

Just as the ammonia molecule has two states with opposing electric dipole moments, so the states of certain solids can exhibit opposing magnetic moments, associated with the spin of the electrons. We have already discussed the properties of paramagnetic ions, and have seen that the energy levels are determined by the effects of the crystalline field and the Zeeman splitting due to the applied DC magnetic field. If we consider a 'free' electron, with no crystalline field splitting, and with a spin of $\frac{1}{2}$, we see that the doublet state is split by the DC

field, and the state with spin parallel to H has a higher energy than that with spin anti-parallel. If we now reverse the direction of H before the populations in the two levels have time to change, we reverse the positions of the levels and obtain an emissive state. This simple method has not so far proved practical, as the populations tend to follow t he change in direction of H.

Another method of achieving this result, the 'inversion' of the levels, is by the method of adiabatic fast passage. We have already seen that application of a strong signal at resonance can saturate two levels. If now the frequency is swept up from well below resonance through the resonance, it is found that the population of the two levels will be inverted. In a paramagnetic salt, instead of sweeping the frequency of the signal, we can equally well vary the resonant frequency of the salt by change of the DC field. The rate of change in field must be such as to accomplish the inversion in a time short compared with the spin lattice relaxation time but long enough for the electrons to follow the change, that is, long compared with the spin-spin relaxation time. This effect was first described by Bloch,[6] who applied it to nuclear spins, and has been considered further by Wittke.[7] The only known case in which it has been used in practice to obtain an emissive condition is by Combrisson, Honig and Townes,[8] who used as active material arsenic and phosphorous impurities in silicon.

A further method of obtaining inversion has been suggested, but not applied in practice. It is known as the '180° pulse' method, and a full understanding of it requires more detailed quantum mechanical treatment than we have attempted here. The theory of transition probabilities given in Section 3 breaks down when the signal power is very large. When this holds, an atom in level 1 will certainly be excited to level 2, and will then return to level 1 by stimulated emission. Thus the probability of an atom being in level 2 is a periodic function of time, and application of a pulse of the correct magnitude and length will achieve exact inversion of the two levels.

The above schemes for 'two level' Masers have been discussed in some detail by a number of authors (see Wittke,[7] Bolef and Chester[9]). We shall not consider them further here, except to state that they are necessarily pulsed devices, the amplification period being but a small fraction of the 'duty cycle', which in turn must be less than the spin-lattice relaxation time. The method to be described in detail is a C.W. method, and is the only one which has been used in practice to obtain solid state maser action. It is known as the 'three level' scheme, and was first suggested by Bloembergen.[10]

Consider a set of three energy levels as in Figure 1. In thermal equilibrium, the populations are such that

$N_1 > N_2 > N_3$

Suppose we now supply a strong radiation field at frequency ν_{13}, and saturate levels 1 and 3. It may then be shown quite easily that this gives an emissive state between either levels 3 or 2, or levels 2 or 1, and stilulated emission can be obtained at one of these frequencies. This system is generally described as 'pumping' at frequency ν_{13} and amplifying at frequency ν_{32} (or ν_{12}). We shall consider in some detail the properties of such a device, and shall see that certain stringent requirements are placed upon the energy level system. These requirements are met by some paramagnetic ions, and it is for this reason that we have outlined the properties of these ions in Section 4.

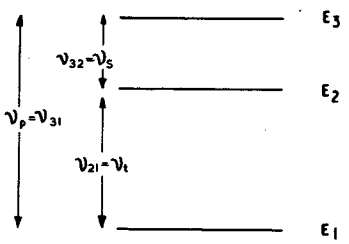

Figure 1
Energy levels of three level Maser

6. The Three Level Maser

A theory of the operation of the three level Maser has been given by Bloembergen,[10] and in more expanded form by Artman.[11] We shall obtain the necessary formulae here in order to show clearly the factors governing the choice of the paramagnetic crystal.

Consider a system of three energy levels, energies E_1, E_2 and E_3. We have seen that paramagnetic ions have $(2S + 1)$ low lying energy levels. We shall select three of these in the present discussion, and assume the other levels to be undisturbed by the radiation fields, and to remain in thermal equilibrium. This assumption simplifies the algebra, and we shall consider later the effect on the operation of a Maser.

Referring to Figure 1, we have already shown that emission is possible between, say, levels 3 and 2, when a pump field is applied across levels 1 and 3. The following discussion can be very easily modified to deal with emission between levels 2 and 1. We shall denote the frequency corresponding to transition 1 to 3 by ν_p, and that for transition 2 to 3 by ν_s, the pump and signal frequencies respectively. For convenience, we denote $\nu_p - \nu_s$ by ν_t.

We can now derive equations for the rate of change of populations in each of the three levels under the action of pump and signal fields and the relaxation processes. The former are proportional to the difference in population between the two levels involved, whilst the relaxation processes are proportional to the deviation of this difference from the value in thermal equilibrium. Denoting the

probability of a transition due to the applied radiation fields by W_p, W_S, and the probability of a transition due to relaxation processes by w_p, w_S, w_t, the rate of change of the populations n_1, n_2, n_3 are given by

$$\begin{aligned}\frac{dn_3}{dt} &= w_p\left[n_1 - n_3 - (n_{10} - n_{30})\right] \\ &\quad + w_S\left[n_2 - n_3 - (n_{20} - n_{30})\right] \\ &\quad + W_p(n_1 - n_3) + W_S(n_2 - n_3)\end{aligned}$$

$$\begin{aligned}\frac{dn_2}{dt} &= w_S\left[n_3 - n_2 - (n_{30} - n_{20})\right] \\ &\quad + w_t\left[n_1 - n_2 - (n_{10} - n_{20})\right] \\ &\quad + W_S(n_3 - n_2)\end{aligned} \qquad (6.1)$$

$$\begin{aligned}\frac{dn_1}{dt} &= w_p\left[n_3 - n_1 - (n_{30} - n_{10})\right] \\ &\quad + w_t\left[n_2 - n_1 - (n_{20} - n_{10})\right] \\ &\quad + W_p(n_3 - n_1)\end{aligned}$$

Here n_{io} represents the population of level i in thermal equilibrium, which we can deduce from the Boltzmann distribution law (2.2). We have already seen that, in the microwave region,

$$h\nu/kT \ll 1,$$

and we shall henceforth expand the exponentials and retain only dominant terms. Under these conditions, if there are $(2S + 1)$ levels with total population N, we have

$$n_{io} - n_{jo} = \frac{N}{2S + 1}\,\frac{E_j - E_i}{kT} = \frac{N}{2S + 1}\,\frac{h\nu_{ij}}{kT} \qquad (6.2)$$

(Equation (2.3) follows from this when there are only two levels).

We can now find the steady state solution of equations (6.1) by equating the rate of change of the populations to zero. The stimulated

emission at frequency ν_s will be proportional to $n_3 - n_2$, and denoting $\nu_s/(\nu_p - \nu_s)$ by R, we have

$$n_3 - n_2 = \frac{E}{R(2S+1)kT} \frac{Nh\nu_s}{} \tag{6.3}$$

where

$$E = \frac{W_p(w_t - Rw_s) - R(w_p w_t + w_p w_s + w_s w_t)}{W_p(w_s + w_t + W_s) + (w_p w_s + w_p w_s + w_s w_t) + W_s(w_p + w_s)} \tag{6.4}$$

We shall refer to E as the efficiency parameter. The power emitted by these centres at frequency ν_s is

$$P_{em} = (n_3 - n_2)\, h\nu_s\, W_s \tag{6.5}$$

If we now consider the crystal to be placed in a resonant microwave cavity, resonant at frequencies ν_p and ν_s, we can describe this emitted power in terms of a negative quality factor Q_M, which we shall call the magnetic Q. Q is defined normally by the relationship

$$\frac{2\pi\nu}{Q} = \frac{\text{Power 'absorbed'}}{\text{stored energy}}$$

The magnetic energy stored in the cavity is simply $(1/8\pi) \int H^2 d\underline{r}$ where H is the signal field strength. We shall define an 'effective volume' of the cavity V_c by the relationship

$$V_c H_s^2 = \int H^2 dv$$

H_s being the signal field strength at the crystal, assumed constant. We shall further define a filling factor F, as

$$F = \text{Volume of crystal} / V_c.$$

Thus, associated with the population difference (6.3) we find a magnetic Q given by

$$-\frac{1}{Q_m} = \frac{4En_oFh^2\nu_s}{R(2S+1)kT} \frac{W_s}{H_s^2} \tag{6.6}$$

where n_o is the density of active ions in the crystal and R stands for $\nu_s/(\nu_p - \nu_s)$ as previously.

The aim in designing a Maser is to obtain a large emission from the crystal, that is, a low magnetic Q. We shall now consider the factors

occurring in (6.6) and show their significance in the choice of a paramagnetic crystal. Consider an 'ideal' case of a three level system, in which the saturation of levels 1 and 3 is complete, and there is a strong relaxation process, maintaining thermal equilibrium between levels 2 and 1. We then see immediately that the excess population $n_3 - n_2$ is

$$\frac{N}{2S + 1} \quad \frac{(\nu_p - \nu_s)}{kT} \tag{6.7}$$

Comparison with equation (6.3) shows that in this case we have E = 1. Incomplete saturation of levels 1 and 3, and a disturbance in thermal equilibrium between levels 1 and 2 reduces the number of active ions, reducing E. Hence the term 'efficiency parameter'.

It is instructive to compare (6.5) with the power absorbed by a crystal at frequency ν_s, for this is an easily measurable quantity. In thermal equilibrium, we have

$$n_2 - n_3 = \frac{N}{2S + 1} \quad \frac{h\nu_s}{kT}$$

Equation (6.3) therefore shows that the number of centres taking part in the stimulated emission is the product of three factors:-

(1) The number taking part in absorption at the signal frequency

(2) a frequency factor $\frac{\nu_p}{\nu_s} - 1$

(3) the efficiency parameter.

We shall consider these in turn and show the conditions necessary to optimise them.

(1) Absorption Properties By analogy with (6.6) the Q factor associated with absorption, Q_a, is given by

$$\frac{1}{Q_a} = \frac{4n_o F h^2 \nu_s}{(2S + 1)kT} \quad \frac{W_s}{H_s^2} \tag{6.8}$$

and W_s is given by (see equation (3.12)

$$W = \frac{\pi^2}{h^2} \quad |\mu_{32}|^2 \; H_s^2 \; T_{2s}^{*} \tag{6.9}$$

T_{2s}^{*} is the inverse of the line width of absorption. It has the

dimensions of time, and is often called the effective spin-spin relaxation time.

Low values of Q_a are obtained by using

(a) High filling factor F, that is, filling as much as possible of the region of high magnetic field in the cavity with the crystal. In principle, it should be possible to attain a filling factor of 0.5. In practice, figures of about 0.1 have so far been used.

(b) High density of ions, n_o, and long spin-spin relaxation time T_2^* - that is, narrow line width. The line width is mainly determined, however, by the interaction of the spins of neighbouring magnetic ions, and is directly proportional to n_o, so the product $n_o T_2^*$ is approximately constant for a given material.[12] At very low dilution, however, other factors such as magnetic field inhomogeneity contribute to the line width, so it is advantageous in such a case to increase the dilution until the line width increases proportionally. Line widths vary considerably in different crystals with the same dilution, and we require here crystals with a narrow line width. Such substances have been studied extensively, since they allow high resolution in absorption experiments. Line widths of 30 to 50 Mc/s are easily obtained with dilution of 1 part in 100.

(c) Low temperature, T. This affects Q_a both explicitly and implicitly in that the line width decreases at low temperatures.

(d) Large magnetic moment matrix element $|\mu_{32}|^2$. In general, this is a complicated function of the Spin Hamiltonian parameters and the direction of the signal field. It is large for an 'allowed' transition, and in an ideal case takes the value (see equation (4.9)

$$\frac{g^2\beta^2}{4}\Big(S(S+1) - M(M+1)\Big) \qquad (6.10)$$

where the transition is from level $S_z = M$ to level $S_z = M+1$. In general, the requirement that $|\mu_{32}|^2$ should be large influences the choice of direction and magnitude of the DC magnetic field and thus only indirectly the choice of the paramagnetic crystal.

(2) Frequency Factor $(\nu_p - \nu_s)/\nu_s$. Best operation is obtained by pumping at a frequency greatly in excess of the signal frequency. The pump frequency is limited, however, by the availability of components and valves. In addition, a large factor ν_p/ν_s causes difficulties in design of a cavity resonant at the two widely differing frequencies.

(3) Efficiency Parameter E We have already observed that when W_p and w_t are both large, E becomes unity. For lower pump powers, maser action is only possible if E is positive, which implies two conditions. The first of these is $\nu_t w_t > \nu_s w_s$. This is simply the condition that emission should occur at ν_s rather than ν_t. We may define the frequencies ν_s and ν_t so that this is satisfied. The second condition is

$$W_p > \frac{\nu_s}{\nu_t} \frac{w_p w_t + w_p w_s + w_s w_t}{w_t - w_s \nu_s/\nu_t} \qquad (6.11)$$

in other words, the pump power must be sufficiently large to cause 'saturation'. Equation (3.12) shows W_p to be

$$W_p = \frac{\pi^2}{h^2} |\mu_{31}|^2 H_p^2 T_{2p}^{*}.$$

H_p^2 can be increased simply by increasing the power input at the pump frequency. However, the power supply is limited by the capabilities of klystrons, and more seriously, high power input will heat the crystal. We shall see that increase in temperature causes increase in the w's, thus more than cancelling any increase in W_p. A large value of H_p^2 at the crystal can only be obtained using reasonable power input by use of a cavity with a high Q at the pump frequency.

A narrow line width for the pump transition is also desirable.

We also require a large matrix element $|\mu_{31}|$. If we consider the simple energy states leading to (6.10), we find that $|\mu_{31}| = 0$, corresponding to a non-allowed transition. Thus, we must use states which do not correspond to single quantum states M - that is, we must have admixture of the original quantum states. This can be achieved in a paramagnetic salt by combining the crystalline field splitting with the effect of a DC magnetic field applied at an angle to the crystal axis. We saw in Section 4 that this 'mixes' the original quantum states. If indeed the splittings due to the crystal field and the DC field are of the same order, the matrix elements μ_{31}, μ_{32} may be of the same order. The reduction in $|\mu_{32}|^2$ from the value (6.10) reduces the power output but is more than compensated by the increase in E for the same pump power.

Consulting equation (6.11), we see that the lower limit for W_p can be reduced by choosing a large ratio ν_p/ν_s. We have already stated above that this ratio should be as large as possible consistent with the design of a cavity with large Q's at both pump and signal frequencies.

The remaining terms in (6.11) concern the transition probabilities w, which describe the spin-lattice relaxation. They are generally described in terms of spin lattice relaxation times, T_{1s}, T_{1p} and T_{1t}, e.g. $w_{1s} = 1/2T_{1s}$. Theoretically these quantities should be calculated by considering in detail the interaction of the lattice with the spins, via the spin-orbit coupling of the ions. A few attempts have been made to do this, but with poor agreement with experiment.[13]

It may at first sight appear that these quantities can be easily measured experimentally. We have already seen in Section 2 that in a two-level system, absorption will saturate at high powers, and so T_1 can be measured. Care must be taken, however, in applying this method to a multi-level system, for saturation at one frequency will affect all the levels via the relaxation processes, and there is no means of separating the various relaxation times. We can thus see that an approximation has been made in the above discussion by considering only the three 'active' energy levels and disregarding the remaining levels. (When S = 1, there are, of course, only three levels). Full treatment of all the levels is possible,[14] and it can be shown that the above conclusions remain substantially true provided the relaxation times are suitably defined. Four independent spin-lattice relaxation parameters occur, in fact, and all these must be measured if reliable theoretical estimates are to be made of the operation of a Maser.

If we tentatively assume that w_s, w_p and w_t are of the same order of magnitude, then (6.11) shows that W_p is proportional to w_p. Thus, long spin lattice relaxation times are desirable to achieve saturation with reasonable power input. Of the paramagnetic ions, those with half filled shells of electrons are known to have this property, Fe^{+++}, Mn^{++}, and Cr^{+++} in the iron transition group, and Gd^{+++} in the rare earth group. The relaxation times are known to decrease rapidly with increase of temperature, meagre experimental and theoretical evidence suggesting that T_1 varies as T^{-7} in many cases. Hence, extremely low temperatures are necessary to achieve saturation; at present, operation at liquid helium temperatures appears necessary.

Long spin relaxation times, although desirable to reduce pump power, also limit the signal power which can be handled. Equation (6.4) shows that W_s begins to reduce the efficiency parameter when it becomes comparable with w_s. A balance between the two effects must be determined by the particular application for which the Maser is intended.

7. Operational Properties of a Maser

(1) Gain

We have seen that it is possible to establish a negative magnetic Q-factor Q_m, describing the paramagnetic crystal. Let us now consider

this crystal placed in a resonant cavity with unloaded Q-factor Q_u describing the ohmic losses in the walls, and with a resonant frequency ν_r, which we shall assume to be also the resonant frequency of the crystal.

Suppose a signal is incident at frequency ν_s through an input waveguide with Q-factor Q_e. In our determination of Q_m, we have assumed that the signal frequency is exactly the resonant frequency. We shall show in fact that the effective bandwidth of the amplifier is very small compared with the bandwidth of the paramagnetic transition, and shall hence neglect the variation of Q_m with frequency.

In the presence of the crystal, the 'effective Q' of the cavity, Q_c, is given by

$$\frac{1}{Q_c} = \frac{1}{Q_u} - \frac{1}{|Q_m|} \tag{7.1}$$

The impedance of the cavity, Z, under the above conditions is given by[15]

$$\frac{Z}{Z_c} = \frac{1/Q_e}{(1/Q_c) + 2i(\nu_s - \nu_r)/\nu_s}$$

where Z_c is the characteristic impedance of the wave guide.

Consider now a reflection cavity, in which the circuit arrangement is shown in Figure 2. The power reflection coefficient or gain is given by

$$G = \frac{(1 - |Z/Z_c|)^2}{(1 + |Z/Z_c|)^2}$$

$$= \frac{(1/Q_c - 1/Q_c)^2 + 2(\nu_s - \nu_r)^2/\nu_r^2}{(1/Q_e + 1/Q_c)^2 + 2(\nu_s - \nu_r)^2/\nu_r^2} \tag{7.2}$$

The maximum gain is clearly when $\nu_r = \nu_s$ and is

$$G_m = \left(\frac{1/Q_e - 1/Q_c}{1/Q_e + 1/Q_c} \right)^2 \tag{7.3}$$

This can be made large by varying the external Q-factor Q_e so that

$$\frac{1}{Q_e} \sim -\frac{1}{Q_c} = \frac{1}{|Q_m|} - \frac{1}{Q_u} \tag{7.4}$$

We see from (7.3) that amplification will occur when Q_c is negative, or

$$Q_u > |Q_m| \tag{7.5}$$

Hence we require a large unloaded Q-factor Q_u and a small magnetic Q-factor $|Q_m|$. Now we have shown, in the three level maser, that

$$\frac{1}{|Q_m|} = \frac{E}{RQ_a}$$

where Q_a is the Q of the ion in the absorbing state at the signal frequency. Hence (7.5) is immediately related to the condition $Q_a \sim Q_u$ which is the condition that the perturbation to the cavity caused by the crystal absorbing power at the signal frequency should be significant.[2] When the inequality

$$\frac{1}{|Q_m|} > \frac{1}{Q_u} + \frac{1}{Q_e}$$

is satisfied, (7.2) breaks down, and the maser will oscillate at the frequency ν_s. Adjustment of the coupling Q-factor Q_e, can prevent this occurring. Alternatively, under this condition the maser will act as a stable oscillator, with an output of perhaps a few microwatts at a lower frequency than the pump frequency ν_p.

(2) Bandwidth

When condition (7.4) holds, and the gain is large, the bandwidth, or the frequency width of (7.2) for half maximum gain, is

$$B = \frac{2\nu_r}{\sqrt{(G_m)} \cdot Q_c} \tag{7.6}$$

$$\sim \frac{2\nu_r}{\sqrt{(G_m)} \cdot |Q_m|} \quad \text{(when the ohmic losses are small)}$$

Thus larger gains imply a lower bandwidth. Using typical figures for S band amplification (ν_s = 2.8 k Mc/s), we find that with a gain of 20 dB (G = 100), the bandwidth is 500 kc/s. Use of a lower $|Q_m|$ increases the bandwidth, and it is conceivable that, with good crystals, the bandwidth could be increased to 5 Mc/s. The bandwidth of the paramagnetic transition is 50 Mc/s, and we thus see that the limitation on bandwidth lies in the associated microwave circuitry.

(3) Noise Factor

The noise factor F of a microwave receiver is defined as the noise power output per unit bandwidth divided by the fraction thereof coming from the source impedance. If the noise powers per unit bandwidth at the input and output are P_{n_1}, P_{n_2} respectively, we may write

$$P_{n_2} = GP_{n_1} + P_a$$

where G is the gain and P_a is the noise power per unit bandwidth inserted by the amplifier. Hence

$$F = \frac{P_{n_2}}{GP_{n_1}} = 1 + \frac{P_a}{GP_{n_1}} \qquad (7.7)$$

Ideally, $P_a = 0$ and $F = 1$. By conventional means, a noise factor of about 4 is obtainable (6 dB). The practical importance of the Maser lies in the fact that very low noise factors should be obtainable - in an ideal case $F = 1.01$, or 0.04 dB.

If the source impedance (assumed to be matched) is at temperature T_s the input noise power is

$$P_{n_1} = kT_s \qquad (7.8)$$

provided the frequency is such that $h\nu$ is less than kT_s. Hence, F depends upon the source temperature. A more convenient quantity to consider is the 'noise input temperature' of the amplifier, T_a, defined as the source temperature for which the output noise power coming from the source is equal to that coming from the amplifier. The noise factor is then

$$F = 1 + \frac{T_a}{T_s} \qquad (7.9)$$

In order to evaluate T_a, we have to consider the entire amplifier, that is, all parts occurring before the final output. We shall consider here in detail only the contribution from the magnetic ions, and shall quote the formula involving the remaining components of the maser.

The noise contributed by the paramagnetic ions arises from the spontaneous emission process considered in Section 2. The output power P_2, arises from stimulated emission, proportional to the input power P_1 and to the population difference $N_2 - N_1$ in the two active states, and from spontaneous emission, proportional to the population of the upper state N_2. Thus

$$P_2 = A(N_2 - N_1)\, P_1 + BN_2 \qquad (7.10)$$

In thermal equilibrium, where source and amplifier have the same temperature T_s, we have already seen that

$$\frac{BN_2}{A(N_2 - N_1)P_1} = \frac{h\nu}{kT_s}$$

Using the known Boltzmann distribution for N_1 and N_2, and (7.8) for P_1 gives

$$B = Ah\nu$$

Substituting into (7.10), dividing by $A(N_2 - N_1)P_1$, and comparing with (7.9) shows that the contribution to T_a arising from the crystal is just

$$T_a = |T_m| \qquad (7.11)$$

where T_m is the negative effective temperature of the ions, defined as in Section 2. When operation is at liquid helium temperatures, T_m will ideally be about $2^{\circ}K$, corresponding to a noise factor, F, for a source at room temperature of about 0.04 dB.

The contribution to the noise from the cavity and the transmission lines has been considered in detail by a number of authors[14,16] and the result may be stated as follows (see Butcher[14]).

$$T_a = |T_m| + X(|T_m| + T_e) \qquad (7.12)$$

where R_m is the negative effective temperature of the ions,

T_e is the mean temperature of the parts of the amplifier in thermal equilibrium

and X is a factor which is small for a good amplifier. When there are no ohmic losses, X is zero, and (7.9) is obtained.

In a reflection cavity, (see Figure 2), assuming the circulator to be ideal and the wave guides to be loss-less, we find

$$X = \frac{|Q_c|}{Q_e}$$

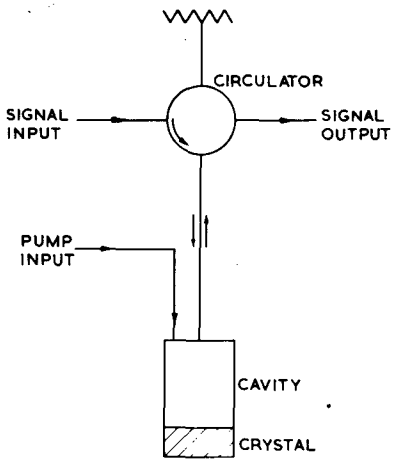

Figure 2 The reflection cavity Maser

and the noise factor is

$$F = 1 + \frac{|T_m|}{T_s} \cdot \frac{|Q_c|}{|Q_m|} + \frac{T_\ell}{T_s} \frac{|Q_c|}{Q_e} \qquad (7.13)$$

where T_ℓ is the mean temperature of the cavity and the crystal lattice.

We see from the above discussion that, theoretically, very low noise factors are obtainable, approaching a lower limit of 0.04 dB. In order to achieve low noise factors, however, (7.12) shows that T_e should be low, that is, the temperature of the 'lossey' parts of the amplifier, including the circulator and transmission lines, should be low.

The possibility of attaining such low noise figures suggests the value of the Maser in radio astronomy. The signals observed from the low temperature source are incoherent, and the significant parameter is the excess noise factor F - 1. In radar detection, coherent signals are examined, and the significant parameter is the signal to noise power ratio at the amplifier output divided by this ratio at the input, which is 1/F. This can approximate very closely to unity in an ideal Maser.

8. Practical Three Level Masers

So far as is known at the time of writing, there are three solid state Masers in existence, only one of which has been operated as a controlled amplifier.

Scovil, Feher and Seidel[17] at B.T.L. used gadolinium ethyl sulphate as active material, diluted with corresponding diamagnetic lanthanum salt. Pump power is applied at J band (17.5 k Mc/s), and the signal at X band (9 k Mc/s.). Gadolinium has eight energy levels, and levels 2, 3 and 4 are used, with a D.C. magnetic field of 2850 gauss. The crystal is doped with cerium, which has a short spin lattice relaxation time, and conditions are so chosen[18] that this reduces the relaxation time T_t, increasing the efficiency parameter E. Measurements by simple saturation methods show that T_s is about 10^{-4} secs, T_t about 10^{-5} secs, and T_{2s} equals 8.10^{-9} secs. No other relaxation terms are measured. A reflection type rectangular cavity was used,

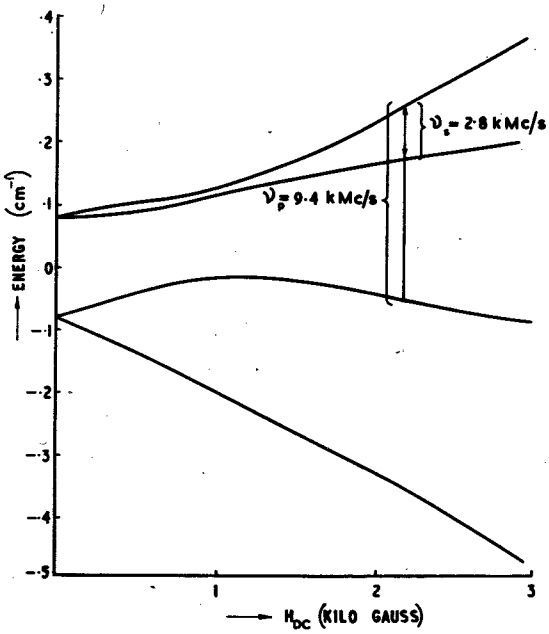

Figure 3 Energy levels in potassium chromicyanide (Magnetic field along c-axis)

the rf fields at J and X band were parallel, and the Q-factors (presumably at liquid helium temperature) were Qp = 1000, Q_S = 8000. The filling factor was about 10%. Oscillation at X band was obtained with a pump power of 100 mW. This high pump power is made necessary by the low spin lattice relaxation times, and the small value of $|\mu_{31}|^2$ (about 1% of $|\mu_{32}|^2$).

The second Maser to operate was at Lincoln Labs. McWhorter and Meyer[19] used potassium chromicyanide, diluted with diamagnetic potassium cobalticyanide. The active chromium ions have four low lying levels (S = 3/2) and Figure 3 shows the levels used in the Maser. Pumping is at X band (9.4 k Mc/s.) and amplification at S band (2.8 k Mc/s.). Measured relaxation times are T_{1S} = 0.2 secs, T_{2S} about 5.10^{-9} secs. A fixed tuned cylindrical cavity was used with low temperature Q_S of about 23,000. The long T_{1S} allows low pump power; maximum efficiency as an oscillator is reached with about 1 mW. input at X band. As an amplifier, stable gains of 20 dB with 250 kc/s bandwidth, and 37 dB with 25 kc/s bandwidth were obtained. Long T_{1S} also implies low signal power handling, and the gain decreased when the signal power exceeded 10^{-10} watts.

The narrow line width and long T_1 of potassium chromicyanide make it a suitable substance for experiment, and operation at X band (pumping at K band) and at 1420 Mc/s. (pumping at X band) have been reported from Columbia and Harvard Universities.

At R.R.E. work is in progress to produce Masers, using gadolinium ethyl sulphate doped with cerium, and potassium chromicyanide, to amplify at centimetre wavelengths. Work is also in progress to investigate various paramagnetic crystals by paramagnetic resonance techniques, and to study

further those crystals which have previously been investigated.[20]

9. Conclusion

The solid state Maser offers a means of low-noise amplification in the microwave region of the spectrum. A great deal of work is required before the device becomes a practical tool. One pressing need is to achieve larger bandwidths. At present, it would not appear possible to achieve bandwidths greater than, say, 5 k Mc/s. for a 20 dB gain at S band. Bandwidths are limited by the Q of the associated circuitry, and one solution may lie in development of the travelling wave Maser.

Before accurate predictions can be made on the performance of a Maser, it is necessary to measure the relaxation parameters, and a considerable field for both experimental and theoretical investigation is opening up.

10. Acknowledgments

This article owes much to many stimulating discussions with Dr. M. H. Oliver and his group, and with Dr. P. N. Butcher. Thanks are also due to Dr. R. A. Smith for allowing free use of lecture notes on some of the material included here.

11. References

1. J. P. Gordon, H. J. Zeiger and C. H. Townes, Phys. Rev. 99, 1264 (1955).

2. B. Bleaney and K. W. H. Stevens, Rep. Prog. Phys. 16, 108 (1953).

3. A. Abragam and M. H. L. Pryce, Proc. Roy. Soc. A, 205, 135 (1951).

4. M. H. L. Pryce, Proc. Phys. Soc. A, 63, 25 (1950).

5. E. U. Condon and G. H. Shortley, Theory of Atomic Spectra, p.48 (1953).

6. F. Bloch, Phys. Rev. 70, 460 (1946).

7. J. P. Wittke, Proc. I.R.E., March 1957, p.291.

8. J. Combrisson, A. Honig and C. H. Townes, Comptes Rendus, 242, 2451 (1956).

9. D. Bolef and P. Chester, Sci. Paper 6-94466-5-P11, Westinghouse Res. Lab. May, 1957.

10. N. Bloembergen, Phys. Rev. 104, 324 (1956).

11. J. O. Artman, Harvard Univ. Cruft Lab. Tech. Rep. 264 (1957).

12. J. H. Van Vleck, Phys. Rev. 74, 1168 (1948).

13. J. H. Van Vleck, Phys. Rev. 57, 426 (1940).

14. P. N. Butcher, Electron Tube Lab. Stanford Univ. Tech Rep. 155-1 (1957).

15. J. C. Slater, Microwave Electronics.

16. R. V. Pound, Ann. Physics, 1, 24 (1957).

17. H. E. D. Scovil, G. Feher and H. Seiden, Phys. Rev. 105, 762 (1957).

18. G. Feher and H. E. D. Scovil, Phys. Rev. 105, 762 (1957).

19. A. McWhorter and J. W. Meyer, CR.M37-14, Lincoln Lab. M.I.T. (1957).

20. See, for instance, K. D. Bowers and J. Owen, Rep. Prog. Phys. 28, 304 (1955).

MAGNETISM AND THE RARE-EARTH METALS

by J. M. Lock

1.1 Introduction

The rare-earths, together with lanthanum, form a group of fifteen elements lying between barium and hafnium in the Periodic Table, and possessing such closely similar chemical properties that it is only since the war that they have become available in anything approaching a state of high purity. In the course of the last ten years considerable progress has been made in the study of their chemical and physical behaviour, but we shall be concerned in this article only with the physical properties of the pure metals themselves.

The story of the rare-earths begins with the discovery in 1794 by J. Gadolin of a new 'earth', or metallic oxide, in the mineral gadolinite, and in 1797 A. G. Ekeberg named the new earth yttria after Ytterby, the place in Sweden where this mineral was found. In 1843 C. G. Mosander found that yttria is really a complex earth containing the oxides of two new elements, erbium and terbium, besides yttria proper, and the second half of the nineteenth century saw the gradual separation of the various rare-earths from the parent minerals. Nevertheless even by 1887 W. Crookes could still say: 'The rare-earths form a group to themselves; chemically, they are so much alike that it takes the utmost skill of the chemist to effect even a partial separation, and their history is so obscure that we do not yet know the number of them'.

We now realise that the chemical properties of elements depend on the different structures of the various shells of electrons which surround their atomic nuclei, and especially on that of the outermost shells. Thus the inert gases helium, neon, argon, krypton and xenon owe their inactivity to the fact that their atoms possess a completely closed outermost shell of 8 electrons, while the alkali metals are characterized by having closed shells surrounded by an outermost shell with only one electron in it. The various shells which can be occupied by electrons are labelled by a principal quantum number, 1, 2, 3, 4 .....

specifying roughly speaking, the dimensions of the shell, and by a letter, s, p, d, f, g..... specifying its shape. Each shell can contain a limited number of electrons only, and as successive electrons are added as we go through the periodic table they fill up the available shells in turn, starting with those of lowest energy and occupying those of higher energy until a place has been found for all the electrons. In specifying the state of an atom it is usual to give the number of electrons in the partially filled outer shells only, completely filled inner shells being left 'understood'. Thus the atom of iron has the 1s, 2s, 2p, 3s and 3p shells completely filled (18 electrons) and has 6 electrons in the 3d shell and 2 in the 4s shell. Its state is specified as $3d^6\ 4s^2$. The total number of electrons in the unionized atom is the atomic number of the element in question, 26 in the case of iron.

Generally speaking the energy of an electron is determined primarily by the principal quantum number of the shell, increasing with this number. But it is also affected by its shape, the energy of the s or spherical type of shell being the lowest. It so happens that by the time we have reached the element lanthanum in the periodic table all the shells up to and including the 4d have been completed, and it has been energetically favourable for the 5s, 5p and 6s shells to have filled up, and even for one electron to enter the 5d before the 4f shell has begun to be occupied. At this point, however, the energy of an electron entering the 4f shell becomes lower than that of one in the 5d, so that before the filling of the latter proceeds further all the 14 available places in the 4f shell are successively filled, giving the fourteen rare-earth elements, each of which has substantially the same outermost shell structure. It is this fact which explains the remarkable similarity between the chemical properties of the different rare-earth elements; any differences between them must be looked for in the different numbers of 4f electrons, and these lie well within the atom, shielded by the larger filled shells 5s, 5p and 6s.

When the atoms of a metal are bound together in the solid state the interactions between them affect the energies of their outermost electrons enormously, enabling some of them, the 'conduction' electrons, to escape entirely from their parent atoms and to wander freely through the metal. This means that partly filled shells in the solid state often have their energies considerably distorted by the presence of neighbouring atoms. Thus in the 'transition' elements partially filled 3d, 4d and 5d shells are surrounded in the free atom by only one or two 4s, 5s and 6s electrons respectively. In the metallic state these become conduction electrons and the d shell is left naked: indeed, some of the d electrons, too, usually go into the conduction band as it is called, and their properties are modified out of all recognition as compared with those in the free atom.

The rare-earth elements, however, provide a unique opportunity for studying partially filled shells in the solid state under conditions not too dissimilar from those prevailing in the free atom. For in them only the 5d and 6s electrons are removed to the conduction band, leaving the 4f electrons still well shielded by the 5s and 5p closed shells, 8 electrons in all. It is this fact which makes these elements of particular interest to the physicist in his attempts to interpret the effects of atomic interactions in the solid state on the behaviour of electrons; for it is rarely possible to work out a satisfactory theory of large interactions before an understanding of smaller ones is reached.

1.2 Preparation and General Physical Properties

Until 1944 the separation of the rare-earths was a formidable task depending largely upon the tedious process of fractional crystallization of suitable salts repeated a few hundred times. It is not surprising, therefore, that the elements were not available at that time in anything like a high state of purity. Nowadays a much more rapid and effective method of separation depending on the ion-exchange process has been developed[1] and all the elements with the exception of promethium (61), which does not exist in nature, have been prepared with impurities of less than a half per cent.

The ion-exchange process depends on the establishment of equilibrium between ions in solution, generally aqueous, and ions attached to a suitable solid matrix. In the separation of the rare-earths the method used is generally as follows. Certain synthetic resins have been made consisting essentially of sulphonated benzene rings linked with aliphatic chains, and in them the hydrogen ion of the sulphonic acid group can readily be replaced by other cations. If hydrogen is attached to these groups the resin is said to be in the hydrogen state, while if a different cation, such as ammonium or a rare-earth ion is attached, then it is said to be in the ammonium or rare-earth state. The polymeric network of the resin makes it highly porous, and enables an ionic solution to diffuse through it freely, exchanging cations with those attached to the resin until equilibrium is established.

A quantity of resin is poured into glass columns which have a suitable porous plate at their base to support the resin bed, and after washing, the resin is converted to the desired cationic state by passing an excess of a solution of the cation through it. Next a solution of the mixed rare-earths (usually the chlorides) is poured into the column, and the rare-earth cations completely replace those attached to the resin at the top of the column, forming a narrow saturated band. The process in which separation occurs now follows: a solution of some negative-ion species (e.g. citric acid) which forms a tight complex with the rare-earths is allowed to flow steadily down through the column.

TABLE 1

Element	Melting point (°C)	Crystal structure	Lattice constants (A.U.) a	c	Density
La	920	hex	3.770	12.159	6.162
		f.c.c.	5.302		6.190
Ce	804	f.c.c.	5.161		6.768
Pr	935	hex	3.672	11.835	6.769
Nd	1024	hex	3.658	11.799	7.007
Pm	-	-	-	-	-
Sm	1152	rhombic	3.621	26.25	7.540
Eu	<900	b.c.c.	4.606		5.166
Gd	~1250	h.c.p.	3.636	5.783	7.868
Tb	1365	h.c.p.	3.601	5.694	8.253
Dy	~1400	h.c.p.	3.590	5.646	8.556
Ho	~1500	h.c.p.	3.577	5.616	8.799
Er	1500-50	h.c.p.	3.559	5.587	9.058
Tm	1550-1650	h.c.p.	3.537	5.555	9.318
Yb	824	f.c.c.	5.486		6.959
Lu	1650-1750	h.c.p.	3.503	5.551	9.849

It sets up a competition for the rare-earth ions between the aqueous and solid phases, an individual ion continually exchanging between the resin and the negative ions in solution. The positive ions in the complexing solution replace the rare-earth ions at the rear edge of the band so that the rare-earth band is driven down the resin bed. Because the stabilities of the different rare-earth complexes vary slightly, the most tightly bound complex moves most rapidly down the column, and after traversing a suitable length complete separation of the different elements can be attained, bands of each issuing one after another at the bottom end of the column.

The separation of their salts having been accomplished, the pure rare-earth metals are generally prepared from the chloride or fluoride by reduction with calcium or lithium in a tantalum crucible contained in a sealed steel bomb, the temperature being raised above the melting-point of the rare-earth. This then collects into a fused mass, while the calcium or lithium halide forms a slag on top which can easily be removed[2].

The pure metals are mostly extremely reactive, the rate of oxidation in air being higher than that of any other metals with the exception of the alkalis and the alkaline earths. Sesquioxides, R_2O_3, are formed, and all the rare-earths are in fact normally trivalent, though some will form less stable compounds

in which they have a different valency. More will be said about this point later. The crystal structures have been investigated fully by Spedding, Daane and Herrmann[3], using the X-ray powder photograph technique, and are generally either hexagonal or cubic close-packed. In some of the hexagonal metals an unusual order of the layer stacking gives rise to a large value of the lattice constant c, and europium is exceptional in having the body-centred cubic structure. Some properties of the rare-earth metals are given in Table 1.

2.1 Paramagnetism

As we have said, differences in the properties of the rare-earths arise mainly from the number of 4f electrons in their atoms and this number will not appreciably affect the chemical properties of the element. It is, however, of paramount importance in determining its magnetic properties as we shall now proceed to show.

The modern theory of magnetism has shown that it is fundamentally of an electrical nature, and is due to the fact that an electron circulating about the nucleus of an atom constitutes a small electric current which in turn possesses a magnetic moment. This moment bears a definite relation to the angular momentum of such an orbital electron, the ratio magnetic moment/angular momentum - or 'magneto-mechanical ratio' - being always equal to e/2mc. The electron itself possesses angular momentum or 'spin' and magnetic moment, but in its case the magneto-mechanical ratio is equal to e/mc, exactly twice the orbital value. According to modern quantum theory angular momentum, whether of orbit or spin, is quantized in units of $h/2\pi$, the orbital and spin quantum numbers for a given electron being designated by ℓ and s respectively. The quantum number ℓ can take integral values 0, 1, 2, 3, (corresponding to s, p, d, f ..... states) up to (n-1) where n is the principal quantum number of the electron (vide supra) while s always has the value $\frac{1}{2}$. The numerical values of the total orbital and spin angular momenta turn out to be $(h/2\pi)\sqrt{\ell(\ell+1)}$ and $(h/2\pi)\sqrt{s(s+1)}$ instead of the simple $\ell h/2\pi$ and $sh/2\pi$ given by the old quantum theory, but this is a detail that need not worry us. The magnetic moment of an atom is thus similarly quantized in units of $eh/4\pi mc$, known as the Bohr magneton, μ_B.

2.2 The Vector Model - Russell-Saunders Coupling

In a many-electron atom it is necessary to know how the effects of the individual electrons add up, and in general this is a very difficult problem. In many atoms, however, (and fortunately in the rare-earths) the interaction between the several spin and orbital angular momenta is small compared with the interaction of the orbital angular momenta among themselves, and of the spins among themselves. In this case the

spins form a resultant (integral or half integral) S, and the orbital angular momenta form a resultant L (integral). These may be regarded as vectors which add together to form a resultant J, the total angular momentum quantum number. J can then assume the range of values

$$J = |L - S|, \quad |L - S + 1|, \ldots\ldots, L + S - 1, L + S$$

Its projection along a particular axis of quantization is the 'magnetic quantum number' M, which can have the values

$$M = -J, -J + 1, \ldots\ldots, J - 1, J$$

and which determines the component of total angular momentum in the direction of this axis. This type of interaction between the individual electrons is called Russell-Saunders Coupling, and is illustrated in Figure 1.

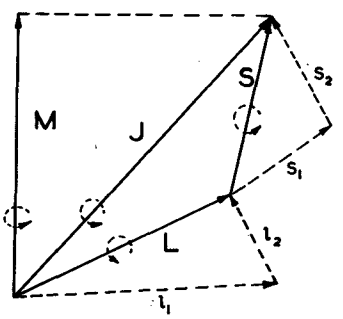

Figure 1 Russell-Saunders coupling

The strong coupling between the ℓ's and between the s's may be taken to mean that in the figure the vectors representing them are precessing rapidly about the vectors representing their resultants L and S. In the same way the coupling between L and S to give J means that these two vectors are precessing about J, while on the application of a magnetic field in the z-direction the whole figure precesses comparatively slowly about M. All such attempts at geometrical pictures should not, however, be taken too seriously, since the actual way in which angular momenta combine has to be worked out by matrix algebra rather than by vector addition.*

We must now calculate the effective magnetic moment of an atom in which Russell-Saunders coupling applies, remembering that the gyromagnetic ratio for spins is double that for orbits. The appropriate vector diagram for the magnetic moments is therefore as shown in Figure 2. In the most usual case the rate of precession of angular momentum about the direction of J is greater than that of the whole figure about the direction of the applied field, and this means that the effective magnetic moment of the atom is given, not by its instantaneous value

---

* For example, the total angular momentum is actually $(h/2\pi)\sqrt{J(J + 1)}$ - as above for ℓ, s - but even in the case where J is 'parallel' to the z-axis its component in this direction, $Mh/2\pi$, is exactly equal to $Jh/2\pi$ and not to $(h/2\pi)\sqrt{J(J + 1)}$.

which is proportional to the length AC in the diagram, but by its projection AB, on the direction of J. Physically, this means that the coupling energy between L and S is large compared with kT where T is the temperature of the atoms concerned. Now

$$AB = (\text{projection of L on J}) + (\text{projection of 2S on J}) = \frac{\mathbf{L}.\mathbf{J}}{|J|} + \frac{2\mathbf{S}.\mathbf{J}}{|J|}$$

Now by the cosine formula we have

$$2\mathbf{L}.\mathbf{J} = L^2 + J^2 - S^2$$

and

$$2\mathbf{S}.\mathbf{J} = S^2 + J^2 - L^2$$

whence we obtain

$$AB = \frac{S^2 - L^2 + 3J^2}{2|J|}$$

$$= J\left\{1 + \frac{S^2 + J^2 - L^2}{2J^2}\right\}$$

Remembering that in the exact matrix calculations quantities like J^2 have to be replaced by $J(J+1)$, the effective magnetic moment of the atom is given by

$$\mu_{eff} = g\mu_B \sqrt{J(J+1)} \qquad (1)$$

where g, the Landé splitting factor as it is called, is given by

$$g = 1 + \frac{S(S+1) + J(J+1) - L(L+1)}{2J(J+1)} \qquad (2)$$

(Note that if S = 0 so that J = L, g = 1: this is the result we should expect if the orbital moment only is effective. If L = 0, so that J = S, g = 2: this is the result we should expect if the spin only is effective).

The susceptibility per atom is now given by the classical formula

$$\chi = \mu_{eff}^2/3kT \qquad (3)$$

expressing the fact that, if interactions between neighbouring atoms can be neglected, a system of paramagnetic atoms such as we have been considering should obey the Curie Law, its susceptibility being

inversely proportional to the absolute temperature.

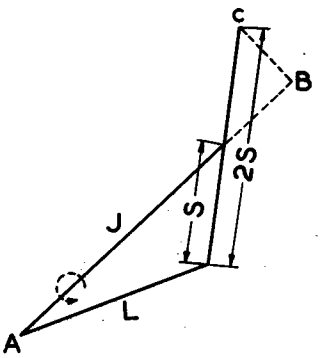

Figure 2 Vector diagram for 'g'

In applying the above theory to a particular rare-earth element we have, of course, to know what values to take for the quantum numbers L, S and J. In magnetism, as opposed to spectroscopy, we only require this information for the state of lowest energy, or ground state as it is called, and here we can make use of a theory due to Hund which tells us that of the various possible states the one with lowest energy is that with the maximum possible value of S, and also the maximum L consistent with this. (This qualification is necessary because of the Pauli Exclusion Principle which tells us that no two electrons in an atom can occupy the same quantum state). Let us see how we can use these rules in a particular case, say neodymium. Here we have three electrons in the 4f shell, for each of which $\ell = 3$, $s = \frac{1}{2}$. Their quantum states will be completely determined by the projections of these orbital and spin moments in a given direction, so Pauli's Principle tells us that these cannot all be the same. According to Hund's Rule, however, the state of lowest energy will be that for which S, the vector sum of the individual s's, is a maximum, namely $3/2$, obtained when the s's are all lined up parallel. No two of the ℓ's may be parallel therefore, and the maximum possible value of L consistent with this is $3 + 2 + 1 = 6$. This arrangement of spins and orbits is illustrated schematically by the vector diagram of Figure 3. You must not expect these lengths really to add up like this - the total orbital momentum L is a matrix, but the scheme gives the right answer!

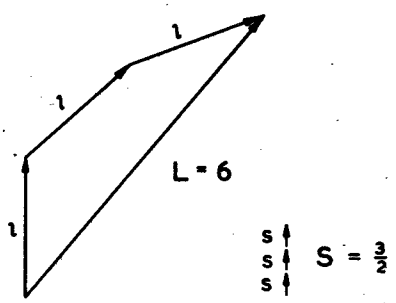

Figure 3 Vector diagram for neodymium

As stated above, J can now assume values ranging from $|L - S|$ to $L + S$, but it can be shown that the value which gives the lowest energy is that with minimum or maximum J according as the number of 4f electrons is less than or greater than 7 (half-filled shell). In the case of neodymium therefore the lowest state has $L = 6$, $S = 3/2$,

Table 2

Element	4f-electrons in +++ ion	Ground state	μ_{eff}/μ_B
La	0	$^1S_0$	0
Ce	1	$^2F_{5/2}$	2.54
Pr	2	$^3H_4$	3.58
Nd	3	$^4I_{9/2}$	3.62
Pm	4	$^5I_4$	2.68
Sm	5	$^6H_{5/2}$	0.84
Eu	6	$^7F_0$	0
Gd	7	$^8S_{7/2}$	7.94
Tb	8	$^7F_6$	9.7
Dy	9	$^6H_{15/2}$	10.6
Ho	10	$^5I_8$	10.6
Er	11	$^4I_{15/2}$	9.6
Tm	12	$^3H_6$	7.6
Yb	13	$^2F_{7/2}$	4.5
Lu	14	$^1S_0$	0

$J = {}^9/_2$. (The energy differences for states with different J is much less than the energy difference produced by changing S or L).

The notation which is used to describe atomic states is that of the spectroscopist. The value of L is denoted by the same capital letter as the small letter used for the angular momentum of an individual electron. Thus S describes the state for which L = 0, P for L = 1, D for L = 2, F for L = 3 and then in successive alphabetical order for

L = 4, 5 ..... etc. A superfix before the capital letter gives the value of 2S + 1, known as the multiplicity of the state: this is because, as we have just said, with given values of L and S (provided L is greater than S which is usually the case) J can take the 2S + 1 values from $|L - S|$ to L + S, the splitting of the energies of these different states being comparatively small, though still usually large compared with kT at room temperature. Finally the value of J is given as a suffix after the capital letter.

Once the ground state of the atom is known its magnetic susceptibility can be predicted from equations (1), (2) and (3). In Table 2 are given the ground states expected for the trivalent ions of the rare-earth elements, and their theoretical effective atomic moments calculated in this way.

2.3 Co-operative Magnetic Phenomena - Ferromagnetism and Anti-Ferromagnetism

The simple Curie Law of equation (3) will only be obeyed if the atoms are free from disturbing interactions with their neighbours other than random thermal jostling. Because of the high degree of shielding provided by the outer electrons we should expect the magnetic behaviour of the 4f shell to agree fairly closely with the simple theory, particularly in chemical compounds in which the nearest neighbouring rare-earth ions are comparatively distant, and in fact this proves to be the case. Certain of the rare-earth salts, such as gadolinium sulphate, $Gd_2(SO_4)_3.8H_2O$, show no observable departure from the Curie Law down to temperatures at least as low as 2°K.

In the metals themselves the magnetic ions are much closer together and interactions between them will be correspondingly larger. These interactions are not directly of a magnetic nature, however - the magnetic interaction energy between two dipoles of moment μ_B separated by a distance of 1 A.U. corresponds roughly to their thermal energy at only 1°K - but are much more closely akin to the forces which bind two hydrogen atoms together in a molecule (covalent binding). They are known as exchange forces, because they arise from the fact that, electrons being identical particles, we cannot distinguish in any way between two states of a system which differ only in that two electrons have exchanged places. In addition, we know from Pauli's Exclusion Principle (vide supra) that the states of the two electrons cannot be identical, so that if their spins are the same (parallel) then their orbital states must differ, while if their orbital states are the same their spins must be different (anti-parallel).

Clearly, therefore, the state of the complete system is closely bound up with the question of the parallelism or anti-parallelism of

the electron spins, and in particular the charge cloud distribution (which is determined by the orbital state of the electrons) may differ in the two cases, this giving rise to a large electrostatic change of energy. In this way, to quote Van Vleck, 'the exchange effect, though entirely orbital in nature, is, because of the exclusion principle, very sensitive to the way in which the spin is aligned, and is formally equivalent to 'cosine coupling' between the spin magnets of the various atoms'. In other words the spin dependence of energy of two atoms varies with the angle between their spins exactly like the classical interaction between two dipoles, only the coupling coefficient is very much larger. The energy of two spins can be written as $-Is_1 \cdot s_2$ (scalar product of the two vectors s_1 and s_2) where the coupling coefficient I, known as the exchange integral, is large.

These considerations form the basis of the Heisenberg theory of ferro-magnetism. Calculation of the exchange integral I for a particular crystal is difficult, as it depends on a knowledge of the exact wave-functions of the electrons involved, but qualitatively it will be larger if the wave-functions overlap appreciably midway between neighbouring atoms, so that states of high orbital eccentricity (ℓ large) will have larger values of I than s states ($\ell = 0$, spherical symmetry). It turns out that I can be either positive or negative, and these two cases correspond respectively to equilibrium conditions in which neighbouring spins are parallel (ferromagnetism) or anti-parallel (anti-ferromagnetism). Roughly speaking the ordered state will set in at a temperature such that $|I|\ zS^2$ is of the order kT, where z is the number of nearest neighbours surrounding each atom, and S the spin quantum number as before. The critical temperature for the appearance of the ordered state is known as the Curie temperature for ferromagnetics, and the Néel temperature for anti-ferromagnetics. Above it the simple theory suggests that the substance should obey not the Curie Law but the Curie-Weiss Law

$$\chi = C/(T - \theta) \qquad (4)$$

where the Curie constant C has the same value, $\mu_{eff}^2/3k$ as for free atoms, while θ should be equal to the Curie temperature, T_c for a ferromagnetic, but should be negative for an anti-ferromagnetic, its relation to the Néel temperature, T_N being determined by the relative magnitudes of nearest and next-to-nearest neighbour interactions.

Below the critical temperature a ferromagnetic substance shows the familiar phenomenon of large magnetic saturation even in small applied fields. According to the theory any microscopic region of such a material is in fact saturated in zero field, all the spins being aligned parallel, but the overall magnetization is much lower than this in

general because of the existence of domains. In the case of anti-ferromagnetic substances the susceptibility reaches a maximum at the Néel temperature, decreasing below it because of the onset of anti-parallel ordering. This type of ordering, unlike the ferromagnetic, can exist in many forms, according to the symmetry of the crystal concerned, and it is only in the simplest types of crystal such as the simple cubic and body-centered cubic that a magnetic arrangement can exist in which the spins on all the nearest neighbours of a given atom are anti-parallel to its spin. In close-packed arrangements, whether hexagonal or cubic, some of the nearest neighbours of a given atom are also nearest neighbours of one another and no simple configuration of anti-parallel spins can exist. Qualitatively, however, the principle remains the same, that as many as possible of the nearest neighbours will have their spins arranged anti-parallel.

2.4 Specific Heat of Magnetic Substances

As we have shown, the ground state of a magnetic atom is characterized by definite values of L, S and J, and just as a state with given L and S leads (if L is greater than S) to $(2S + 1)$ possible values of J whose energies are separated usually by an amount greater than kT at room temperature, so also a state with given J is split in a suitable field into $(2J + 1)$ separate energy levels, separated by a much smaller amount, usually less than kT at room temperature. At any given temperature the number of atoms occupying each energy level has an equilibrium value determined by the classical Boltzmann statistics, and the redistribution of atoms among these levels as the temperature is changed leads to a specific heat anomaly. At absolute zero all the atoms will be at the lowest energy level while at temperatures high compared with the total splitting all levels will be occupied in equal numbers. The total rise in entropy for one mole of the substance should therefore be $R \ln (2J + 1)$, and the set of levels considered as one state characterized by J is said to be $(2J + 1)$-fold degenerate. The temperature range over which this entropy is spread, and thus over which the specific heat behaves anomalously, will depend on the process by which this degeneracy is removed. There are two main mechanisms involved.

(1) In Stark splitting the levels are separated by the electric field in the crystal. In this case we can still think of each atom separately as existing in the field provided by its neighbours, and can thus make use of single atomic quantum states and energy levels. The degree to which the degeneracy can be removed is limited by the symmetry of the crystal (in general strongly anisotropic fields are more effective in removing the degeneracy than isotropic fields), and by a law discovered by Kramers[4] which tells us that, for states with an odd number of electrons, each level must remain at least doubly degenerate

under the influence of electric fields. The specific heat anomaly produced by this type of splitting is a smooth 'Schottky bump', of the form

$$\frac{C}{R} = \frac{\Delta^2}{k^2T^2} \quad \frac{e^{-\Delta/kT}}{(1+e^{-\Delta/kT})^2}$$

for a simple splitting into two energy levels separated by an amount Δ (see Figure 4(a)).

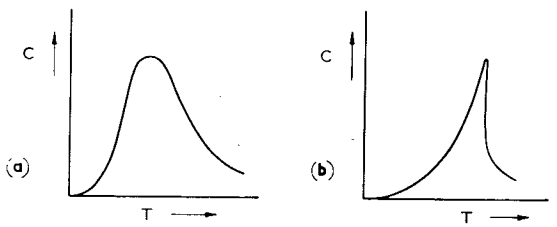

Figure 4(a) Schottky bump
Figure 4(b) Co-operative phenomenon

(2) The second mechanism is the onset of one of the two co-operative phenomena - ferromagnetism or anti-ferromagnetism. In this case, because we are dealing with the interactions between a number of atoms, we can no longer label them by single atomic states and the picture of the splitting of the degeneracy into individual levels breaks down. The specific heat anomaly above the critical temperature should fall sharply (see Figure 4(b)), because of the sharp onset of the ordered state on cooling, but, according to the relative magnitudes of the Stark splitting and the ferro or anti-ferromagnetic exchange energy, the co-operative phenomenon may or may not account for the whole $R \ell n (2J + 1)$ of entropy. If the Stark splitting is large then the crystal may already have lost a proportion of its entropy by the time the temperature is sufficiently low for the onset of magnetic ordering. In this case the transition to a Curie-Weiss Law of susceptibility will be gradual, this law holding only at temperatures above where the specific heat anomaly has died out. If Stark splitting alone occurs then at lower temperatures the susceptibility generally shows only a smooth transition from one value of μ_{eff} to another, though, if the lowest of the split levels is a singlet, temperature independent para-magnetism sets in at the lowest temperatures.

3.1 Experimental Behaviour

We are now in a position to describe the actual behaviour of the various rare-earths in relation to the theoretical models that we have been considering. We are concerned here only with the pure metals themselves, but it is worth while reiterating the fact that the salts of the rare-earths, and particularly those in which the element concerned is in a state of high dilution (i.e. atoms well separated), obey the simple theory of paramagnetism down to extremely low temperatures. The pure metals represent the opposite extreme of high concentration of the magnetic ions, and are thus well suited for the study of interatomic interactions under the simplest possible conditions (no foreign atoms present, 4f shell well shielded by outer electrons). Free use has been made in this section of the review article by Spedding, Legvold, Daane and Jennings, 'Some Physical Properties of the Rare-Earth Metals' in 'Progress in Low Temperature Physics', Vol. II, edited by C. J. Gerter (North Holland Publishing Company, Amsterdam).

3.2 Lanthanum

The Paramagnetic susceptibility of lanthanum is small,[5] rising steadily from about 0.73×10^{-6} e.m.u. g^{-1} at room temperature to about 1.5×10^{-6} e.m.u. g^{-1} at 6°K. Below this temperature, usually at about 4.4°K, lanthanum becomes a super-conductor which makes susceptibility measurements extremely difficult. These facts are both in agreement with the view that in metallic lanthanum, as in the free atom, the 4f shell is empty. The feeble paramagnetism is probably due to the conduction band of electrons, formed from atomic 5d and 6s states.

The specific heat of lanthanum has been measured from 2°K up to 200°K[6] and shows no anomalous behaviour apart from the usual peak at the superconducting transition temperature. The Debye characteristic temperature for the specific heat is about 132°K at 4.5°K.

3.3 Cerium

The crystal structure of cerium at room temperature is normally face-centred cubic with a lattice constant of 5.161 A.U., although some samples contain a proportion of the hexagonal close-packed structure. When cooled below 85°K[7,8] or subjected at room temperature to a pressure of 15,000 atmospheres a phase change occurs in which, while remaining face-centred cubic, the lattice constant diminishes by about 6%. This effect shows a remarkable hysteresis with temperature, the room temperature form re-appearing only at about 180°K. It is accompanied by a similar hysteresis effect in the susceptibility[5,10] and this is shown in Figure 5 in which the reciprocal of the susceptibility is plotted against absolute temperature. This is a convenient method of plotting, since a

substance which obeys the Curie-Weiss Law (4) will then give a straight line graph. The appearance of the low temperature contracted phase is accompanied by a drop in the paramagnetic susceptibility, but this is not reversible, the susceptibility only rising to its initial value on warming to nearly 200°K. A further curious feature is observed, namely, that the extent of this thermal hysteresis gradually decreases if the specimen is taken through a large number of thermal cycles, cooling it to 20°K and warming it to room temperature successively. This is shown by the shrinkage of the hysteresis loop in Figure 5 and its almost complete disappearance after about 100 such cycles. The

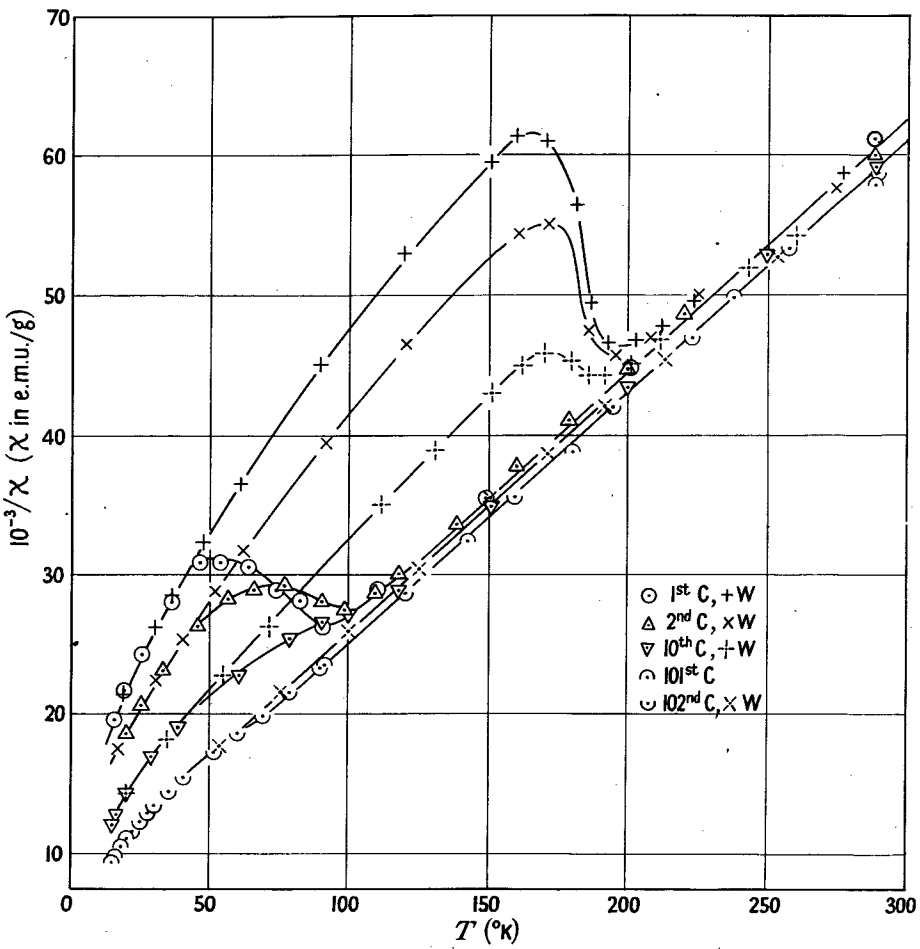

Figure 5 Cerium. Variation of $1/\chi$ with T.
C = cooling, W = warming

theoretical interpretation of the phase change[6,7] seems to be that, on cooling, a proportion of the 4f electrons go into the conduction band and thus cease to contribute appreciably to the paramagnetism of the metal, while the irreversible behaviour must presumably be attributed to the existence of two minima in the free energy, the relative heights of which vary with temperature qualitatively as shown in Figure 6.

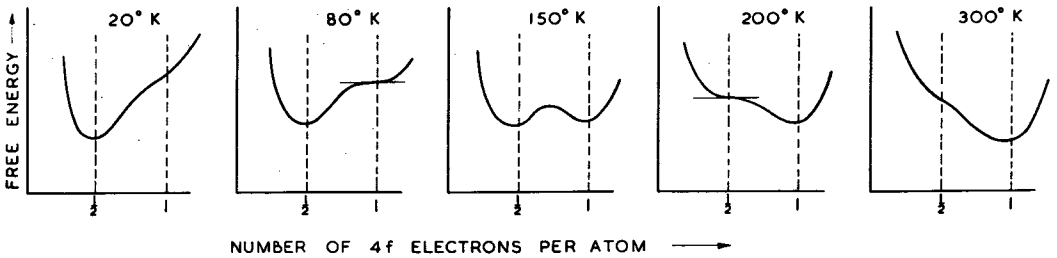

Figure 6 Schematic diagram of free energy of cerium

A specimen examined after a large number of thermal cycles has exhibited a small proportion of the hexagonal close-packed structure which was not present in the virgin material,[5] and it is possible that the appearance of the foreign structure introduced strains large enough to modify the free energy picture given above sufficiently to inhibit the 4f to conduction band transition.

The reversible portion of the susceptibility plot of Figure 5 from room temperature down to about 100°K is a straight line within the limits of experimental error, so that in the absence of the phase change mentioned above cerium obeys the Curie-Weiss Law quite closely over this range, and the effective atomic moment deduced from the Curie constant is about 2.51 μ_B, in good agreement with the theoretical value given in Table 2 (2.54 μ_B). The continuation of the susceptibility measurements below 20°K is shown in Figure 7 and specific heat measurements[11] over the same range in Figure 8. The chief feature here is the peak in the specific heat accompanied by a maximum in the susceptibility at 12.5°K, the magnitudes of these being governed by the extent of the 4f to conduction band transition. It is probable that cerium becomes anti-ferromagnetic at this temperature, the sharpness of the fall in the specific heat on the high temperature side of the peak and the existence of the maximum in the susceptibility both suggesting the onset of this type of magnetic ordering. The relative heights of the specific heat peaks, and the relative magnitudes of the susceptibilities both suggest that in the virgin specimen just about half of the available 4f electrons go into the conduction band on cooling. The

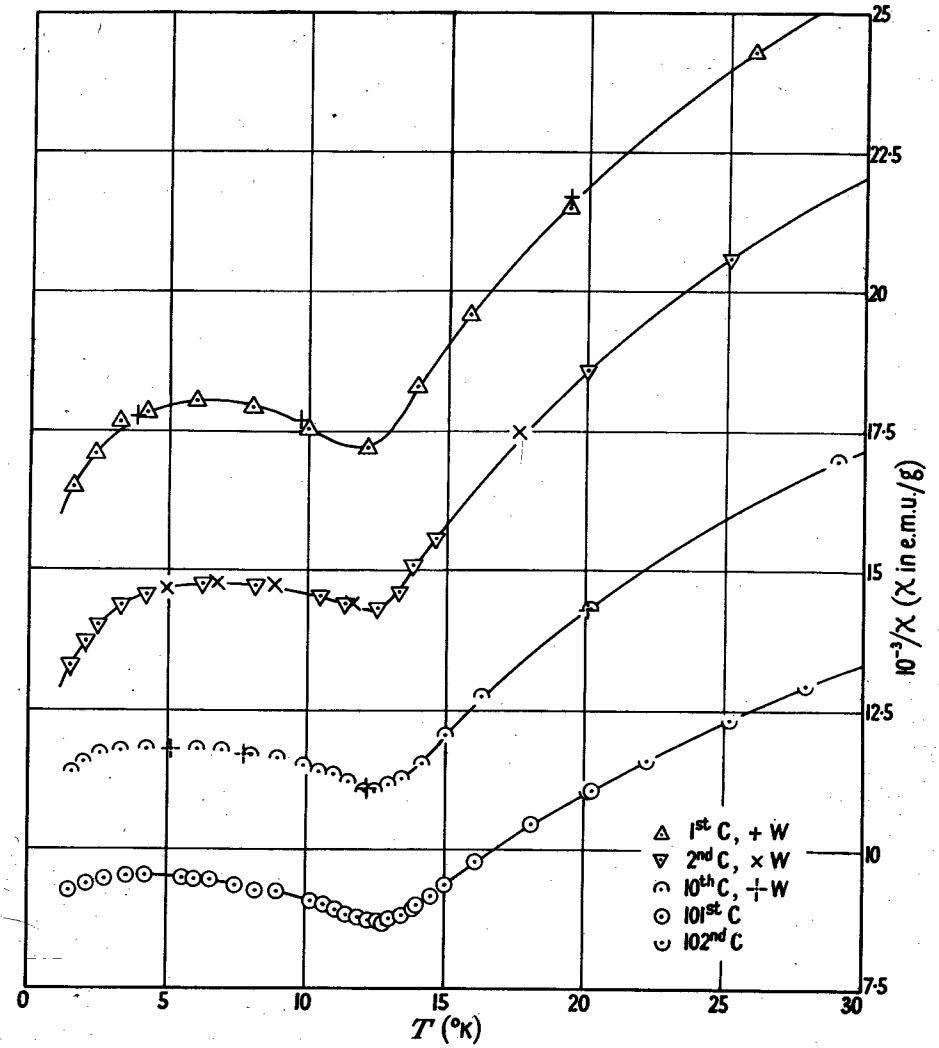

Figure 7 Cerium. Variation of $1/\chi$ with T.
C = cooling, W = warming

entropy involved in the specific heat anomaly for the highest peak has been estimated by subtracting the specific heat of lanthanum as an approximation to the lattice contribution in cerium. It turns out that only about R ℓn 2 of entropy is removed by the appearance of anti-ferromagnetism and, since previous measurements on cerium up to 200°K[6,12] suggest that the room temperature excess entropy is the full R ℓn(2J + 1), (= R ℓn 6) the remainder is probably removed by Stark splitting at a

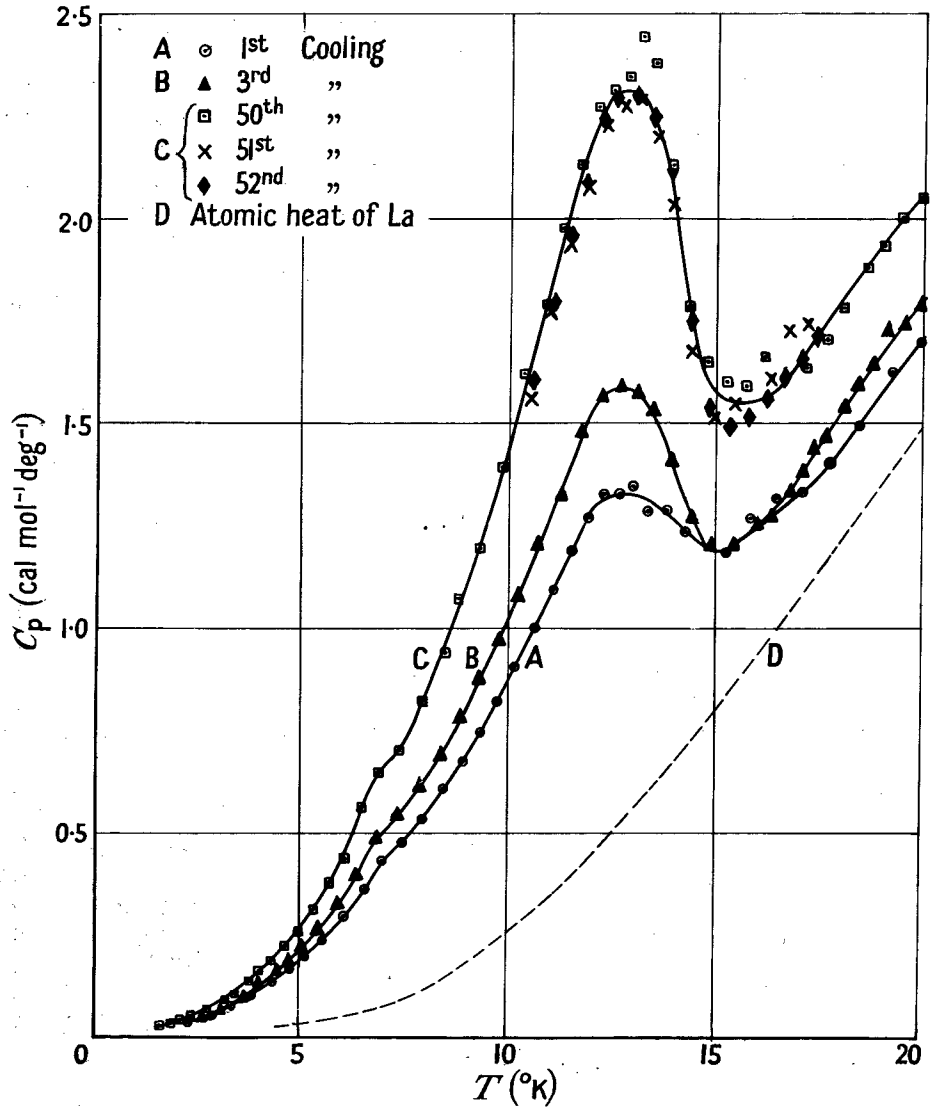

Figure 8 Atomic heat of cerium

higher temperature. This would suggest that anti-ferromagnetism in cerium is a co-operative phenomenon involving only a lowest doublet state in the Stark splitting pattern: this state must be doubly degenerate according to Kramer's Theorem.[4]

3.4 Praseodymium

Praseodymium seems to be the least anomalous of all the rare-earth metals! Down to 100°K its susceptibility[5] obeys the simple Curie Law, $\chi = 11.25/T \times 10^{-3}$ e.m.u. g^{-1}, from which its effective atomic moment is calculated to be 3.56 μ_B, in very close agreement with the theoretical value of 3.58 μ_B for the free ion Pr^{3+} in the $^3H_4$ state (see Table 2). At lower temperatures the susceptibility first increases slightly faster than predicted by the Curie Law and then gradually flattens out at a value of 1.55×10^{-3} e.m.u. g^{-1} below about 4°K. This is completely in agreement with the supposition that the only mechanism involved is Stark splitting of the ground state leaving a lowest singlet energy level. The specific heat anomaly should thus be of the Schottky type, and this is in fact found, a spread-out maximum centred at about 30°K being obtained after subtraction of the lattice contribution.[6] All of the excess entropy, R ℓn 9, can be accounted for in this way when the temperature has reached about 200°K.

3.5 Neodymium

The susceptibility of neodymium has been measured by various workers above 20°K.[13,14,15] The R.R.E. measurements[5] (Figure 10) indicate that it obeys the equation $\chi = \frac{9.47 \times 10^{-3}}{T - 4.3} + 5.0 \times 10^{-6}$ e.m.u. g^{-1} above about 35°K, indicating an effective atomic moment of 3.3 μ_B compared with the theoretical value of 3.62 μ_B for the $^4I_{9/2}$ state. The constant term in the susceptibility can be approximately accounted for on the assumption that the next highest J level, $^4I_{11/2}$, is separated from the ground level by a rather small amount, about 630 cm^{-1}. At higher temperatures the deviation of the $1/\chi$ vs.T plot from a straight line becomes somewhat less marked and a separation of 1350 cm^{-1} is required to account for it.[15] In fact it is doubtful whether the measurements in either case really lead to a reliable value of the energy gap, since it is not known what temperature dependence of susceptibility should be assumed for the d-conduction electrons.

Below 20°K two anomalies appear in the susceptibility,[5] a maximum at 7.5°K and a sharp change in temperature coefficient at 20°K. These temperatures correspond closely to those at which maxima are found in the specific heat,[6] and the lower of them is almost certainly due to the onset of anti-ferromagnetism. The upper anomaly remains at present unexplained, since it seems too sharp to be a simple Stark splitting effect, and in addition this would scarcely account for the sudden change in the gradient of the $1/\chi$ versus T plot. The excess entropy at room temperature again seems to agree with the theoretical value R ℓn 10,[12] and as the two low temperature anomalies account for only about R ℓn 4, the remainder must appear at higher temperatures, probably

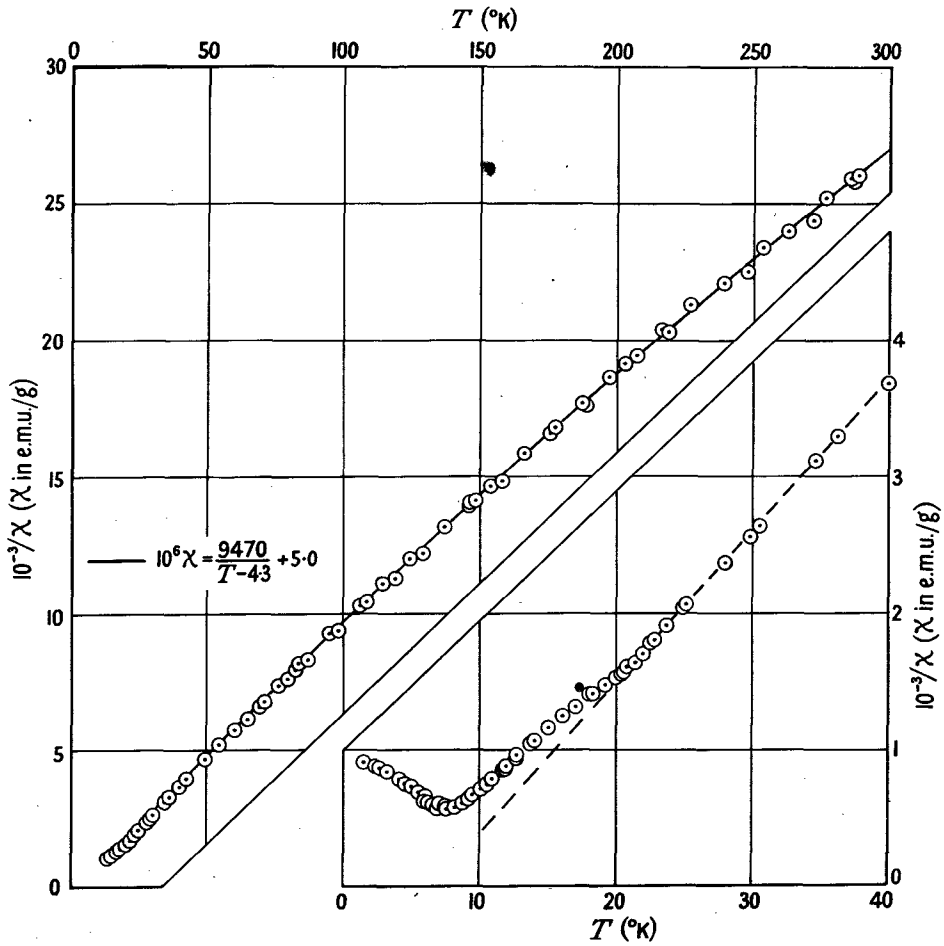

Figure 9 Neodymium. Variation of 1/χ with T

in the form of a spread out Schottky bump as in praseodymium. (It should e emphasized that calculation of the excess magnetic entropy from experimental specific heat results is inevitably rather rough, owing to the uncertainty in the lattice contribution. One can only make some plausible assumption as to how this varies from element to element, the most reasonable being that the Debye characteristic temperature is a linear function of the atomic number, so that it can be interpolated using the known values for lanthanum and gadolinium).

Recently some susceptibility measurements have been made on single crystals of neodymium,[16] and markedly non-linear magnetization curves

have been found at 4.2°K. A fairly sharp 'knee' was observed at about 9000 gauss, and this seems to confirm that the crystal was anti-ferromagnetic at this temperature; it is curious, however, that the effect was found to be comparatively insensitive to the direction of the applied field relative to the crystal axes. A possible explanation may be that the high temperature measurements indicated that the atomic moments favour an alignment perpendicular to the hexad axis, so that there must be at least three equivalent directions in the (0001) plane parallel to which the moments tend to line up. This would certainly reduce the anisotropy of the non-linear effect.

3.6 Promethium

This element does not appear to exist in nature. In the future it is possible that sufficient quantities might become available from atomic piles.

3.7 Samarium

The crystal structure of samarium is very curious,[17] the stacking of layers perpendicular to the c-axis being in the order a b c b c a c a b, giving a unit cell with $4\frac{1}{2}$ times the normal hexagonal close packed lattice constant. This seems to imply an extraordinarily long-ranged 'memory' of the atoms extending over 9 close-packed layers, and may be related therefore to the exceptionally sharp onset of anti-ferromagnetism which apparently occurs at 14.8°K. The susceptibility[5] and specific heat[18] results for samarium obtained at R.R.E. are shown in Figures 10 and 11. The specific heat shows a sharp peak at 13.6°K, and the susceptibility reaches a maximum at 14.8°K, dropping off very rapidly just below this temperature, but rising again below 7°K. There is also a slight kink in the curve between 110°K and 150°K but above this temperature the susceptibility varies by less than 2% and there is no temperature range in which it obeys anything like a Curie-Weiss law. The magnetic contribution to the entropy at 20°K is probably between 0.7 and 0.9 cal mole^{-1} deg^{-1} (depending on the value assumed for the lattice specific heat) and this must be compared with R ℓn 2 = 1.38 cal mole^{-1} deg^{-1}.

The case of samarium is, however, complicated by two features. In the first place, although the normal ground state of the Sm^{3+} ion is $^6H_{5/2}$, the L, S coupling in this state is rather weak and the next highest J level (J = 7/2) is split from it by only a small energy gap, probably about 1000 cm^{-1}, as in the case of neodymium. Redistribution of the atoms between the different J levels as the temperature rises will give a distinctly non-linear plot of $1/\chi$ vs.T, and seems to account reasonably well for the observed susceptibility of various salts of samarium.[19] In the metal anti-ferromagnetic interactions will

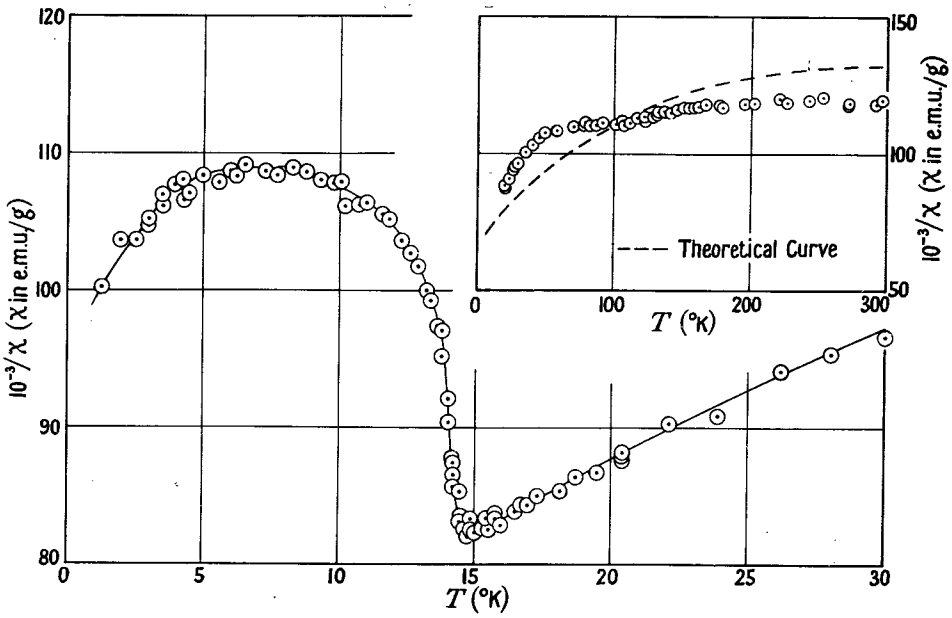

Figure 10 Samarium. Variation of 1/χ with T

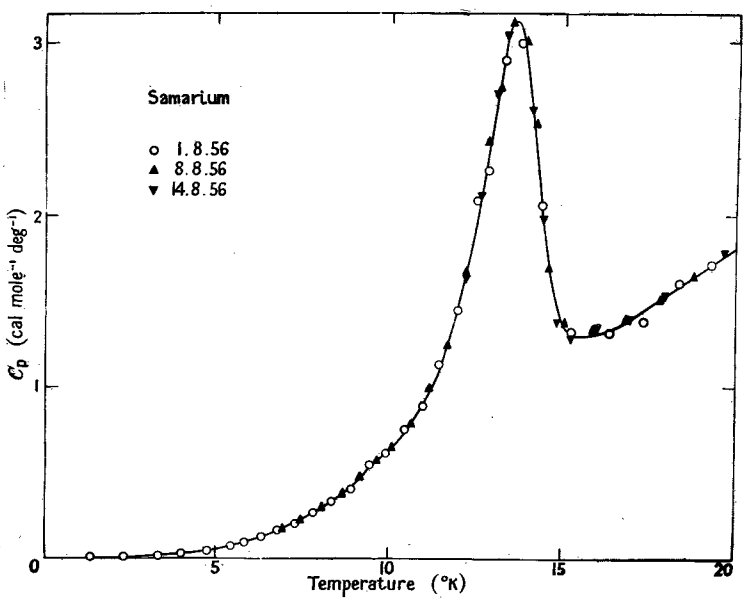

Figure 11 Atomic heat of samarium

certainly affect the theoretical dependence of χ on T and no rigorous treatment is possible. Qualitatively, however, it seems likely that the effect will be to shift the theoretical curve to the left parallel to the T-axis (cf. the effect of such interactions in changing the simple Curie law $\chi = C/T$ to the Curie-Weiss Law $\chi = C/(T-\theta)$, where θ is negative for anti-ferromagnetic substances). Figure 10 shows the theoretical curve of $1/\chi$ vs T moved to the left in this way, assuming $\theta = -60^{\circ}K$, and it is seen that in order of magnitude at least the calculated susceptibility agrees with experiment.

Secondly it is known that samarium has divalent tendencies in its chemical behaviour, though not so markedly as ytterbium and europium. The reason for this tendency in these three elements is the high stability associated with the completely filled and the half-filled 4f shell, in both of which cases the atom is in an S state. The elements which in the trebly ionized condition would have one too few electrons to half fill or completely fill this shell (namely europium and ytterbium respectively) thus tend to prefer to lose only two electrons in combination, forming dipositive ions. Samarium, too, shows this tendency, though having two electrons too few in the tri-positive state, it cannot achieve the S condition even when only doubly ionized, and its divalent compounds are therefore less stable than those of europium and ytterbium. Nevertheless it is possible that a proportion of the atoms in the metal possess six rather than five 4f electrons. They would thus have a $^7F_0$ ground state and make a much smaller contribution to the paramagnetic susceptibility and the entropy than the tripositive ions. Experimentally Figure 10 shows that below $100^{\circ}K$ the susceptibility is anomalously low while the contribution to the entropy at $20^{\circ}K$ due to the magnetic anomaly is certainly less than we should expect if every atom were taking part in the anti-ferromagnetic ordering.

3.8 Europium

Table 1 shows that europium crystallizes in the body-centred cubic system, and has the anomalously low density for the rare-earths of 5.17 gm cm^{-3}. It almost certainly has only two electrons in the conduction band, the 4f shell taking up the extra electron to give the $^8S_{7/2}$ state as in Gd^{3+}. Susceptibility measurements have been made on a relatively pure specimen recently by La Blanchetais and Trombe.[20] They find that it obeys a Curie-Weiss Law with $\theta = 108^{\circ}K$ and an effective atomic moment of 7.12 μ_B, as compared with a theoretical value of 7.94 μ_B for the $^8S_{7/2}$ state. Strong saturation phenomena appear below $104^{\circ}K$, and the metal appears to become ferromagnetic with a Curie temperature of about $90^{\circ}K$. The fact that this is much lower than that of gadolinium ($289^{\circ}K$) is attributed to the more open structure of the metal, each atom having only 8 nearest neighbours, at a distance of 4.0 A.U. compared with 12 in the close-packed system at a distance of 3.64 A.U. in gadolinium.

3.9 Gadolinium

In 1935 Trombe discovered that gadolinium becomes ferromagnetic at about room temperature[21] (289°K), the first pure element to show this property other than the 'classical' metals iron, nickel and cobalt. It also has a far higher saturation moment per atom at 0°K, closely approximating to the theoretical value $Jg\mu_B = 7\mu_B$ for perfect alignment of the spins, while for iron, which has the highest moment of the iron group, the experimental value is $2.2\mu_B$. This high saturation moment is not of great practical value, however, since the large atomic weight of gadolinium, 156.9 compared with 55.8 for iron, reduces the actual saturation intensity of magnetization to about 1950 gauss compared with about 1700 gauss for iron.

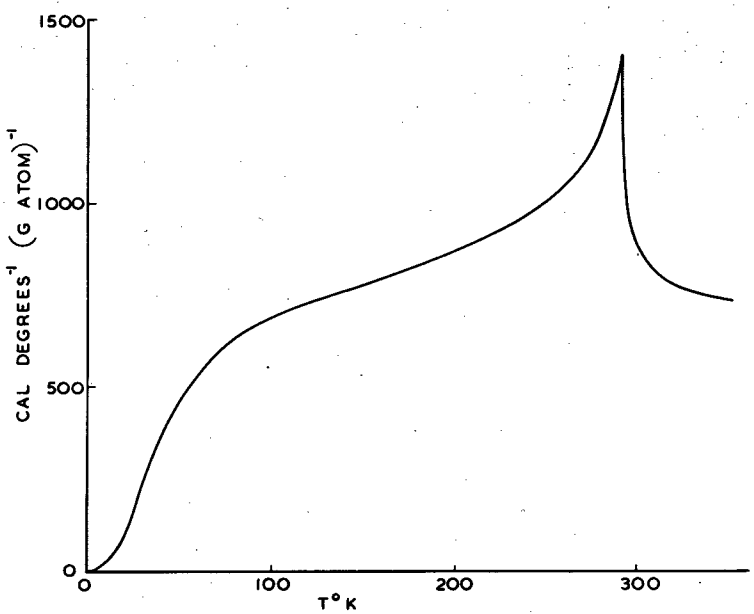

Figure 12 Atomic heat of gadolinium

The susceptibility of gadolinium above the Curie point was also measured by Trombe, and obeys a Curie-Weiss law with an effective moment of $7.95\mu_B$, very close to the theoretical value given in Table 2.

(Note: The fact that the saturation moment of a ferromagnetic at absolute zero is only $Jg\mu_B$, whereas its effective moment in the

paramagnetic state above the Curie point is $g\mu_B\sqrt{J(J+1)}$ is due to the point mentioned in Section 2.2, that the component of angular momentum parallel to the z-axis even when J is 'parallel' to it is exactly equal to $Jh/2\pi$, while the total angular momentum is always $(h/2\pi)\sqrt{J(J+1)}$. Physically we may regard this as being due to the quantum uncertainty about the exact direction of J - even at absolute zero there is a sort of zero-point wobble of J so that even when most nearly parallel to the z-axis its component parallel to that axis is smaller than its absolute value.)

Specific heat measurements[22] show the expected sharp drop just above the Curie temperature (see Figure 12). It is difficult to estimate the excess entropy due to the ferromagnetic ordering, since about 80% of the total entropy at 300°K is due to the lattice, but it is probably about 3.3 cal $mole^{-1}$ deg^{-1} at this temperature compared with $R \ln (2J + 1) = 4.13$ cal $mole^{-1}$ deg^{-1}. As can be seen from Figure 12 the specific heat peak has the usual tail above the Curie point and probably the remaining entropy appears at higher temperatures. The existence of the tail is usually supposed to imply the persistence of short-range magnetic order above the critical temperature, and it is invariably present both in ferromagnetic and anti-ferromagnetic substances.

3.10 Terbium to Thulium

Table 1 shows that the elements from gadolinium to thulium should all have extremely large effective atomic moments, greater than 7 μ_B. In addition to possessing enormous paramagnetic susceptibilities it is not surprising, therefore, to find that they all exhibit magnetic anomalies at comparatively high temperatures. Like gadolinium, terbium, dysprosium, holmium and erbium all become ferromagnetic, but the last three of these show an additional magnetic anomaly on raising the temperature in which there is a transition from the ferromagnetic to the anti-ferromagnetic state before the disordered paramagnetic state appears. Corresponding to these transitions the specific heats of these elements show two well marked peaks, as is shown in Figure 13 for the case of dysprosium.[23] Making a plausible assumption about the lattice specific heat, the magnetic contribution to the entropy is found to be 5.56 cal $mole^{-1}$ deg^{-1} at 300°K, compared with $R \ell n (2J + 1) = 5.51$ cal $mole^{-1}$ deg^{-1}. The upper peak in the specific heat occurs at 174°K, and Figure 14 shows how the susceptibility reaches a maximum at about this temperature,[24] corresponding to the paramagnetic to anti-ferromagnetic transition. Above 220°K the susceptibility obeys the Curie-Weiss law with an effective moment of 10.2 μ_B per atom close to the theoretical value for Dy^{3+} (10.6 μ_B).

Thulium shows a marked peak in susceptibility at 51°K and presumably becomes anti-ferromagnetic at that temperature. Specific

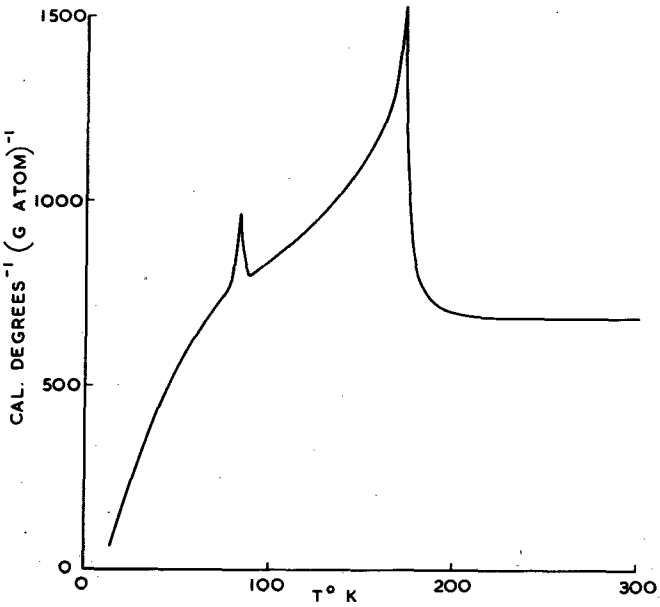

Figure 13 Atomic heat of dysprosium

heat data are not yet available. It shows no signs of a transition to ferromagnetism at lower temperatures.

The properties of the six elements from gadolinium to thulium are summarized in Table 3. The effective moments in the temperature range above the magnetic anomalies are in good agreement with the theoretical values for the tripositive ions in every case.

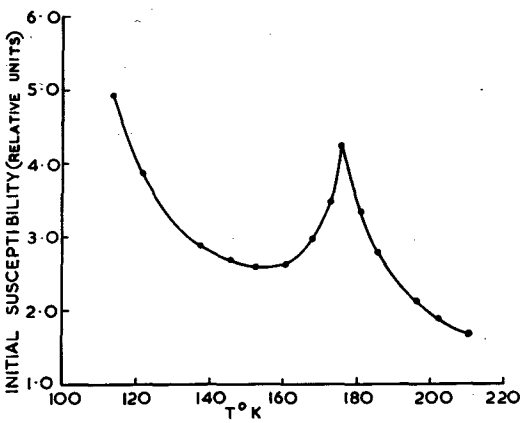

Figure 14 Susceptibility of dysprosium

TABLE 3

Element		Gd	Tb	Dy	Ho	Er	Tm
Effective atomic moment above magnetic anomalies (μ_B)	Theor.	7.94	9.7	10.6	10.6	9.6	7.6
	Expt.	7.95	9.0	10.2	10.9	10.0	7.6
Magnetic order		F	F	AF F	AF F	AF F	AF
Transition temp. (^oK)		289	230	174, 83	132, 20	78, 20	51

3.11 Ytterbium and Lutetium

These two elements are only weakly paramagnetic, the room temperature susceptibilities being less than 10^{-6} emu g^{-1} in both cases. The former has been studied at R.R.E. down to 1.2^oK,[25] and after a suitable correction for the paramagnetism of the conduction electrons has been made, its susceptibility obeys a Curie-Weiss law up to about 40^oK with a Curie constant having the very small value of 56×10^{-6} emu g^{-1}. The most probable interpretation of this result is that in ytterbium metal the majority of the atoms are in the ^1S state with a filled 4f shell, only about 1 atom in 260 being in the $^2F_{7/2}$ state with 13 electrons in the 4f shell. This is in agreement with the greater stability of the filled shell mentioned in Section 3.7.

Lutetium has a filled 4f shell plus three conduction electrons, and there is no possibility of a paramagnetic contribution from the 4f electrons in its case.

4.1 Discussion

The experimental susceptibilities of the rare-earth metals at temperatures above those at which the various anomalies occur lend support to the idea that the 4f shell can be treated in the first approximation by the simple theory of free magnetic ions obeying Russell-Saunders coupling.

The variety and complexity of the different individual types of magnetic behaviour exhibited among the group at low temperatures is in striking contrast to the similarity of the chemical behaviour of the elements, and only in one or two of the simpler cases can it be said that the behaviour of the metals over the whole range of temperature has begun to be understood. The true magnetic anomalies occur at

considerably lower temperatures in the first half of the group (omitting europium from this for the reason mentioned above), so that it seems that the magnitude of the interactions between neighbouring atoms increases roughly with their effective moment. Apart from this it does not seem possible to make any general statement about the types of magnetic interaction which occur, and we must await further experimental and theoretical work for a deeper understanding of the group. It is probable that the preparation of single crystal specimens, particularly of the uniaxial hexagonal metals, would be a great advance, since it would enable such important measurements to be made as the directions of easy magnetization in the ferromagnetic elements. Also a study of anti-ferromagnetism in the rare-earths by neutron diffraction is urgently needed, since this provides the only completely unambiguous proof of this type of ordering.

References

1. F. H. Spedding and J. E. Powell, J. Metals, 6, 1131 (1954).

2. F. H. Spedding and A. H. Daane, J. Metals, 6, 504 (1954).

3. F. H. Spedding, A. H. Daane and K. W. Herrmann, Acta Cryst., 9, 559 (1956).

4. H. A. Kramers, Proc. Acad. Sci. Amst., 33, 959 (1930).

5. J. M. Lock, Proc. Phys. Soc. B, 70, 566 (1957).

6. D. H. Parkinson, F. E. Simon and F. H. Spedding, Proc. Roy. Soc. A, 207, 137 (1951).

7. A. H. Schuch and J. H. Sturdivant, J. Chem. Phys., 18, 142, (1950).

8. M. Foex and F. Trombe, C. R. Acad. Sci., Paris, 217, 501, (1943).

9. A. W. Lawson and Ting-Yuan Tang, Phys. Rev., 76, 307 (1949).

10. C. H. La Blanchetais, C. R. Acad. Sci., Paris, 220, 392 (1945).

11. D. H. Parkinson and L. M. Roberts, Proc. Phys. Soc. B, 70, 471 (1957).

12. R. E. Skochdopole, M. Griffel and F. H. Spedding, J. Chem. Phys., 23, 2258 (1955).

13. F. Trombe, C. R. Acad. Sci., Paris, 198, 1591 (1934).

14. J. F. Elliott, S. Legvold and F. H. Spedding, Phys. Rev., 94, 50 (1954).

15. L. F. Bates, S. J. Leach, R. G. Loasby and K. W. H. Stevens, Proc. Phys. Soc. B, 68, 181 (1955).

16. D. R. Behrendt, S. Legvold and F. H. Spedding, Phys. Rev., 106, 723 (1957).

17. A. H. Daane, R. E. Rundle, H. G. Smith and F. H. Spedding, Acta Cryst., 7, 532 (1954).

18. Lois M. Roberts, Proc. Phys. Soc. B, 70, 434 (1957).

19. S. Freed, J. Amer. Chem. Soc., 52, 2702 (1930).

20. C. H. La Blanchetais and F. Trombe, C. R. Acad. Sci., Paris, 243, 707 (1956).

21. Urbain, Weiss and Trombe, C. R. Acad. Sci., Paris, 200, 2132 (1935).

22. M. Griffel, R. E. Skochdopole and F. H. Spedding, Phys. Rev., 93, 657 (1954).

23. M. Griffel, R. E. Skochdopole and F. H. Spedding, J. Chem. Phys., 25, 75 (1956).

24. J. F. Elliott, S. Legvold and F. H. Spedding, Phys. Rev., 94, 1143 (1954).

25. J. M. Lock, Proc. Phys. Soc. B, 70, 476 (1957).

GAIN-BANDWIDTH RELATIONSHIP FOR AMPLIFIERS

by I. A. D. Lewis

1. Introduction

Many factors are relevant when the performance of a thermionic valve amplifier is to be specified. The properties of voltage gain, transient response, noise factor, maximum power output and linearity have a relative importance which varies with the particular application but in the present article we are only going to consider the voltage gain G and the bandwidth B.

Again, a valve has many properties, any combination of which may decide its suitability for a given purpose. The equivalent noise resistance, lead inductance, input damping at high frequencies, and the maximum linear current swing may well influence the choice but we shall only attempt to take into account here the mutual conductance g_m and the input and output capacitances C_g and C_a respectively.

Our aim is to find a relationship between the quantities B, G, the valve parameters g_m, C_a, C_g, and some function which is characteristic of the type of interstage coupling circuit employed in the amplifier. We shall first see how such a relationship may be arrived at (no proof being offered) and then show that the expression fits in a number of particular cases. The usefulness of the formula for (i) comparing different coupling circuits using the same type of valve, (ii) comparing different valves in the same circuit and (iii) deriving optimum conditions, will then be demonstrated.

2. Definition of Bandwidth

Considering a low-pass (video) amplifier, we may take for B the frequency at which the voltage gain falls to $1/\sqrt{2}$ (-3 dB) of its normal value (on the assumption that the amplitude response curve is fairly smooth). Phase distortion also exists, which, in the case of a distributed amplifier for example, is not always directly correlated

with the amplitude response. Although such distortion may be compensated, at least on paper, by the inclusion of all-pass networks which correct the phase without affecting the amplitude response, phase distortion must be kept in mind as one of the possible fundamental circuit limitations. Thus, in addition to the amplitude bandwidth B_a, say, we must reckon with a phase bandwidth B_p which may be defined as the frequency at which the phase change departs by $\pm$ 1/2 radian from the value which would obtain were the phase response curve of the amplifier perfectly linear (again we suppose that the curve is smooth). For the useful bandwidth B we take whichever is the smaller of the two quantities B_a and B_p i.e.

$$B = \min \begin{Bmatrix} B_a \\ B_p \end{Bmatrix} \tag{1}$$

3. The Gain-Bandwidth Formula

Consider now a single amplifying stage comprising the anode/cathode circuit of one valve, a coupling network, and the grid/cathode circuit of a second valve. For unit change in potential on the grid of the first valve we have a change in anode current of g_m which, on flowing through some sort of effective load resistance R, produces a potential change of g_mR at the grid of the second valve. When N identical stages are connected in cascade the normal overall gain is simply

$$G = (g_mR)^N \tag{2}$$

If the quantity C is representative of the valve capacitances, then C is effectively in parallel with R and a reduction in gain, and a non-linear phase change, occur at a high frequency f such that $1/2\pi fC \sim R$. The bandwidth is accordingly proportional to 1/CR. Clearly, as the number of stages is increased the bandwidth falls and this shrinkage may be taken into account by introducing functions $S_a(N)$ and $S_p(N)$ such that

$$B_a \propto \frac{S_a(N)}{CR} \quad , \quad B_p \propto \frac{S_p(N)}{CR} \tag{3}$$

where the "shrinkage" functions are peculiar to the type of coupling network.

From (1) and (3) we have

$$B \propto \frac{1}{CR} \min \begin{Bmatrix} S_a(N) \\ S_p(N) \end{Bmatrix}$$

$$\therefore \quad R \propto \frac{1}{CB} \min \left\{ \begin{matrix} S_a(N) \\ S_p(N) \end{matrix} \right\}$$

Substituting in (2) we get

$$G \propto \left[\frac{g_m}{BC} \min \left\{ \begin{matrix} S_a(N) \\ S_p(N) \end{matrix} \right\} \right]^N$$

The precise formula, generalized to include distributed amplifiers turns out to be:-

$$G = \left[\frac{NF}{nB} \min \left\{ \begin{matrix} S_a(N) \\ S_p(N) \end{matrix} \right\} \right]^n \qquad (4)$$

where $F = \dfrac{g_m}{4\sqrt{C_a C_g}}$ is the valve figure of merit

(Wheeler's wide-band index)

N = total number of valves arranged in n distributed groups in cascade, each group containing N/n valves.

The formula appears to be valid under the following general conditions:-

(i) interstage transformers, of suitably chosen ratio, are admitted;

(ii) the formula applies to band-pass amplifiers, with the same values of S_a and S_p as obtain for the corresponding low-pass case, provided the (full) bandwidth B is very much less than the centre frequency;

(iii) the load resistance is at our disposal, i.e. grounded-grid and transistor circuits are excluded, unless the input resistance of each stage happens to supply just the value of R that is required by the preceding stage;

(iv) stray capacitances, and any extra capacitance due to the coupling network (as might be introduced by the use of

transmission line sections for example), which are thrown directly across the valve electrodes must be included in the values assigned to C_a and C_g which occur in F;

(v) there is no feedback between the valve input and output circuits.

The formula (4), subject to the above restrictions, seems to be the best compromise between simplicity and generality that can be achieved. A satisfactory feature is the fact that the overall performance of the amplifier (G,B) can be related to the quantity F, which depends upon the valve alone, and the S functions, which depend upon the type of coupling circuit alone (and the total number of valves).

4. Resistance Coupling

The equivalent circuit to be analysed is shown in Figure 1. The transformer is assumed to be perfect.

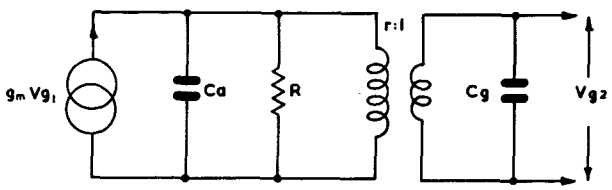

Figure 1 Resistance coupling

It may readily be shown that the (complex) voltage gain function $\overline{G} = V_{g_2}/V_{g_1}$ for N such stages in cascade is

$$\overline{G} = \left[\frac{g_m R/r}{1 + j\omega(C_a + C_g/r^2)R}\right]^N$$

The low-frequency gain is

$$G = (g_m R/r)^N \qquad (5)$$

and on putting

$$x = \omega(C_a + C_g/r^2) \qquad (6)$$

we can therefore write

$$\overline{G} = \frac{G}{(1 + jx)^N}$$

The amplitude bandwidth is determined by the relation

$$|\overline{G}| = G/\sqrt{2}$$

and thus

$$(1 + x^2)^{\frac{N}{2}} = \sqrt{2}$$

$$\therefore \quad x = \sqrt{2^{1/N} - 1} = x_a \quad \text{say} \tag{7}$$

Using the series $a^y = 1 + y.\ln a + \frac{(y.\ln a)^2}{2} + \ldots.$

we find

$$x_a \simeq \sqrt{\frac{\ln 2}{N}} \left(1 + \frac{\ln 2}{4N} + \ldots..\right)$$

$$\therefore \quad x_a \simeq \frac{0.833}{N^{\frac{1}{2}}} \tag{8}$$

to an accuracy of within 4% for $N \geqslant 4$.

Let us now work out the phase bandwidth. The phase change β (radians) per stage is given by

$$\tan \beta = x$$

and the phase error, per stage, is

$$\beta - \omega\left(\frac{d\beta}{d\omega}\right)_{\omega=0} = \beta - x\left(\frac{d\beta}{dx}\right)_{x=0} = \beta - x(\cos^2\beta)_{x=0}$$

$$= \beta - x = \tan^{-1}x - x$$

The phase bandwidth is defined by setting the overall phase error, for N stages, equal to $\pm$ 1/2 radian, i.e. we have to find the solution x_p of the equation

$$\tan^{-1}x_p - x_p = \pm 1/2N \tag{9}$$

This equation may be solved graphically; alternatively we can use the series $\tan^{-1}x = x - x^3/3 + x^5/5 - \ldots$, which applies provided $x^2 \leqslant 1$, and we then find

$$x_p \backsimeq \left(\frac{3}{2N}\right)^{\frac{1}{3}} \left[1 + \frac{1}{5}\left(\frac{3}{2N}\right)^{\frac{2}{3}} + \ldots\ldots\right]$$

$$\therefore \quad x_p \backsimeq \frac{1.15}{N^{\frac{1}{3}}} \tag{10}$$

to an accuracy of within 10% for $N \geqslant 4$.

Now B_a and B_p are given by (6) in terms of x_a and x_p. Thus we have, for the useful bandwidth,

$$B = \min\begin{Bmatrix} B_a \\ B_p \end{Bmatrix} = \frac{1}{2\pi(C_a + C_g/r^2)R} \min\begin{Bmatrix} x_a \\ x_p \end{Bmatrix}$$

$$\therefore \quad R = \frac{1}{2\pi B(C_a + C_g/r^2)} \min\begin{Bmatrix} x_a \\ x_p \end{Bmatrix}$$

Substituting for R in (5) we have

$$G = \left[\frac{g_m}{2\pi B(rC_a + C_g/r)} \min\begin{Bmatrix} x_a \\ x_p \end{Bmatrix}\right]^N$$

By differentiating with respect to the transformer turns ratio r it can easily be shown that this has a maximum value when $r = \sqrt{C_g/C_a}$; thus, finally

$$G = \left[\frac{g_m}{4\pi B\sqrt{C_a C_g}} \min\begin{Bmatrix} S_a \\ S_p \end{Bmatrix}\right]^N$$

where $S_a = x_a$ as given by (7)

and $S_p = x_p$ as given by (9).

This result is in agreement with the formula (4) for the case when $N = n$. The S functions are plotted in Figure 4(I).

5. Π-Section Coupling

The circuit to be analysed is depicted in Figure 2. The (complex) voltage gain function for N such stages in cascade is

$$\overline{G} = \left\{ \frac{g_m R/r}{1 - \omega^2 LC_a + j\left[\omega(C_a + C_g/r^2)R - \omega^3 LC_a C_g R/r^2\right]} \right\}^N \qquad (11)$$

and, as before, the low frequency gain is

$$G = (g_m R/r)^N \qquad (12)$$

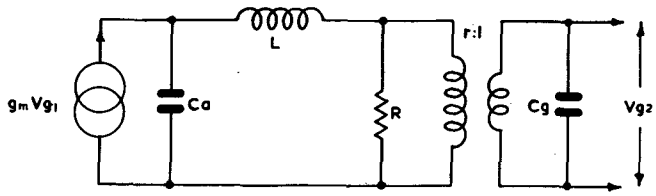

Figure 2 π-section coupling

It follows that the gain at angular frequency ω is

$$|\overline{G}| = \frac{G}{(1 + a_2\omega^2 + a_4\omega^4 + a_6\omega^6)^{N/2}} \qquad (13)$$

where

$$a_2 = (C_a + C_g/r^2)^2 R^2 - 2LC_a$$

$$a_4 = L^2 C_a^2 - 2LC_a C_g (C_a + C_g/r^2) R^2/r^2$$

$$a_6 = L^2 C_a^2 C_g^2 R^2/r^4$$

By suitable choice of the circuit parameters we can make a_2 and a_4 zero, i.e. have an amplitude response which is maximally flat. The conditions are

$$r = \sqrt{3C_g/C_a} \qquad (14)$$

$$L = \frac{8}{9} C_a R^2$$

The expression (13) for the gain then reduces to

$$|\overline{G}| = \frac{G}{(1 + x^6)^{N/2}}$$

where $$x = \frac{2}{3}\omega C_a R \qquad (15)$$

The amplitude bandwidth is given by

$$(1 + x^6)^{N/2} = \sqrt{2}$$

$$\therefore \quad x = (2^{1/N} - 1)^{1/6} = x_a \text{ say} \qquad (16)$$

Using the series expansion, as before, we can write

$$x_a \simeq \left(\frac{\ln 2}{N}\right)^{\frac{1}{6}} \left[1 + \frac{\ln 2}{12N} + \ldots\right]$$

$$\therefore \quad x_a \simeq \frac{0.941}{N^{1/6}} \qquad (17)$$

to an accuracy of within about 1% for $N \geqslant 4$.

In order to find the phase bandwidth we must first substitute the conditions (14) into (11). The phase change β per stage then turns out to be given by

$$\tan\beta = \frac{2x - x^3}{1 - 2x^2}$$

The phase error, per stage, is thus

$$\beta - x\left(\frac{d\beta}{dx}\right)_{x=0} = \beta - 2x(\cos^2\beta)_{x=0} = \tan^{-1}\left[\frac{2x - x^3}{1 - 2x^2}\right] - 2x$$

and accordingly the phase bandwidth is defined by the solution x_p of

$$\tan^{-1}\left[\frac{2x_p - x_p^3}{1 - 2x_p^2}\right] - 2x_p = \pm 1/2N \qquad (18)$$

This equation has been solved graphically; alternatively, using series expansions we find

$$x_p \simeq \frac{3}{2N}^{\frac{1}{3}}\left[1 - \frac{2}{5}\left(\frac{3}{2N}\right)^{\frac{2}{3}} + \ldots\right]$$

$$\therefore \quad x_p \simeq \frac{1.15}{N^{1/3}} \tag{19}$$

to an accuracy of within 10% for $N \geq 12$.

B_a and B_p are given by (15) in terms of x_a and x_p, thus, for the useful bandwidth we have

$$B = \min\begin{Bmatrix} B_a \\ B_p \end{Bmatrix} = \frac{3}{4\pi C_a R} \min\begin{Bmatrix} x_a \\ x_p \end{Bmatrix}$$

$$\therefore \quad R = \frac{3}{4\pi B C_a} \min\begin{Bmatrix} x_a \\ x_p \end{Bmatrix}$$

Substituting in (12) and using (14) we then find

$$G = \left[\frac{g_m}{4\pi B \sqrt{C_a C_g}} \min\begin{Bmatrix} S_a \\ S_p \end{Bmatrix} \right]^N$$

where $S_a = \sqrt{3} x_a$ as given by (16)

$S_p = \sqrt{3} x_p$ as given by (18)

The formula (4) again applies with $N = n$; the S functions are plotted in Figure 4 (II).

6. Distributed Amplifier with Constant-k Filter Sections

The circuit of a pair of cascaded distributed amplifier stages is shown in Figure 3. The anode line of the first group is matched into the grid line of the second group by means of an ideal transformer of step-down voltage ratio $r = \sqrt{Z_{oa}/Z_{og}}$.

For the amplifier to function, the delay per section in the grid and anode lines must be the same, i.e.,

$$L_g C_g = L_a C_a$$

accordingly, since $Z_o = \sqrt{L/C}$, we find

$$r = \sqrt{C_g/C_a} \tag{20}$$

The angular cut-off frequency is given by

$$\omega_c = 2/\sqrt{L_g C_g} = 2/\sqrt{L_a C_a} \qquad (21)$$

and is the same for both lines.

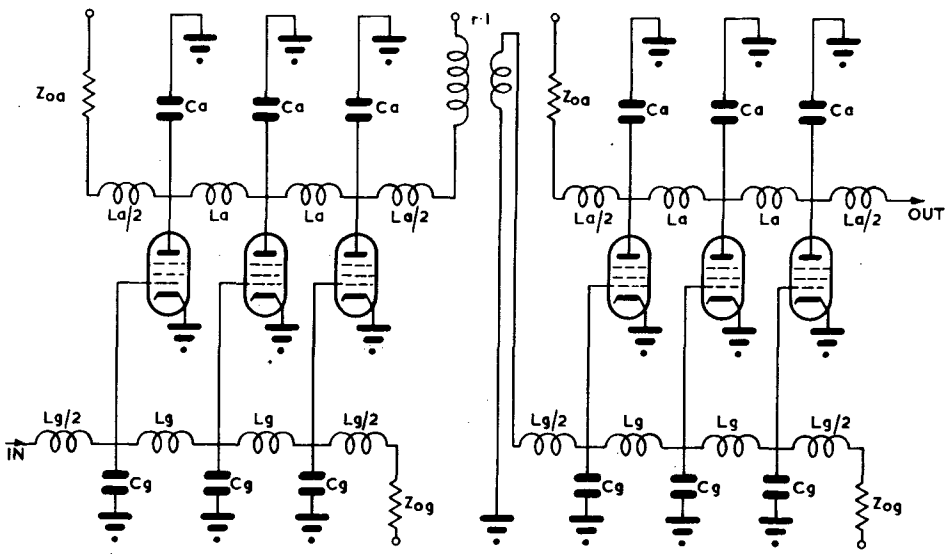

Figure 3 Distributed amplifier

If we ignore the variations in amplitude response due to the change of characteristic impedance of the filter sections with frequency, and also due to coil loss and valve damping, we can take the amplitude bandwidth to be simply

$$B_a = \omega_c/2\pi \qquad (22)$$

The phase change per section can be shown to be given by

$$\sin \beta/2 = x$$

where

$$x = \omega/\omega_c \qquad (23)$$

Now for an amplifier containing N valves in all, each element of signal, which goes to build up the output, traverses N filter sections

independent of the path followed. Thus the phase error for N valves is

$$N\left[\beta - x\left(\frac{d\beta}{dx}\right)_{x=o}\right] = N\left[2\sin^{-1}x - x\left(\frac{2}{\sqrt{1-x^2}}\right)_{x=o}\right] = 2N\left[\sin^{-1}x - x\right]$$

The phase bandwidth is accordingly given by the solution x_p of the equation

$$\sin^{-1}x_p - x_p = \pm 1/4N \qquad (24)$$

This equation may be solved graphically; alternatively, using the series $\sin^{-1}x = x + x^3/6 + 3x^5/40 + \ldots$ (provided $x^2 \leq 1$) we can write

$$x_p \simeq \left(\frac{3}{2N}\right)^{\frac{1}{3}}\left[1 - \frac{6}{40}\left(\frac{3}{2N}\right)^{\frac{2}{3}} + \ldots\right]$$

$$\therefore \quad x_p \simeq \frac{1.15}{N^{\frac{1}{3}}} \qquad (25)$$

to an accuracy of within about 7% for $N \geq 4$.

Employing (22) and (23), we therefore have for the useful bandwidth

$$B = \frac{\omega_c}{2\pi}\min\left\{\begin{matrix}1\\x_p\end{matrix}\right\} \qquad (26)$$

Now the normal overall gain for n groups is

$$G = \left[\frac{N}{n}\cdot\frac{g_m Z_{oa}}{2}\cdot\frac{1}{r}\right]^n$$

Substituting for r from (20) and using (21) we find $Z_{oa}/r = 2/\omega_c\sqrt{C_aC_g}$. Employing this result, and substituting for ω_c from (26) we finally get

$$G = \left[\frac{Ng_m}{4\pi nB\sqrt{C_aC_g}}\min\left\{\begin{matrix}S_a\\S_p\end{matrix}\right\}\right]^n$$

where $S_a = 2$

and $S_p = 2x_p$ as determined by 24.

Our conclusion is in agreement with the formula (4); the function S_p is plotted in Figure 4(III).

7. Distributed Amplifier with m-Derived Filter Sections

The arguments and formulae of the preceding section again apply except that (i) we must use a different expression for the phase change per section, namely

$$\sin \beta/2 = mx \left[1 + (m^2 - 1)x^2\right]^{-1/2}$$

and (ii) put in the values C_{ak} and C_{gk}, for the prototype constant-k sections, in place of the actual capacitances C_a and C_g.

The capacitances are linked by the relations

$$C_{ak} = C_a/m \quad , \quad C_{gk} = C_g/m$$

Thus our formula for the gain becomes

$$G = \left[\frac{2m\, N g_m}{4\pi n B \sqrt{C_a C_g}} \min\left\{\frac{1}{x_p}\right\}\right]^n \qquad (27)$$

with a new form for x_p.

Using the series for $\sin^{-1} y$ the phase change, per section, is found to be

$$2 \sin^{-1} mx \left[1+(m^2-1)x^2\right]^{-1/2} = 2mx + \frac{mx^3}{3}(3-2m^2) + \frac{mx^5}{20}(8m^4-20m^2+15) + \ldots$$

By choosing $m = \sqrt{3/2} = 1.23$ we can make the x^3 term vanish i.e. make the phase response more linear. Using this value of m, the required values of x_p are the solutions of

$$\sin^{-1} \frac{1.23x_p}{\sqrt{1 + x_p^2/2}} - 1.23x_p = \pm 1/4N \qquad (28)$$

This equation has again been solved graphically; series expansion yields

$$x_p \doteq \left(\frac{10}{3}\sqrt{\frac{2}{3}} \cdot \frac{1}{N}\right)^{\frac{1}{5}} \left[1 - \frac{1}{21}\left(\frac{10}{3}\sqrt{\frac{2}{3}} \cdot \frac{1}{N}\right)^{\frac{2}{5}} + \ldots\right]$$

$$\therefore \quad x_p \approx \frac{1.22}{N^{\frac{1}{5}}} \qquad (29)$$

to an accuracy of within about 4% for $N \geqslant 4$.

Let us return now to (27). On taking the factor $2m = 2.46$ inside the curly brackets we have finally

$$G = \left[\frac{N}{nFB} \min\left\{ \begin{matrix} S_a \\ S_p \end{matrix} \right\} \right]^n$$

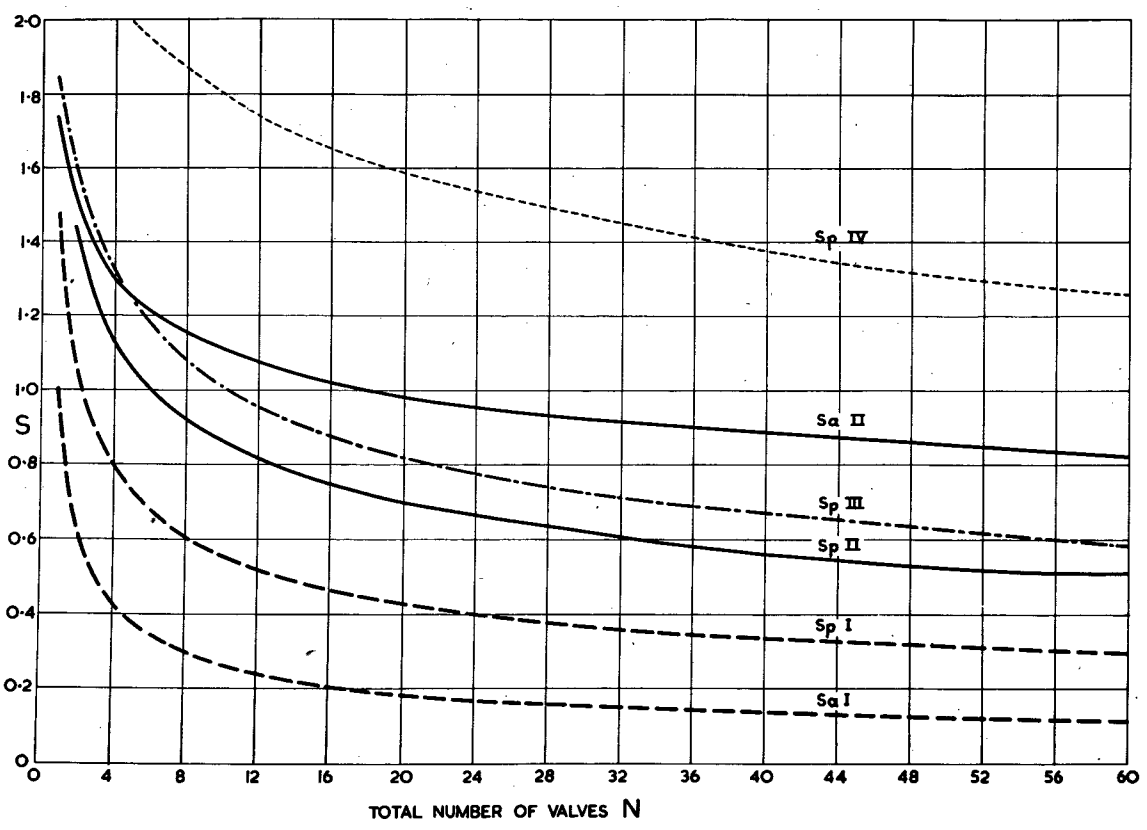

Figure 4 Graphs of S(N)

where $S_a = 2.46$

$S_p = 2.46x_p$ as given by (28)

The function S_p is plotted in Figure 4(IV).

8. Deductions from the Formula

It is obvious at once from (4) that we should use a valve with as high a figure of merit as possible (other things being equal) and select a network with a large S_a and S_p. The values of F for some useful valves are tabulated below, and the graphs of the S functions for a number of circuits have already been given - see Figure 4.

	g_m (mA/V)	C_g (pF)	C_a (pF)	$F = \frac{g_m}{4\pi\sqrt{C_a C_g}}$ (Mc/s)
EF91 CV138	$7\frac{1}{2}$	10	3	110
E180F CV3998	$16\frac{1}{2}$	11	4	200
E2133 CV2276	19	$10\frac{1}{2}$	$4\frac{1}{2}$	220

N.B. About 25% has been added to the (cold) input capacitance quoted by the manufacturers to allow for the increase which occurs when the valve is operating; a further 1 pF has been added to both input and output figures to allow for wiring strays.

The graphs show that amplitude bandwidth is the limitation in simple resistance coupling whereas phase bandwidth is the limiting factor in all other cases; it must be emphasized, however, that the effect of grid damping, which is not taken into account in the formula, is important in the case of distributed amplifiers. The improvement of π-section coupling over simple resistance coupling is demonstrated, and the marked superiority of m-derived over constant-k lines in a distributed amplifier is also made evident.

The formula (4) leads to a number of other results. Writing $S = \begin{Bmatrix} S_a(N) \\ S_p(N) \end{Bmatrix}$ for short and differentiating we have

$$\frac{1}{n}\frac{dG}{G} + \frac{dB}{B} = \frac{dF}{F} + \frac{dS}{S} + \frac{dN}{N} + \left(\frac{1}{n}\ln G - 1\right)\frac{dn}{n} \qquad (30)$$

By means of this equation we can work our precisely the changes (provided they are small) which will occur in G or B when the type of valve, the type of circuit, or the number of valves is altered. Note that when the type of coupling circuit is fixed we put $dS = \frac{dS}{dN} \cdot dN$ or if the number of groups N/n is fixed we employ the relation $\frac{dN}{N} = \frac{dn}{n}$. Two or three results are of special interest.

(i) Suppose we decide to have only one valve per group (i.e. consider a non-distributed amplifier). Then N = n and (4) and (30) become respectively

$$G = (FS/B)^N \tag{31}$$

$$\frac{1}{N}\frac{dG}{G} + \frac{dB}{B} = \frac{dF}{F} + \frac{dS}{S} + \frac{\ln G}{N}\frac{dN}{N} \tag{32}$$

If the bandwidth, the type of valve, and the type of circuit are fixed we put $dB = dF = 0$ and $dS = \frac{dS}{dN} \cdot dN$; (32) then reduces to

$$\frac{1}{N}\frac{dG}{G} = \left(\frac{N}{S}\frac{dS}{dN} + \frac{\ln G}{N}\right)\frac{dN}{N}$$

For G to be a maximum, for variations in N, we write $dG = 0$, then

$$\ln G = -\frac{N^2}{S}\frac{dS}{dN}$$

$$\therefore G = \left[e^{-\frac{N}{S}\frac{dS}{dN}}\right]^N \tag{33}$$

or, on substituting in (31)

$$e^{-\frac{N}{S}\frac{dS}{dN}} = FS/B \tag{34}$$

Since F, B and the type of circuit have been chosen this equation enables us to find N and S. Relation (31) (or (33)) then yields the maximum possible gain that can be obtained with a given F and B and type of circuit, no matter how many valves we are prepared to employ.

Now in all the types of circuit analysed so far we have found that

$$\left.\begin{matrix} S_a \\ S_p \end{matrix}\right\} \longrightarrow K/N^p \qquad (35)$$

for sufficiently large values of N. The constants K and p, obtained from equations (8), (10), (17), (19), (25), (29), are summarized in the table

	K	p
Resistance coupling, amplitude	0.833	1/2
.. .. , phase	1.15	1/3
Π-section .. , amplitude	1.63	1/6
.. .. , phase	1.98	1/3
Distributed, constant-k, phase	2.29	1/3
.. , m-derived , phase	2.99	1/5

We note at once from (35) that

$$\frac{N}{S}\frac{dS}{dN} \longrightarrow -p$$

thus (34), together with (35), yields for the maximum useable number of (single) stages

$$e^p = \frac{FS}{B} = \frac{KF}{BN^p}$$

$$\therefore \quad N = \frac{1}{e}\left(\frac{KF}{B}\right)^{1/p} = N_o \text{ say} \qquad (36)$$

The corresponding maximum overall gain (33) becomes

$$G = e^{N_o p} \qquad (37)$$

Suppose, by way of example, we want an amplifier of 100 Mc/s. bandwidth and wish to use a valve with F = 200 Mc/s. Using the values K = 0.833, p = 1/2, given in the table for simple resistance coupling, and putting F/B = 2 we find

$N_o \simeq 1$ from (36)

and $G \simeq e^{\frac{1}{2}}$ from (37)

On the other hand, for π-section coupling, $K = 1.98$ and $p = 1/3$,

whence $N_0 \simeq 23$

and $G \simeq e^8$

(ii) Suppose now that we have a distributed amplifier in which F, B, N and S are all given. We want to find the best way in which the valves should be grouped. Putting $dG = 0$ for the gain to be a maximum we find from (30) that

$$\left(\frac{\ln G}{n} - 1\right)\frac{dn}{n} = 0$$

$$\therefore \quad G = e^n$$

i.e. we have an optimum arrangement when the gain per group is equal to $e = 2.7$. Substituting in (4) we then have

$$\frac{NFS}{nB} = e \qquad (38)$$

Taking approximately the same numerical values as before, i.e. $F = 200$ Mc/s., $N = 24$ valves in all, and $G = e^8$, let us see what bandwidth may be realized using an m-derived distributed amplifier. First, we must group the valves in the optimum manner such that the gain per group is equal to e. It follows that we must take $n = 8$; this value is quite suitable since it makes N/n an integer. Putting $F = 200$ Mc/s., $N = 24$, $n = 8$ in (38) and using (35) with $K = 2.99$ and $p = 1/5$ we find

$$B = \frac{NFS}{ne} = \frac{KFN^{(1-p)}}{ne} = 350 \text{ Mc/s.}$$

Thus, by using a distributed system we get three and a half times the bandwidth for the same gain, and about the same number of valves, compared with the value obtainable with π-section coupling.

(iii) Finally, consider a distributed amplifier with F and B given and the type of network fixed. Let us suppose that the valves are grouped in the optimum manner such that $G = e^n$.

Putting $dS = \frac{dS}{dN} dN$ in (30) we find

$$\frac{1}{n}\frac{dG}{G} = 1 + \frac{N}{S}\frac{dS}{dN}\frac{dN}{N}$$

By making use of (35) we see that

$$\frac{dG}{dN} \propto 1 - p$$

Now $0 < p < 1$ thus dG/dN is always positive, i.e. we can increase the gain as much as we please by adding more valves.

We know that if the bandwidth required from an amplifier is so large that the gain per stage of a simple amplifier (one valve per stage) would turn out to be less than unity, then it would be futile to connect the valves in cascade and that one would of necessity have to use a distributed system. The above considerations show that if the gain per stage falls below e (rather than unity) a distributed system is called for.